"On the Shoulders of Giants"
A brief excursion into the history of mathematical programming [1]

R. Tichatschke
Department of Mathematics, University Trier, Germany

Similar to many mathematical fields also the topic of mathematical programming has its origin in applied problems. But, in contrast to other branches of mathematics, we don't have to dig too deeply into the past centuries to find their roots. The historical tree of mathematical programming, starting from its conceptual roots to its present shape, is remarkably short, and to quote ISAAK NEWTON, we can say:

"We are standing on the shoulders of giants".

The goal of this paper is to describe briefly the historical growth of mathematical programming from its beginnings to the seventies of the last century and to review its basic ideas for a broad audience. During this process we will demonstrate that optimization is a natural way of thinking which follows some extremal principles.

1 The Giants

Let us start with LEONHARD EULER.

Leonhard Euler (1707-1783)

1727: Euler has been appointed professor (by Daniel Bernoulli) at the University of Saint Petersburg, Russia.

Member of St. Petersburg's Academy of Science and since 1741 member of the Prussian Academy of Science, Berlin.

1744: *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici latissimo sensu accepti*.
(*Method to find curves, possessing some property in the most or smallest degree or the resolution of the Isoperimetric problem considered in the broadest sense.*)
In this work he has established the variational analysis in a systematic way.

[1]Part of a lecture held at the University of Trier on the occasion of the Year of Mathematics 2008 in Germany; published in: Discussiones Mathematicae, vol. 32, 2012.

He is the most productive mathematician of all times (his oeuvre consists of 72 volumes) and as one of the first he captured the importance of the optimization. He wrote [33]: ”Whatever human paradigm is manifest, it usually reflects the behavior of maximization or minimization. Hence, there are no doubts at all that natural phenomena can be explained by means of the methods of maximization or minimization.”

It is not surprising why optimization appears as a natural thought pattern. Thousands of years human beings have sought solutions for problems which require a minimal effort and/or a maximal revenue. This approach has contributed to the growth of all branches of mathematics. Moreover, the thought of optimizing something has entered nowadays many disciplines of science.

Back to EULER. He has delivered important contributions on the field of optimization in both theory and methods. His characterization of optimal solutions, i.e. the description of necessary optimality conditions, has founded the *variational analysis*. This topic treats problems, where one or more unknown functions are sought such that some definite integral, depending on the chosen function, attains its largest or smallest value.

$$\int_{t_0}^{t_1} L(y(t), y'(t), t)dt \to \min!, \qquad y(t_0) = a,\ y(t_1) = b. \tag{1}$$

A famous example is the *Brachistochrone problem*:

Problem: *Find the path (curve) of a mass point, which moves in shortest time under the influence of the gravity from point $A = (0,0)$ to point $B = (a,b)$:*

$$\mathcal{J}(y) := \int_0^a \sqrt{\frac{1 + y'^2(x)}{2gy(x)}}dx \ \to\ \min!, \qquad y(0) = 0,\ y(a) = b.$$

This problem had been formulated already in 1696 by JOHANN BERNOULLI, and it is known that he always quarreled with his brother JACOB BERNOULLI, who found the correct solution to this problem, but was unable to prove it. In 1744 Euler answered this question by proving the following theorem.

Theorem *Suppose $y = y(t),\ t_0 \leq t \leq t_1$, is a C^2-solution of the minimization problem (1), then the (Euler)-equation holds:*

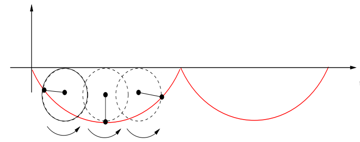
$$\frac{d}{dt}L_{y'} - L_y = 0.$$

In the case of the Brachistochrone this equation has the particular form (because L does not depend on time t):

$$\frac{d}{dt}(y'L_{y'} - L) = 0.$$

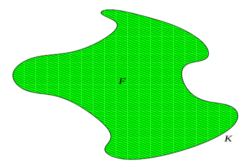
Solving this differential equation one gets the sought solution as arc of a cycloid.

Solution: Cycloid
$x(t) = c_1 + c(t - \sin t); \quad y(t) = c(1 - \cos t).$
$0 \le t \le t^*.$
The constants c, c_1 and t^* are determined by the boundary conditions.



The cycloid describes the behavior of a tautochrone, meaning that a mass point $(x(t), y(t))$ sliding down a tautochrone-shaped frictionless wire will take the same amount of time to reach the bottom no matter how high or low the release point is. In fact, since a tautochrone is also a brachistochrone, the mass point will take the shortest possible time to reach the bottom out of all possible shapes of the wire.
EULER is also one of the first who used *methods of discrete approximation* for solving variational problems. With this method he has solved, for instance, the well-known *Isoperimetric problem*:

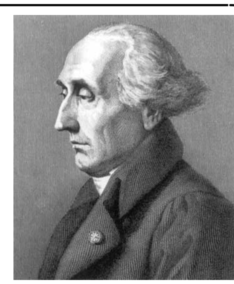
Problem: *Under all closed curves K of length L, enclosing the area F, find the one which maximizes the area F.*



Solution: K - circle with circumference L.

In today's language of optimization this problem can be considered as a maximization problem subject to a constraint, because the length L is understood as a restriction.
More than 200 years later C. CARATHÉODORY (1873-1950) has described Euler's variational analysis as "one of the most beautiful mathematical works, which has been ever written" [13].

Joseph Louis Lagrange (1736-1813)



1755: Professor for mathematics at the Royal Artillery School in Turin.

1757: He is one of the founders of the Academy of Science in Turin.

1766: Director of the Prussian Academy of Science in Berlin and successor of Euler.

Accomplisher of the building of Newton's mechanics, worked also in selestical mechanics, Algebra and number theory.

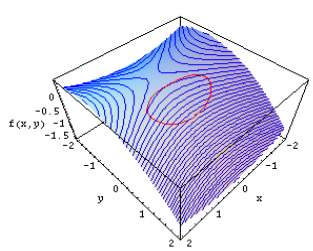
1762: *Multivariable Variational Analysis*,
1788: *Méchanique analytique*.

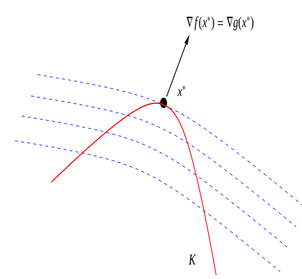
In 1762 LAGRANGE simplified Euler's deduction of the necessary optimality conditions and was able to generalize these conditions (so called Euler-Lagrange-equation) for multivariate functions [70], [71]. His starting point has been the equations of motions in the mechanics. Dealing with the movement of mass points on curves or areas, one has to add to Newton's equation so-called forces of pressure to keep the points at the curve or area. This apparatus is rather clumsy. Following the ingenious idea of Lagrange it became much more elegant – by inserting a suitable system of coordinates – to eliminate all the constraints completely. Newton's equation of mechanics (second law: $a = F/m$, i.e. acceleration a of a body is parallel and directly proportional to the net force F and inversely proportional to the mass m) cannot be translated to more sophisticated physical theories like electrodynamics, universal relativity theory, theory of elementary particles etc. But the Lagrange approach can be generalized to all field theories in physics. The corresponding variational description is *Hamilton's principle of stationarity*, named after WILLIAM ROWAN HAMILTON (1805-1865). It proves to be an extremal principle and describes a generalization of different physical observations. In 1746 PIERRE LOUIS MAUPERTUIS was the first who discussed a universal valid principle of nature *behaving extremal or optimal*. For instance, a rolling ball is locally always following the steepest descent; the difference of the temperature in a body is creating a thermal stream in the direction of the lowest temperature or a ray of light shining through different media is always taking the path with the shortest time. (*Fermat's principle*).

EULER AND LAGRANGE contributed essentially to the mathematical formulation of these thoughts. CARL GUSTAV JAKOB JACOBI (1804-1851) wrote in this respect: "While Lagrange was going to generalize Euler's method of variational analysis, he observed how one can describe in one line the basic equation for all problems of analytical mechanics."[49].

Lagrange principle:

$$\min \{f(x) : g(x) = 0\} \quad \rightsquigarrow \quad L(x,\lambda) := f(x) + \lambda g(x) \to \min_{(x,\lambda)\in\mathbb{R}^{n+1}}$$





By crossing of the level lines $f(x) = const$ the value of the objective function f is changing; it becomes (locally) extremal if curve K touches at x^* tangentially such a level line, i.e., the tangents of both curves coincide at x^*, hence their normal vectors ∇f and ∇g are co-linear at x^*:

Euler-Lagrange formalism: x^* is solution $\Rightarrow$ $\exists\, \lambda^*$ such that

$$\begin{array}{rcl} L_x(x^*,\lambda^*) = 0 & \Leftrightarrow & \nabla f(x^*) + \lambda^* \nabla g(x^*) = 0 \\ L_\lambda(x^*,\lambda^*) = 0 & \Leftrightarrow & g(x^*) = 0 \end{array}$$

The description of many physical problems has been simplified by LAGRANGE'S formalism. Today it is a classical tool in optimization and finds its application wherever extrema subject to equality constraints have to be calculated.

Probably, this is the right place to mention that the Euler-Lagrange-equations are *necessary conditions* for a curve or a point to be optimal. However, in using these conditions, historically many errors were made which gave rise to mistakes for decades.
It is as in PERRON'S paradoxon:

Let N be the largest positive integer. Then for $N \neq 1$ it holds $N^2 > N$, contradicting that N is the largest integer.
Conclusion: $N = 1$ is the largest integer.

Implications as above are devastating, nonetheless they were made often. For instance, in elementary algebra in old Greece, where problems were solved beginning with the phrase: "Let x be the sought quantity".

In variational analysis the Euler equation belongs to the so-called necessary conditions. It has been obtained by the same pattern of argumentation as in Perron's paradoxon. The basic assumption that there exists a solution is used for calculating a solution whose existence is only postulated. However, in the class of problems, where this basic assumption holds true, there is no wrongdoing. But, from where do we know that a concrete problem belongs exactly to this class? The so-called necessary condition does not answer this question. Therefore, a "solution", obtained by these necessary Euler conditions, is still not a solution, but only a candidate for being a solution.
It is surprising that such an elementary point of logic went unnoticed for a long time. The first who criticized the Euler-Lagrange method was KARL WEIERSTRASS (1815-1897) almost one century later. Even GEORG FRIEDRICH BERNHARD RIEMANN (1826-1866) made the same unjust assumption in his famous *Dirichlet principle* (cf. [39]).

While at that time the resolution of several types of *equations* was a central topic in mathematics, one was mainly interested in finding unique solutions. Solving of *inequalities* arose a marginal interest only. Especially solving of inequalities by algorithmic methods wasn't playing almost any role.
FOURIER [38] was one of the first who described a systematic elimination method for solving linear inequalities, similar – but in its realization much more complicated – to Gauss elimination, which was already known by the Chinese people 300 years earlier, of course without CARL FRIEDRICH GAUSS's (1777-1855) knowing.

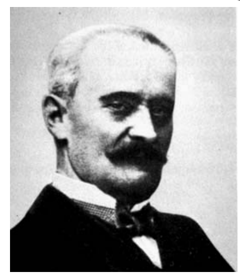
Jean Baptiste Joseph de Fourier (1768-1830)

1797: Professor for analysis and mechanics at the École Polytechnique Paris, successor of Lagrange.
1832: *Théorie analytique de la chaleur*.
(Analytic theory of the heat).

First systematic foundation of (Fourier) series and (Fourier) integrals for solving differential equations.
A memorial plaque of him can be found at the Eiffel tower in Paris.

He was a very practical-minded man. In 1802 Napoleon appointed him prefect of the

department Isère in the south of France. In this position he had to drain the marshes near Lyon. In 1815 Napoleon (after his return from island Elba) installed him as prefect of the department Rhône. He was working lifelong as secretary of the French Academy of Science.
Among the few who worked with inequality systems was FARKAS, born near Klausenburg (nowadays Cluj-Napoca, Romania). He investigated linear inequalities in mechanics and studied theorems of the alternative [34].
Probably 40 years later these results proved to be very helpful in the geometry of polyhedra and in the duality theory of linear programming.



Julius Farkas (1847-1930)

1887: Professor in Kolozsvár (Romania)
1902: *Grundsatz der einfachen Unglei- chungen*, J. f. Reine und Angew. Math. 124, 1-27.

Theorem: Given $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$.

$$\{x \in \mathbb{R}^n : Ax \leq b, x \geq 0\} \neq \emptyset \quad \Leftrightarrow \quad \{u \in \mathbb{R}^m : u \geq 0,\ \ A^T u \geq 0,\ \ u^T b < 0\} = \emptyset,$$

i.e., of these two linear inequality systems always exactly one is solvable.

In connection with linear inequality systems also MINKOWSKI has to be named, who used linear inequalities for his remarkable geometry of numbers and developed together with HERMANN WEYL (1885-1955) the structural assembling of polyhedra [81].

Herman Minkowski (1864-1909)

1892: Assistance professor at the University Bonn,

1894: Professor at the University Königsberg and since 1896 at the Polytechnikum Zürich, where Albert Einstein was one of his students.

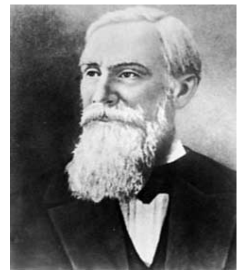
- Geometry of numbers,
- Geometrization of the special relativity theory,
- Theory of convex bodies.

Theorem: *Let P be a polyhedral set, L_P its lineality space and $P^0 = P \cap L_P^{\perp}$.*
Denote $S = \{x^1, ..., x^q\}$ the extremal points and $T = \{y^1, ..., y^r\}$ the extremal rays of P^0.
Then

$$P = L_P + conv(S) + cone(T).$$

To the roots of the theory of optimization belong also the works of CHEBYSHEV, better known from his contributions to approximation theory.

Pafnuti Lvovich Chebyshev (1821-1894)



1850: Associate professor at St. Petersburg's University,

1860: Professor, there.

Basic contributions to probability theory, theory of numbers and approximation theory.

Chebyshev-problem:

$$\min_x \max_{t\in T} |a(t) - \sum_i x_i f_i(t)|.$$

In the simplest version of such a continuous approximation problem one is looking for the uniform approximation of a given continuous curve $a(t)$ by a system of linearly independent functions $f_i(t)$. In today's terminology one would say we are dealing with a non-smooth convex minimization problem, ore more exactly with a *semi-infinite problem*. Hence, CHEBYSHEV can be regarded as one of the first who considered this kind of optimization problems. For some special cases he found analytic solutions, known as Chebyshev polynomials.
Similar to EULER he also understood the significance of extremal problems. He wrote [115]: "In all practical human activities we find the same problem: How to allocate our ressources such that as most as possible profit can be attained?"

In Russia two students of CHEBYSHEV, namely MARKOV and LYAPUNOV, carried on with the investigations of extremal problems.
MARKOV is mainly known for theory of stochastic processes.

Andrey Andreyevich Markov (1856-1922)

1886: Assistance professor at St. Petersburg's University,

member of the Russian Academy of Science.

Famous for his works on number theory and probability theory (Markov chains, Markov processes etc).

Problem of moments:

$$\begin{aligned} \min \quad & \int_a^b t^n f(t)dt, \\ s.t. \quad & 0 \le f(t) \le t, \quad \forall\, t \in [a,b], \\ & \int_a^b t^i f(t)dt = c_i, \quad i = 1,...,n-1. \end{aligned}$$

In 1913 he studied sequences of letters in novels to detect the necessity of independence of the law of large numbers. According to that law, the average of the results obtained

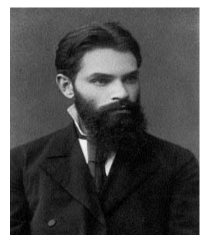
from a large number of trials should be close to the expected value, and would tend to become closer as more trials are performed.
The so-called *stochastic Markov process* became a general statistical tool, from which future developments can be determined by current knowledge. But MARKOV studied also so-called *moment problems* for optimizing the moments of a distribution function or stochastic variables [1], [67]. This kind of problems can be formulated as constrained optimization problems with integral functions, where, in distinction to a variational problem, no derivatives appear.

At the first glance LYAPUNOV's investigations are not connected with optimization, because he studied stability theory for differential equations [99].

Aleksandr Mikhailovich Lyapunov (1857-1918)

1895: Associate professor at the University of Kharkov,

founder of stability theory for differential equations.



Theorem: *Solution $x(t)$ of the equation $\dot{x} = f(x)$ is stable if there exists a function $V(x)$ such that*

$$\langle \nabla V(x), f(x) \rangle < 0.$$

We can take an inverse point of view and interpret the result as follows: The differential equation in Lyapunov's theorem is a time-continuous method for minimizing the (Lyapunov-) function $V(x)$. Today the *Lypunov method* is a systematical tool for investigating convergence and stability of numerical methods in optimization.

2 The Pioneers in Linear Optimization

There exist two isolated roots of linear optimization, which can be traced back to GASPARD MONGE [82] and CHARLES-JEAN DE LA VALLÉE POUSSIN [95].

Gaspard Monge (1746-1818)

1765: Professor for mathematics and 1771 for physics in Mézières,
1780: Professor for hydrodynamics in Paris,
1794: Founder of the École Polytechnique in Paris.

1782: *Continuous mass transport under minimal costs*, Application de l'analyse à la géometrie, Paris.

In 1780 MONGE became member of the French Academy of Science. In the days when 1789 the French Revolution began, he was a supporter of it and at the moment of proclamation of the French Republic in 1792 he was appointed Minister of navy. In this position he was jointly responsible for the death sentence of King Ludwig XVI. Among several physical discoveries, for instance theory of mirage, he rendered outstanding services to the creation of the *descriptive geometry*, to which also belongs his work on continuous mass transport. His idea is seen as an early contribution to the linear transport problem, a particular case of the linear programming problem.

The second root is attributed to VALLÉE POUSSIN.

Charles-Jean de La Vallée Poussin (1866-1962)

1892: Professor for mathematics an der Université Louvain

1911: *Sure la méthode de l'approximation minimum*, Anales de la Societé Scientifique de Bruxelles, No 35, pp. 1-16.

1920: First president of the International Mathematical Union.

In the years 1892 - 1894 he attended lectures of CAMILLE JORDAN, HENRI POINCARÉ, ÉMILE PICARD in Paris and of AMANDUS SCHWARZ, FERDINAND FROBENIUS in Berlin. With his paper, published in the Anales of Brussels Society of Science, he is rated as one of the founders of linear optimization.

By the way, concerning the contributions of MONGE and VALLÉE POUSSIN, in 1991 DANTZIG wrote disparagingly [75] (page 19): "Their works had as much influence on the development of Linear Programming in the forties, as one would find in an Egyptian pyramid an electronic computer built in 3000 BC".

In the forties of the last century, as a matter of fact, optimization – as we understand this topic today – was developed seriously and again practical problems influenced the directions of its outcome. Doubtless, time was ripe for establishing such rapid development.

In the community of the optimizers are to name three forceful pioneers: L.V. KANTOROVICH, T.C. KOOPMANS and G.B. DANTZIG.

Koopmans, Dantzig, Kantorovich (Laxenburg, 1976)

In 1939, for the first time, KANTOROVICH solved a problem of linear optimization. Shortly afterwards, F.L. HITCHCOCK published a paper about a transportation problem. However at that time the importance of these papers was not recognized entirely.

1926 - 1930 KANTOROVICH studied mathematics at the University of Leningrad. At the age of 18 he obtained a doctorate in mathematics. However, the doctor degree was awarded to him only in 1935, at that time the academic titles had been re-introduced in the Soviet society [73]. In the forties a rapid development of the functional analysis was set up. Here we have to mention the names of HILBERT, BANACH, STEINHAUS and MAZUR, but also KANTOROVICH.

Leonid Vitalevich Kantorovich (1912-1986)

1934: Professor at the University of Leningrad

- Linear Optimization (1939),
- Optimality conditions for extremal problems in topological vector spaces (1940),
- Functional analytic foundation of descent methods,
  Convergence of Newton's method for functional equations (1939-1948).

1939: *Mathematical Methods for Production Organization and Planning*, Leningrad, 66 pages.
1940: *On an efficient method for solving some classes of extremum problems*, DAN SSSR 28.
1959: *Functional Analysis*, Moscow, Nauka.

Before he attained his majority of twenty-one years, he had published fifteen papers in major mathematical journals and became a full professor at Leningrad University. He was a mathematician in the classical mold whose contributions were mainly centered

on functional analysis, descriptive and constructive function theory and set theory as well as on computational and approximate methods and mathematical programming. So he made significant contributions to the building of bridges between functional analysis, optimization and numerical methods.
At the end of the thirties he was concerned with the mathematical modeling of the production in some timber company and developed a method, which later on was recognized as equivalent to the *dual simplex method*. In 1939 he published a small paperback (only 66 pages) [52], with the exact title (in English translation): "A mathematical method of the production planning and organization and the best use of economic operating funds". Neither the notions *Linejnaja Optimizacija (Linear Optimization)* nor *simplex method* were ever mentioned in this booklet.
In contrast to the publicity of DANTZIG's results in the western countries, KANTOROVICH's booklet received only a small echo within mathematicians and economists in the East. The western world, caused by the iron curtain, didn't have any knowledge of that publication and in the Soviet Union there were probably two reasons for ignoring it. First, there was no real need for mathematical methods in a totalitarian system. Although the central planning of the national economy stood theoretically in the foreground of all social processes, the system was founded essentially on administration. Second, it should be mentioned that this booklet was not written in the usual mathematical language, therefore mathematicians had no reason to read it.
What is really known is his book on *Economical Calculation of the Best Utilization of Resources* [56], published in 1960 (with an appendix by G.S. RUBINSTEIN), but at that time in the West the essential developments were almost finished. In this monograph one can find two appendices about the mathematical theory of Linear Programming and their numerical methods. Curiously, in doing justice to the Marxist terminology, therein the *dual variables* are denoted by *objectively substantiated estimates* but not as *prices*, because in the Soviet thinking prices had not to be imposed by the market but by the Politburo.
As already mentioned, KANTOROVICH contributed significantly to functional analysis [54], [55]. His functional-analytic methods in optimization are well-known and contain ideas and techniques, which have been in the progress of development thirty years before the preparation of the *theory of convex analysis*.
As already mentioned, in 1975 he was honored, together with KOOPMANS, with the Nobel price for economics. Quotation of the Nobel committee: "For contributions to the theory of the optimal allocation of operating capital".

KOOPMANS was an US-American economist and physicist with roots in the Netherlands, who tackled the problems of resource allocations.
Koopmans was highly annoyed that DANTZIG could not participate in that price.

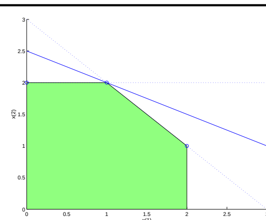
Tjalling Charles Koopmans (1910-1985)

1948: Professor at the Yale University,

1968: Professor at the Stanford University.

1942: *Exchange Ratios between Cargoes on Various Routes (Non-Refrigerated Dry Cargoes)*, Memorandum for the Combined Shipping Adjustment Board, Washington, D.C.
1951: *Activity Analysis of Production and Allocation*, Wiley, New York.
1971: *On the Description and Comparison of Economic Systems*, (with J. Michael Montias) in: Comparison of Economic Systems, Univ. of California Press, 1971, pp. 27-78.

In the mid-forties DANTZIG became aware of the fact that in many practical modeling problems the economic restrictions could be described by linear inequalities. Moreover, replacing the "rule of thumb" by a goal function, for the first time he formulated deliberately a problem, consisting explicitly of a (linear) objective function and (linear) restrictions in form of equalities and/or inequalities. In particular, hereby he established a clear separation between the goal of the optimization, the set of feasible solutions and, by suggesting the *simplex method*, the method of solving such problems.



DANTZIG studied mathematics at the universities of Maryland and Michigan, because his parents could not afford a study at a more distinguished university. In 1936 he got his B.A. in mathematics and physics and switched to the University of Michigan in order to earn a doctorate. In 1937 he finished advanced studies, receiving a M.A. in mathematics.
After working two years as a statistician at the *RAND Corporation* in Washington, in 1939 he started his PhD-study at the *University of California, Berkeley*, which he interrupted when the USA was joining the Second World War. He entered the *Air Force* and became (1941 – 1946) leader of the *Combat Analysis Branch* at the headquarter of the US-Air Force. In 1946 he continued with his PhD-studies and obtained a doctorate under the supervision of JERZY NEYMAN. Thereafter he was working as a mathematical consultant at the Ministry of Defense.

Georg B. Dantzig (1914-2005)

1960: Professor at the University of California, Berkeley,

1966: Professor at the Stanford University.

1966: *Linear Programming and Extensions*, Springer-Verlag, Berlin.
1966: *Linear Inequalities and Related Systems*, Princeton Univ. Press, Princeton, NJ.
1969: *Lectures in Differential Equations*, Van Nostrand, Reinhold Co., New York.

The breakthrough in Linear Programming was made in 1947 with DANTZIG's paper: *Programming in a Linear Structure*. One can read by DANTZIG [75] page 29, that during the summer of 1948 KOOPMANS suggested him to make use of a shorter title, namely *Linear Programming*. The notion *simplex method* goes back to a discussion between DANTZIG and MOTZKIN, the latter held the opinion that simplex method describes most excellently the geometry of changing from one vertex of the polyhedra of the feasible solutions to another.

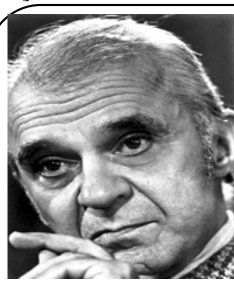
In 1949, exactly two years after the first publication of the simplex algorithm, KOOPMANS organized the first *Conference on Mathematical Programming* in Chicago, which was later counted as number "zero-conference" in a sequence of *Mathematical Programming Conferences*, taking place up today. Besides, well-known economists like ARROW, SAMUELSON, HURWICZ AND DORFMAN, also mathematicians like ALBERT TUCKER, HAROLD KUHN, DAVID GALE, JOHANN VON NEUMANN, THEODORE S. MOTZKIN and others attended this event.

In 1960 DANTZIG became professor at the *University of California at Berkeley* and in 1966 he switched to a chair for Operations Research and Computer Science at *Stanford-University*. In 1973 he was one of the founders and the first president of the *Mathematical Programming Society* (MPS). Also by MPS the Dantzig price has been created and awarded to colleagues for outstanding contributions in mathematical programming. In 1991 the first edition of the *SIAM Journal on Optimization* was dedicated to George B. Dantzig.

Back to the Nobel price awarded to KOOPMANS and KANTOROVICH. In 1975 the Royal Swedish Academy of Science granted the price for economy to equal parts to KANTOROVICH and to KOOPMANS for their contributions to optimal resource allocation. The prize money at that year amounted to 240.000 US-dollars. Immediately after this ceremony KOOPMANS traveled to *IIASA* (The International Institute for Applied Systems Analysis) in Laxenburg, Austria. One can read in MICHEL BALINSKI [75], page 12, at that time director of the IIASA, that on the occasion of a ceremonial meeting KOOPMANS submitted 40.000 $ of the prize money as a present to the IIASA. Therefore, he indeed accepted only one third of the whole amount of the prize money for himself.

By the way, for a long time it was unclear whether it would be permitted that KANTOROVICH could accept the Nobel price of economy, because some years before when BORIS PASTERNAK (known for his novel ”Doctor Shiwago”) had the honor to get the Nobel price for literature, the Soviet authorities forced him to reject. Also the Nobel price award to the physicist ANDREJ SACHAROV, one of the leading Soviet dissidents at that time, had been seen as an unfriendly act. But because one was unable to change SACHAROV’s mind to accept this recognition, one refused his journey to Stockholm and expelled him as a member from the Soviet Academy of Sciences. It is known that KANTOROVICH, together with some physicists, voted against this expulsion.

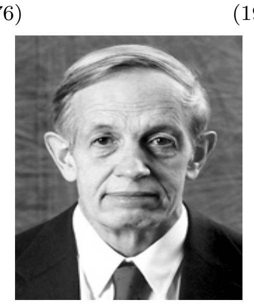
It is worth mentioning that in the framework of ”Linear Programming” and ”Operations Research” in its widest sense five more scientists have been awarded with the Nobel price: in 1976 WASILIJ LEONTIEV (Input-Output-Analysis), in 1990 HARRY MARKOWITZ (Development of the theory of portfolio selection), in 1994 REINHARD SELTEN and JOHN NASH (Analysis of the equilibrium in non-cooperative game theory) and in 2007 LEONID HURWICZ (Development of the basics of economic design). HURWICZ turned at the time of his awarding 90 years and has been the oldest price winner up to now.
An historical overview about the development of Operations Research can be found in [40].



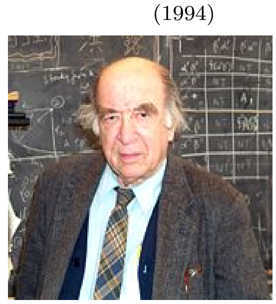
Wasilij Leontiev
(1976)

Harry Markowitz
(1990)

Reinhard Selten
(1994)



John F. Nash
(1994)



Leonid Hurwicz
(2007)

Now, let us sidestep to game theory and his founder JOHANN VON NEUMANN . He achieved outstanding results on several mathematical fields.

Johann von Neumann (1903-1957)

1926-1929: Associate professor at the Humboldt University, Berlin,
1933-1957: Professor at the Institute for Advanced Studies, Princeton,
1943: Collaborator at the Manhattan project in Los Alamos.

- Quantum mechanics,
- Theory of linear operators in Hilbert spaces,
- Theory of shock waves,
- Computer architecture.

1928: *Zur Theorie der Gesellschaftsspiele*, Math. Ann. 100, 295-320.
1944: *The Theory of Games and Economic Behavior*, Springer.

Already in 1928 a paper of the mathematician ÉMILE BOREL (1871-1956) on min-max properties inspired him to ideas, which led to one of the most original design later on, the game theory. In the same year he proved the min-max theorem [85] on the existence of optimal strategies in a *zero-sum game*. Together with the economist OSKAR MORGENSTERN he wrote in 1944 the famous book "The Theory of Games and Economic Behavior" [86], dealing also with n-person games ($n > 2$), which are important generalizations in economy. These contributions made him the founder of game theory, which he applied less to classical salon games, rather than to situations of conflict and decision with incomplete knowledge of the intensions of the opposing players.

When on December 12, 1941, in Pearl Harbor the Japanese Air Force scuttled most of the American Pacific Fleet, it was the time of birth of the application of game theory for military purposes. Later, from the analysis of the debacle it became clear that the recommendations, given by experts and founded on game theoretical considerations, were dismissed by the Pentagon. This gave game theory an extraordinary impetus and up to now the mathematical research on game theory is subject of secrecy in a great extent.

3 The Beginnings of Nonlinear Optimization

In the fifties, apart from Linear Programming, several new research directions in the area of extremal problems have been developed, which are summarized today under the keyword *Mathematical Programming*.

Concerning optimization problems subject to equality constraints we mentioned already that the optimality conditions are going back to EULER and LAGRANGE. Nonlinear inequality constraints have been considered first in 1914 by BOLZA [11] and in 1939 by KARUSH [59]. Unfortunately, these results have been forgotten for a long time.

O. Bolza:
1914: Über Variationsprobleme mit Ungleichungen als Nebenbedingungen,
Mathem. Abhandlungen 1-18.
W. Karush:
1939: Minima of functions of several variables with inequalities as side conditions,
MSc Thesis, Univ. of Chicago.
F. John:
1948: Extremum problems with inequalities as subsidiary conditions,
Studies and Essays, Presented to R. Courant on his 60th Birthday, Jan. 1948,
Interscience, New York, 187-204.
M. Slater:
1950: *Lagrange Multipliers Revisited*, Cowles Commisssion Discussion Paper, No 403.

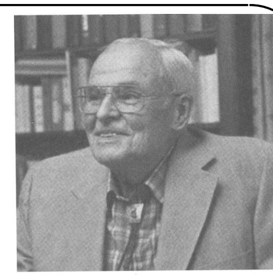
In 1948 FRITZ JOHN [50] considered problems with inequality constraints, too. He did not assume any *constraint qualifications*, up to the fact that all functions should be continuously differentiable.
The term *constraint qualification* can be traced back to KUHN AND TUCKER [69] and says (and this makes the treatment of problems under nonlinear inequalities so difficult) that a suitable local approximation of the feasible domain is guaranteed (mathematically speaking: linearization cone and tangential cone have to coincide).
A discussion here about the historical development of the *constraint qualification* would go to far. But we refer to an earlier paper of MORTON SLATER [103]. He found a useful sufficient condition for the existence of a saddle point, without assuming that the saddle function is differentiable.

In developing the theory of nonlinear optimization ALBERT W. TUCKER played an outstanding role.

Albert W. Tucker (1905-1995)

1933: Professor at the Princeton University.

- Topology,
- Mathematical programming (duality theory),
- Game theory (prisoner's dilemma).



Albert W. Tucker
1905–1995

Afternoon tea club:
J. Alexander, A. Church, A. Einstein, L. Eisenhart, S. Lefschetz, J.v. Neumann,
O. Veblen, H. Weyl, E. Wigner, A. Turing, u.a.

TUCKER graduated in 1932 and since 1933 he was a fellow of the Mathematical Department at the Princeton University. Actually in the fifties and sixties he became a successful chairman of the department. He is known for his work in *duality theory* for linear and nonlinear optimization, but also in game theory. He introduced the well-known *prisoner's dilemma* which is a bi-matrix game with non-constant profit sum.

His famous students were MICHEL BALINSKI, DAVID GALE, JOHN NASH (Nobel price 1994), LLOYD SHAPLEY (Nobel price 2012) and ALAN GOLDMAN. Every year the *Mathematical Programming Society* grants the Tucker price for outstanding student achievements.

In the thirties the department in Princeton was famous for its tea afternoon sessions, bringing together scientists and giving reason for inspiring discussions. Members of this club were ALBERT EINSTEIN, JOHANN V. NEUMANN, HERMANN WEYL and at the beginning also ALAN TURING, a student of the logician ALONZO CHURCH. During the Second Wold War TURING developed the ideas of Polish mathematicians and was able to crack a highly complicated German radio code *ENIGMA*[2] and later on, working at the University of Manchester, he invented the *Turing machine*.

Starting in 1948 until about 1972 at Princeton, under the leadership of TUCKER, a project was sponsored by the *Naval Research Office*, drawing up optimality conditions for different classes of non-linear optimization problems and formulating the duality theory for convex problems. In this project among others HAROLD KUHN, a student of Ralph Fox, DAVID GALE AND LLOYD SHAPLEY were involved.

Due to ALBERT TUCKER and HAROLD KUHN Lagrange's multiplier rule has been generalized to problems with inequality constraints [68].

Harold W. Kuhn and Albert W. Tucker:
1951: Nonlinear programming, *Proceedings of the Second Berkeley Symposium on Mathem. Statistics and Probability*,
Univ. of California Press, Berkeley, 481-492.

$$(P) \qquad \begin{cases} f(x) \to \min,\ x \in \mathbb{R}^n \\ g_i(x) \le 0,\ \ (i = 1, \cdots, m) \end{cases}$$

$$(S) \qquad \exists\, \tilde{x} \text{ with } g_i(\tilde{x}) < 0\, \forall\, i = 1, \cdots, m$$

Theorem: (necessary and sufficient optimality conditions)
In (P) let $f,\ g_i,\ (i = 1, \cdots, m)$ be convex functions and Slater's condition (S) be satisfied. Then:

$$x^* \text{ is a global minimizer of } (P) \quad \Leftrightarrow \quad \exists \quad \begin{array}{l} \lambda_i^* \ge 0\ (i = 1, \cdots, m), \text{ such that} \\ \lambda_i^* g_i(x^*) = 0\ (i = 1, \cdots, m), \\ L(x, \lambda^*) \ge L(x^*, \lambda^*)\ \forall\, x \in \mathbb{R}^n, \end{array}$$

where $\lambda^* \in \mathbb{R}^m_+$ - (Lagrange multiplier to x^*);
$g(x) = [g_1(x), \cdots, g_m(x)]^T$;
$L(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle$ - (Lagrange function of (P)).

1957: Linear and nonlinear programming, *Oper. Res. 5*, 244-257.

What about Linear Programming at that time? Particular attention was payed to its

[2] Marian Rejewski together with two fellow Poznan' University mathematics graduates, Henryk Zygalski and Jerzy Różycki, solved 1932 the logical structure of one of the first military versions of Enigma.

commercial applications, although no efficient computers were at hand. One of the first documented applications of the simplex method was a diet problem by G.J. STIEGLER [105], with the goal of modeling a possibly economical food composition for the US-Army, guaranteeing certain minimum and maximum quantities of vitamins and other ingredients. The solution of this linear program with 9 inequalities and 77 variables kept busy 9 persons, and required computational work of approximately 120 man days.

Historic overview on LP-computations by means of simplex algorithms on computers: (cf. [88])

Year	Constraints	Remarks
1951	10	
1954	30	
1957	120	
1960	600	
1963	2 500	
1966	10 000	structured
1970	30 000	structured

Nowadays structured problems with Millions of constraints are solved, for instance in aviation industries.
In this context one can read the following curiosity in LILLY LANCASTER [72]: It is well known that the simplex algorithm carries out the pivoting in dependence of the reduced costs. However, the prices for spices, as a rule, are higher as those for other commodities. Therefore, the spice-variables appeared mostly as nonbasic variables, hence they got the values zero. The result was that the optimized food was terrible tasteless. In this paper it is described how the LP-model was changed stepwise by altering the constraints in order to get tasty food.

In 1952 CHARNES, COOPER AND MELLON [14] successfully used the simplex method in the oil industry for optimal cracking of crude oil into petrol and high quality oil.

The first publication on solving linearly constrained systems iteratively can be traced back to HESTENES AND STIEFEL [46]. In 1952 they suggested a *conjugate gradient method* to determine a feasible point of a system of linear equations and inequalities.

In the fifties in the US the development of network flow problems has been started. The contribution of FORD AND FULKERSON [37] consisted in connecting flow problems with graph theory. Until now *combinatorial optimization* is benefiting from this approach.

1959-60 DANTZIG AND WOLFE [23] were working on *decomposition principles*, i.e., the decomposition of structured large scale LP's into master and (several) subproblems. Nowadays these ideas allow parallelization of computational work on the level of the subproblems and enable a fast resolution of such problems with hundreds of thousands of variables and constraints.
The dual variant of this decomposition method was used in 1962 by BENDERS [5] for solving *mixed integer problems*.

M. R. Hestenes; E. Stiefel:
1952: *Methods of conjugate gradients for solving linear systems*,
J. Res. Natl. Bur. Stand. 49, 409-436.

R. Gomory:
1958: Outline of an algorithm for integer solutions to linear programs,
Bull. Amer. Math. Soc. 64, 275-278.

G.B. Dantzig; Ph. Wolfe:
1960: *Decomposition principle for linear programs*, Oper. Res. 8, 101-111.

J.F. Benders:
1962: *Partitioning procedures for solving mixed-variables programming problems*,
Numer. Math. 4, 238-252.

L.R. Ford; D.R. Fulkerson:
1962: *Flows in Networks*, Princeton University Press, Princeton, N. J., 194 p.

The study of *integer programming problems* has been started in 1958 by RALPH GOMORY [44]. Unlike the earlier work on the traveling salesman problem by FULKERSON, JOHNSON AND DANTZIG on the usage of cutting planes for cutting off non-optimal tours in the *traveling salesman problem*, GOMORY showed how to generate "cutting planes" systematically. These are extra conditions which, when added to an existing system of inequalities, guarantee that the optimal solution consists of integers. Today such techniques, combining cutting planes with *branch-and-bound-methods*, belong to the most efficient algorithms for solving applied problems in integer programming.

In the Soviet Union first results on matrix games have been published by VENTZEL [111] and VOROBYEV [112] and a very popular Russian textbook on Linear Programming has been written by YUDIN AND GOLSTEIN [113].

S.I. Zukhovitzkij:
1956: *On the approximation of real functions in the sense of Chebyshev* (in Russian),
Uspekhi Matem. Nauk, 11(2), 125-159.

E.S. Ventzel; N.N. Vorobyev:
1959: *Elements of Game Theory* (in Russian), Moscow, Fizmatgiz.

D. B. Yudin; E.G. Golstein:
1961: *Problems and Methods of Linear Programming* (in Russian), Moscow, Sov. Radio.

E. Ya. Remez:
1969: *Foundation of Numerical Methods for Chebyshev Approximation* (in Russian), Kiev, Naukova Dumka.

Approximately at the same time, papers of two Ukrainian mathematicians, ZUKHOVITSKIJ [116] and REMEZ [96], became known. They suggested numerical methods for solving best-approximation problems in the sense of Chebyshev. These are simplex-like algorithms for solving the underlying linear *semi-infinite optimization problem*.

The initial works about this topic are essentially older than mentioned here. For instance, REMEZ's algorithm for the numerical solution of Chebyshev's approximation problem was presented by REMEZ in 1935 on the occasion of a meeting of the Mathematical Department of the Ukrainian Academy of Sciences.

Important results in the field of *control theory*, i.e. optimization under constraints described by differential equations, were initiated with the beginnings of space travel. 1957 was the year when the Soviets were shooting the first rocket, the *Sputnik*, in the outer space.
One of the basic problems in *optimal control* consists in the transfer of the state $x(t)$ of some system, described by differential equations, from a given start into some target domain $\mathcal{T}$. Hereby the controlling is carried out by some control function $u(t)$ which belongs to a certain class of functions and minimizes a given objective functional.
A typical problem of that kind can be found in space travel: Transfer of a controllable object to some planet in shortest time or with smallest costs.

Terminal control problem:

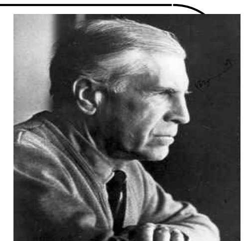
$$\begin{aligned} \min \quad & \int_0^T f^0(x(t), u(t))dt; \\ \frac{dx}{dt} \quad = \quad & f(x,u), \quad x(0) = x_0, \quad x(T) \in \mathcal{T}(T); \\ u(t) \quad \in \quad & U_{ad} = \{u(\cdot) : \text{ measurable }, u(t) \in \Psi \text{ for } t_0 \leq t \leq t_1\}. \end{aligned}$$

(x - state vector, u - control vector, $f = (f^1, ..., f^n)$).

Hence, again the question about necessary optimality conditions arises, which have to be satisfied by an optimal control function, i.e. we are dealing with a strong analogue to variational analysis. In the latter theory the Euler-Lagrange-equations were necessary showing that certain functions prove to be candidates for optimal functions; here in control theory the analogous necessary optimality conditions for a control problem are described by *Pontryagin's maximum principle*.

PONTRYAGIN lost his eyesight at the age of 14 years. Thanks to his mother, who read mathematical books to him, he became a mathematician despite of his blindness.
We have to thank him for a series of basic results, first of all in topology. His excursus into applied mathematics by investigating control problems has to be valued merely as a side effect of his research program.

In 1954 the *maximum principle* was formulated as a thesis by PONTRYAGIN and proved in 1956 in a joint monograph with BOLTYANSKIJ AND GAMKRELIDZE [91]. Still it proves to be fundamental for modern control theory.



Lev Semenovich Pontryagin (1908 - 1988)

1934: Steklov Institute, Moscow,

1935: Head of the Department of Topology and Functional Analysis at the Steklov Institute.

- Geometric aspects of the topology,
- Duality theory of the homology,
- Co-cycle theory in topology (Pontryagin classes),
- Maximum principle.

1956: L.S. Pontryagin, V.G. Boltyanskij, R. V. Gamkrelidze: *Mathematical Theory of Optimal Processes*, Moscow, Nauka.

GAMKRELIDZE, one of the coauthors of the mentioned monograph, indicated later on PONTRYAGIN's proof of the maximum principle as "in some sense sensational". Exceptional on this proof is the usage of topological arguments, namely the *cut theory*, which goes back to the American S. LEFSCHETZ.

BOLTYANSKIJ [9], in his extensive widening of the maximum principle to other classes of control problems, has shown that topological methods, in particular homotopy results, are very useful in control theory. Meanwhile there exist proof techniques belonging to convex analysis and establishing the maximum principle, too. One of the first authors in this area was PSHENICHNYJ [93].

Maximum principle of a terminal control problem:

Consider dual state variables (Lagrange multipliers) $\lambda_0(t), ..., \lambda_n(t)$, the *Hamiltonian function*

$$H(\lambda, x, u) = \sum_{i=0}^{n} \lambda_i(t) f^i(x(t), u(t))$$

and the adjoint system (for some feasible process $(x(t), u(t)), 0 \le t \le T$)

$$\frac{d\lambda_i}{dt} = -\frac{\partial H}{\partial x^j} = -\sum_{i=0}^{n} \lambda_i \frac{\partial f^i(x,u)}{\partial x^j}, \quad j = 1, ..., n,$$

$$\lambda_0 = -1, \quad \lambda_1(T) = \lambda_2(T) = ... = \lambda_n(T) = 0.$$

Theorem: *If the process $(x^*(t), u^*(t))$ is optimal, then there exist absolutely continuous functions $\lambda_i^*(t)$, solving the adjoint system almost all on $[0, T]$ and at each time $\tau \in [0, T]$ the following maximum condition is satisfied:*

$$H(\lambda^*(\tau), x^*(\tau), u^*(\tau)) = \max_{u \in U_{ad}} H(\lambda^*(\tau), x^*(\tau), u).$$

The transfer of the maximum principle into a discrete setting made some difficulties. Among the numerous papers, dealing at that time with this questions, there are more than a few which are incorrect (see references in [10]).

Another important principle, which found its application mainly in the theory of discrete optimal processes, is the *principle of dynamic programming* and can be traced back to BELLMAN [6].

Richard Bellman (1920 - 1984)

1947: Collaborator in the area of theoretical physics in Los Alamos,
1952: Switch to the Rand Corporation, development of dynamical programming,
1965: Professor for mathematics, electro-technique and medicine at the University of Southern California.

- Bellmann's optimality principle,
- Algorithm of Bellman and Ford (shortest path in graphs),
- Bio-informatics.

1957: *Dynamic Programming*, Princeton Univ. Press.

In physics this principle was known for a long time, but under another name: *Legendre transformation*. There, the transition from a global (at all times simultaneously) to a time-dependent (dynamical) way of looking at things corresponds to the transition of the *Lagrange*-functional into the *Hamilton*-functional by means of the *Legendre transformation*.
In control theory and similar areas this approach can be used, for instance, to derive an equation (*Hamilton-Jacobi-Bellman equation*) where its solution amounts in the optimal objective value of the optimal process.
Hereby the argumentation is more or less as follows: If a problem is time-dependent, one can consider the optimal value of the objective functional at a certain time. Then one is asking, which equation has to be fulfilled at the optimal solution such that the objective functional is staying optimal also at a later date. This consideration leads to the *Hamilton-Jacobi-Bellman equation*. That way one can divide the problem into time-steps instead of solving it at the whole.

Bellmann's optimality principle for a discrete process:

The value of the objective functional at the k-th level is optimal if for each at the k-th level chosen x_k the objective value of the $(k-1)$-th level is optimal.
Denote

ξ	state vector, descibing the state of the k-level process,
$\Lambda_k(\xi)$	optimal value of the k-level process, in dependence of the state k,
ξ, x_k	variables (or vectors of variables) which have to be determined at k-th level.

Assumption: After choosing x_k and ξ let the vector of state-variables, corresponding to the $(k-1)$-th level, be given by some transformation $T(\xi, x_k)$.

$$\Lambda_k(\xi) = \max_{x_k}\{f_k(\xi, x_k) + \Lambda_{k-1}[T(\xi, x_k)],\ k = 1, 2, ..., n\}.$$

System of recurrent formulas, in which $\Lambda_k(\xi)$ can be determined for $(k = 1, \cdots, n)$ if $\Lambda_{k-1}(\eta)$ is known for the problem at the $(k-1)$-th level.

About five years later an intensive study of control problems has started, beginning with the papers of ROXIN [100], NEUSTADT [87], BALAKRISHNAN [3], HESTENES [47], HALKIN [45], BUTKOVSKIJ [12], BERKOVITZ [7] and others.

E. Roxin:
1962: *The existence of optimal controls*, Michigan Math. J. 9, 109-119.
L.W. Neustadt:
1963: *The existence of the optimal control in the absence of convexity conditions*, J. Math. Anal. Appl. 7, 110-117.
A. V. Balakrishnan:
1965: *Optimal control problem in Banach spaces*, J. Soc. Ind. Appl. Math., Ser. A, Control 3, 152-180.
M.R. Hestenes:
1966: *Calculus of Variations and Optimal Control Theory*, Wiley, New York.
H. Halkin:
1966: *A maximum principle of Pontryagin's type for nonlinear differential equations*, SIAM J. Control 4, 90-112.
A. G. Butkovskij:
1969: *Distributed Control Systems*, Isd. Nauka, Moscow.
L.D. Berkovitz:
1969: *An existence theorem for optimal control*, JOTA 4, 77-86.

The beginnings of *stochastic optimization* can be found in the literature from 1955 on. At that time the application of observed coincidences (in some parts) of data in LP's has been discussed, for instance, by DANTZIG [21] and G. TINTNER [110].
One was investigating several problems: Compensation problems (recourse), distributed problems (among others also distribution of the optimal value under given common distribution of LP-data) or problems with probability restrictions (chance-constraints).

Under special structural assumptions (with respect to the data and their given distribution) first applications were considered by VAN DE PANNE AND POPP [89] and TINTNER [110]. Also there can be found first solution techniques for stochastic problems by means of quadratic optimization, for instance, by BEALE [4]. DANTZIG AND MADANSKY [22] described techniques which use two-stage stochastic programs and in CHARNES AND COOPER [15] one can find programs with constraints, which have to be satisfied with certain probabilities.

The state of the art in stochastic programming till the mid-seventies has been described in the monographs by KALL [51] and ERMOLIEV [32].

G. B. Dantzig:
1955: *Linear programming under uncertainty*, Management Sci., 1: 197-206.
G. Tintner:
1955: *Stochastic linear programming with applications to agricultural economics*, in: 2nd Symp. Linear Programming, vol.2: 197-228.
A. Charnes and W.W. Cooper:
1959: *Chance-constrained programming*, Management Sci., 5:73-79.
E.M.L. Beale:
1961: *The use of quadratic programming in stochastic linear programming*, RAND Report P-2404, The RAND Corporation.
G.B. Dantzig and A. Madansky:
1961: *On the solution of two-stage linear programs under uncertainty*, in: Proc. 4th Berkeley Symp. Math. Stat. Prob., Berkeley, pp. 165-176.
C. van de Panne and W. Poop:
1963: *Minimum-cost cattle feed under probabilistic problem constraint*, Management Sci., 9:405-430.
P. Kall:
1976: *Stochastic Linear Programming*, Springer-Verlag, Berlin.
Y.M. Ermoliev:
1976: *Stochastic Programming Methods*, Nauka, Moscow.

As we can see there was a time of great activities, but the results in essence were still isolated and could not be understood as a part of an uniquely united branch. The situation changed dramatically in the sixties and seventies. Time was ripe to create a complete picture of *Mathematical Programming*, which immediately led to a kaleidoscope of new contributions.

4 The 60s and 70s

The main directions of the investigation in these years were: General theory of nonlinear optimization, numerical methods for nonlinear optimization problems, non-smooth optimization, global optimization, discrete optimization, optimization on graphs, stochastic optimization, dynamic optimization, and variational inequalities.
The understanding of the common nature of different optimization problems was the first breakthrough in this period. Although in different papers there existed different approaches for analyzing specific nonlinear problems, still these did not lead to a common technique for obtaining optimality criteria. Moreover, the mentioned papers were dealing exclusively with the finite dimensional case, with the exception of the paper of BOLZA [11].

The transition to infinite-dimensional settings was forced essentially by the papers of

DUBOVITZKIJ AND MILYUTIN [31] and PSHENICNYIJ [94]. At the latest at that time it became clear that the functional analytical foundation of duality theory in mathematical programming in general spaces can be deduced from the geometric form of the HAHN-BANACH theorem.

Abram Ya. Dubovitzkij, Aleksej A. Milyutin:

1963: *Extremum problems under constraints*, Doklady AN SSSR, 149(4), 759-762.

Theorem: *Let $C_1, \cdots, C_m$ be convex cones in a Hilbert space H, C_i $(i = 1, \cdots k \leq m)$ be polyhedral and $C = C_1 \cap \cdots \cap C_m$. If*

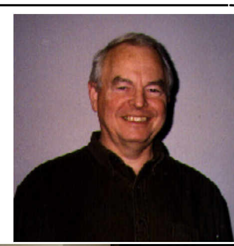
$$\left(\cap_{i=1}^{k} C_i\right) \cap \left(\cap_{i=k+1}^{m} \text{int } C_i\right) \neq \emptyset,$$

then it holds for the dual cone of C:

$$C^* = C_1^* + \cdots + C_m^*.$$

Dubovitzkij-Milyutin's formalism delivered the breakthrough: The characterization of the dual cone of the intersection of finitely many cones is an efficient tool for a unified approach to necessary optimality conditions. Till today this formalism is commonly used for treating different classes of optimization problems as well as in finite-dimensional and in infinite-dimensional spaces. In particular, it delivered new criteria for some difficult problems, for instance, control problems with phase constraints [8]. In the process of working out the *theory of convex analysis*, see for instance the monographs of ROCKAFELLAR [97] (for finite-dimensional spaces) and IOFFE AND TICHOMIROV [48] (for Banach spaces), these investigations were pushed on and became deeper.

R. Tyrrel Rockafellar:
1970: *Convex Analysis*, Princeton Univ. Press, Princeton.



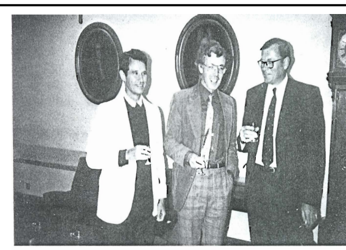
Alexander D. Ioffe, Vladimir M. Tichomirov:
1974: *Theory of Extremum Problems*, Nauka, Moscow.

Parallel to the development of a general theory in nonlinear optimization it became clear that numerical methods can be handled in an united framework, too.

Publications of FLETCHER AND REEVES [36] concerning conjugate gradient methods or papers of LEVITIN AND POLYAK [77] describe several *gradient- and Newton-like methods* for unconstraint optimization problems and their expansion to constraint problems. FIACCO AND MCCORMICK [35] published first results for *penalty-methods*. In the seventies and eighties the books and papers of DENNIS AND SCHNABEL [29], GOLDFARB [42], POWELL [92], DAVIDON [24], only to mention a few, are based on these numerical developments.

General theorems about convergence and rates of convergence for numerical algorithms in finite-dimensional and infinite-dimensional spaces have been proved and a great number of applications of nonlinear (especially global optimization problems), control problems, semi-infinite problems etc. have been considered. These analytic-numeric developments prolonged successfully over many years and numerous monographs appeared.

Robert Fletcher, C.M. Reeves:
1964: *Function minimization by conjugate gradient*, Computer Journal, 7, 149-154.



Davidon, Fletcher, Powell

Evgenij S. Levitin, Boris T. Polyak:
1966: *Minimization methods under constraints*, Zhurn. Vychisl. Matem. i Matem. Fiz. 6, 787-823.

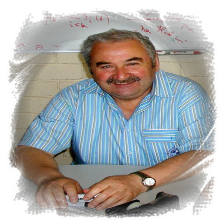
Polyak

Smooth optimization problems, involving differentiable functions, allow to apply descent methods with the help of gradient- and Newton-approximations. The Ukrainian mathematician NAUM SHOR [102] was the first who transferred this approach to non-smooth problems. In his PhD-thesis he suggested a subgradient method for non-differentiable functions and used it for numerically solving of a program, which is dual to a transport-like problem. Later this approach, named *bundle-methods*, was developed further by SCHRAMM AND ZOWE [101], LEMARECHAL[74] and KIWIEL [62].

In the seventies *non-smooth analysis* (this notion is due to F. CLARKE [16]) became a well-developed branch of analysis. Now some parts of this theory are more or less complete. Subdifferential calculus for a class of convex functions and minimax-theory belong to these parts and the latter was decisively developed by Clarke, Demyanov and Gianessi [17] and Demyanov and Rubinov (see, for instance, [25] – [28]).

Naum Z. Shor:
1962: *Application of the subgradient method for the solution of network transport problems*, Notes of Sc. Seminar, Ukrainian Acad. of Science, Kiew.

Vladimir F. Demyanov, Alexander M. Rubinov:
1968: *Approximate Methods for Solving Extremum Problems*, Leningrad, State University

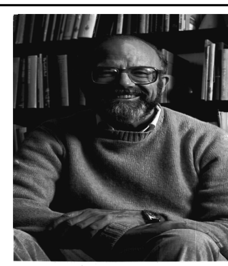


For a long time it was unknown whether linear programs belong to the class of problems which are difficult to solve (in non-polynomial time) or to the class of more easily solvable problems (in polynomial time).

Victor Klee (1925–2007)

1957: Professor at University of Washington, Seattle,

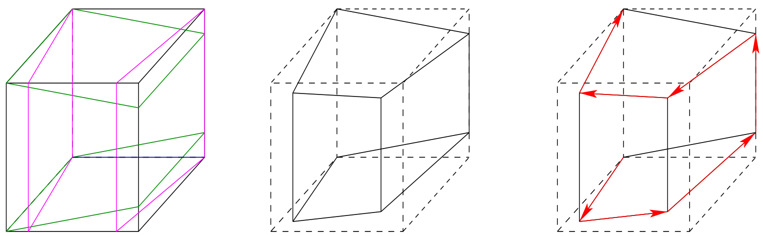
1995: Honorary doctor at the University Trier.



- Convex sets,
- Functional analysis (Kadec-Klee-Theorem),
- Analysis of algorithms, Optimization, Combinatorics.

In 1970 KLEE, who is also well-known in Functional Analysis, constructed some examples together with GEORG MINTY [63] showing that the classical simplex algorithm needs in the worst case an exponential number of steps. Because the number of vertices will grow exponentially if the dimension of a certain distorted standard cube increases exponentially, in the worst case all vertices of the cube must be visited in order to go to the optimal vertex.

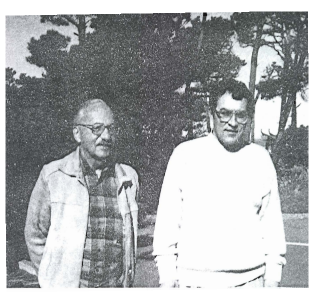
Construction of the Klee-Minty cube for $n = 3$. Now there exists an unique path along the edges of the cube, which hits all vertices and increases monotonously with respect to the objective function $\langle c, x\rangle$ with $c^T = (0, 0, -1)$.



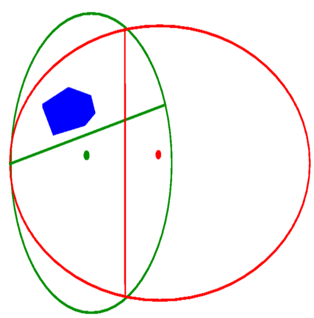
In 1979 LEONID KHACHIYAN [60] published the *ellipsoid method*, a method for determining a feasible point of a polytope.

Leonid Khachiyan:

1979: *A polynomial algorithm in linear programming*, Dokl. Akad. Nauk SSSR 244, 1093-1096.



Dantzig, Khachian (Asilomar, 1990)

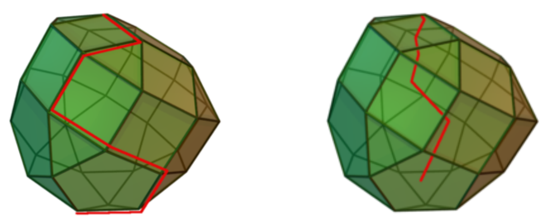


This method was initially proposed in 1976 and 1977 by YUDIN AND NEMIROVSKIJ [114] and independently of those by NAUM SHOR for solving convex optimization problems.
In 1979 KHACHIYAN modified this method and in doing so he developed the first polynomial algorithm for solving linear programs. It is a matter of fact that, by means of the optimality conditions, a linear program can be transformed into a system of linear equations and inequalities, hence one is dealing with the finding of a feasible point of a polyhedral set. However, for practical purposes this algorithm was not suitable.
The basic idea of the algorithm is the following: Construct some ellipsoid containing

all vertices of the polyhedron. Afterwards check whether the middle point of the ellipsoid belongs the polyhedron. If so, one has found a point of the polyhedron and the algorithm stops. Otherwise, construct the half-ellipsoid, in which the polyhedron should be included and put some smaller, new ellipsoid around the polyhedron. After a number of steps, depending polynomially on the code-length of the linear program, one has found a feasible point of the polyhedron or the polyhedron is empty.

In the mid-eighties, precisely in 1984, NARENDRA KARMARKAR [58] and others started developing *interior point methods* for solving linear programs.
In this connection we should mention the fate of a paper by DIKIN [30]. He was a student of KANTOROVICH and in his PhD thesis, at the advice of his supervisor, he suggested some procedure for solving linear programs numerically, although he failed to prove convergence estimates. This work did not find an interest and was forgotten until the late eighties. At that time it became clear that KARMAKAR's algorithm was very similar to DIKIN's method.

Narendra Karmarkar:
1984: *A new polynomial-time algorithm for linear programming*, Combinatorica 4, 373-395.



Simplex-method — interior point method

Interior point methods approach an optimal vertex right through the interior of a polyhedron, whereas the simplex method is running along the edges and vertices of the polyhedron. The significance of the interior-point approach consisted mainly in the fact that it was the first polynomial algorithm for solving linear programs having the potential to be useful also in practice. But the essential breakthroughs, making the interior point methods competitive to the simplex algorithm, took place in the nineties.
Advantages of these methods consist, in contrast to simplex methods, in their easy adaption for solving quadratic or certain nonlinear programs, so-called *semi-definite programs*, and their application to large-scale, sparse problems. One disadvantage is that, by adding a constraint or variable into the linear program, a so-called "warm-start" cannot carried out as efficiently as in simplex methods.

Now, let us once more come back to variational analysis. Also *variational problems, in particular variational inequalities*, have their origin in the calculus of variations associated with the minimization of functionals in infinite-dimensional spaces. The systematic study of this subject began in the early 1960s with the seminal work of the Italian mathematician GUIDO STAMPACCHIA and his collaborators, who used variational inequalities (*VI's*) as analytic tool for studying free boundary problems defined by nonlinear partial differential operators arising from unilateral problems in elasticity and plasticity theory and in mechanics. Some of the earliest books and papers in variational inequalities are LIONS AND STAMPACCHIA [78], MANCINO AND STAMPACCHIA [79] and STAMPACCHIA [104]. In particular, the first theorem of existence and uniqueness of the solutions of a VI was proved in [104]. The books by BAIOCCHI AND CAPELO [2] and KINDERLEHRER AND STAMPACCHIA [61] provide a thorough introduction to the application of VI's in infinite-dimensional function spaces. The book by GLOWINSKI, LIONS AND TRÉMOLIÈRE [41] is among the earliest references to give a detailed numerical treatment of such VI's. Nowadays there is a huge literature on the subject of infinite-dimensional VI's and related problems.

The development of mathematical programming and control theory has proceeded almost contemporarily with the systematical investigation of *ill-posed problems* and their numerical treatment. It was clear from the very beginning that the main classes of extremal problems include ill-posed problems. Among the variational inequalities, having important applications in different fields of physics, there are ill-posed problems, too. Nevertheless, up to now, in the development of numerical methods for finite- and infinite-dimensional extremal problems, ill-posedness has not been a major point of consideration. As a rule, conditions ensuring convergence of a method include assumptions on the problem which warrant its well-posedness in a certain sense. Moreover, in many papers exact input data and exact intermediate calculations are assumed. Therefore, the usage of standard optimization and discretization methods often proves to be unsuccessful for the treatment of ill-posed problems.
In the first methods dealing with ill-posed linear programming and optimal control problems, suggested by TIKHONOV [107, 108, 109], the problem under consideration was regularized by means of a sequence of well-posed problems involving a regularized objective and preserving the constraints from the original problem.
Essential progress in the development of solution methods for ill-posed variational inequalities was initiated by a paper of MOSCO [84]. Based on TIKHONOV's principle, he investigated a stable scheme for the sequential approximation of variational inequalities, where the regularization is performed simultaneously with an approximation of the objective functional and the feasible set. Later on analogous approaches became known as iterative regularization.
A method using the stabilizing properties of the proximal-mapping (see MOREAU [83]) was introduced by MARTINET [80] for the unconstrained minimization of a convex functional in a Hilbert space. ROCKAFELLAR [98] created the theoretical foundation to the further advances in iterative proximal point regularization for ill-posed VI's with monotone operators and convex optimization problems. These results attracted the attention of numerical analysts, and the number of papers in this field was increasing rapidly during the last decades (cf. [106]).

5 Conclusion

Meanwhile optimization has spread out to a powerful stream fed by ideas and works of thousands of mathematicians. New directions of the development were opened up

like *robust programming*, *variational inequalities and complementarity problems, mathematical programming with equilibrium constraints*, *PDE constrained optimization*, *shape optimization* and much more.
Likewise the area of applications in mathematical programming has widened continuously, new optimization models in *medicine and neurology*, *drug design*, *biomedicine*, to name only a few, appeared besides classical application areas in economy and natural sciences. At the 21-th *International Symposium on Mathematical Programming* 2012 in Berlin one could count 40 parallel sessions with approximately 1700 talks and about 2000 participants. Remember, at the congress in Chicago 1949 there were at most two dozens of talks. In the face of these figures it becomes clear that it is almost impossible to name all new developments and one has to restrict oneself in displaying the available results.

This paper is an attempt to describe the early beginnings and some selected mathematical ideas in optimization. Hereby mainly the progress in the American and Russian schools is pointed out. In my opinion these schools have distinctly influenced the development of mathematical programming. However, I am aware that the selected material and topics reflect mostly my personal view. It is completely obvious that other important trends in mathematical optimization have been neglected and substantial contributions of other nations are unmentioned.

6 Acknowledgement

I would like to express my gratitude to V.F. DEMYANOV who provided me with some photos of Russian mathematicians from his collection *Faces and Places*. Other photos I have taken from the books [75], [40] and from the web-sites of Wikipedia.

Literatur

[1] AKHIEZER, N.I., *The Classical Problem of Moments*, Fizmatgiz, Moscow, 1961.

[2] BAIOCCHI C. ,AND CAPELO, A., *Variational and Quasi-Variational Inequalities: Applications to Free Boundary Problems*, J. Wiley & Sons, Chichester, 1984.

[3] BALAKRISHNAN, A.V., Optimal control problems in Banach spaces, *J. Soc. Ind. Appl. Math., Ser. A, Control* 3 (1965) 152-180.

[4] BEALE, E.M.L., The use of quadratic programming in stochastic linear programming, *RAND Report P-2404*, RAND Corporation, 1961.

[5] BENDERS, J.F., Partitioning procedures for solving mixed-variables programming problems, *Numer. Math.* 4 (1962) 238-252.

[6] BELLMAN, R., *Dynamic Programming*, Princeton Univ. Press, Princeton, 1957.

[7] BERKOVITZ, L.D., An existence theorem for optimal control, *JOTA* 4 (1969) 77-86.

[8] BOLTYANSKIJ, V. G., Method of tents in the theory of extremum problems, *Uspekhi Matem. Nauk*, 30 (1975) 3-55.

[9] BOLTYANSKIJ, V. G., *Mathematische Methoden der optimalen Steuerung*, Akad. Verlagsgesellschaft Geest & Portig, Leipzig, 1971.

[10] BOLTYANSKIJ, V. G., *Optimale Steuerung diskreter Systeme*, Akad. Verlagsgesellschaft Geest & Portig, Leipzig, 1976.

[11] BOLZA, O. Über Variationsprobleme mit Ungleichungen als Nebenbedingungen, *Math. Abhandlungen*, 1 (1914) 1-18.

[12] BUTKOVSKIJ, A.G., *Distributed Control Systems*, Isd. Nauka, Moscow, 1969.

[13] CARATHÉODORY, C., *Calculus of Variations and Partial Differential Equations of the First Order*, New York, Chelsea Publishing Co., 1982.

[14] CHARNES, A., COOPER, W.W. AND MELLON, B., A model for optimizing production by reference to cost surrogates, *Econometrica*, 23 (1955) 307-323.

[15] CHARNES, A. AND COOPER, W.W., Chance-constrained programming, *Management Sci.*, 5 (1959) 73-79.

[16] CLARKE, F.H., *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[17] CLARKE, F.H., DEMYNOV, V.F. AND GIANESSI, F., *Nonsmooth Optimization and Related Topics*, Plenum Press, New York, 1989.

[18] DANTZIG, G.B., *Linear Programming and Extensions*, Princeton, 1963.

[19] DANTZIG, G.B., *Linear Inequalities and Related Systems*, Princeton Univ. Press, Princeton, 1966.

[20] DANTZIG, G.B., *Lectures in Differential Equations*, Van Nostrand, Reinbold Co., New York, 1969.

[21] DANTZIG, G.B., Linear programming under uncertainty, *Management Sci.*, 1 (1955) 197-206.

[22] DANTZIG, G.B. AND MADANSKI, A., On the solution of two-stage linear programs under uncertainty, in: *Proc. 4th Berkeley Symp. Math Stat. Prob.* Berkeley, 1961.

[23] DANTZIG, G.B. AND WOLE, PH., Decomposition principles for linear programs, *Oper. Res.*, 8 (1960) 101-111.

[24] DAVIDON, W.C., *Variable Metric Methods for Minimization*, A.E.C. Res. and Develop. Report ANL-5990, Argonne National Laboratory, Illinois, 1959.

[25] DEMYANOV, V.F. AND RUBINOV, A.M., *Approximate Methods for Solving Extremum Problems*, Leningrad State Univ., Leningrad, 1968.

[26] DEMYANOV, V.F. AND MALOZEMOV, V.N., *Introduction to Minimax*, Nauka, Moscow, 1972.

[27] DEMYANOV, V.F. AND VASILIEV, L.V., *Nondifferentiable Optimization*, Nauka, Moscow, 1981.

[28] DEMYANOV, V.F. AND RUBINOV, A.M., *Fundations of Nonsmooth Analysis and Quasidifferential Calculus*, Nauka, Moscow, 1990.

[29] DENNIS, J. AND SCHNABEL, R., Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, 1983.

[30] DIKIN, I., An iterative solution of problems of linear and quadratic programming, *Doklady AN SSSR*, 174 (1967) 747-748.

[31] DUBOVITSKIJ, A. YA. AND MILYUTIN, A.A., Extremum problems under constraints, *Doklady AN SSSR*, 149 (1963) 759-762.

[32] ERMOLIEV, Y.M., *Stochastic Programming Methods*, Nauka, Moscow, 1976.

[33] EULER, L., *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici latissimo sensu accepti*, Opera Omnia, Series 1, Volume 24, 1744.

[34] FARKAS, J., Theorie der einfachen Ungleichungen, *Journal für die Reine und Angewandte Mathematik* 124 (1902) 1-27.

[35] FIACCO, A.V. AND MCCORMICK, G.P., *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, 1968.

[36] FLETCHER, R. AND REEVES, C.M., Function minimization by conjugate gradient, *Computer J.* 7 (1964) 149-154.

[37] FORD, L.R. AND FULKERSON, D.R., *Flows in Networks*, Princeton Univ. Press, Princeton, 1962.

[38] FOURIER, J.B., *Analyse des Équations Déterminées*, Navier, Paris, 1831.

[39] GARDING, L., The Dirichlet problem, *Math. Intelligencer* 2 (1979) 42-52.

[40] GASS, S.I. AND ASSAD, A.A., An annotated timeline of operations research: an informal history, *Int. Ser. in OR & Management Sci.* 75 (2005).

[41] GLOWINSKI, R. AND LIONS, J.L. AND TRÉMOLIÈRES, R., *Numerical Analysis of Variational Inequalities*, North-Holland, Amsterdam, 1981.

[42] GOLDFARB, D., Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints, *SIAM J. Appl. Math.* 17 (1969) 739-764.

[43] GOLDMAN, A.J. AND TUCKER, A.W., *Theory of linear programming*, in: Kuhn and Tucker (eds.), *Linear Inequalities and Related Systems* Princeton, 1956, pp. 53-97.

[44] GOMORY, R., Outline of an algorithm for integer solutions to linear programs, *Bull. Amer. Math. Soc.* 64 (1958) 275-278.

[45] HALKIN, H., A maximum principle of Pontryagi, *SIAM J. Control* 4 (1966) 90-112.

[46] HESTENES, M.R. AND STIEFEL, E., Methods of conjugate gradients for solving linear systems, *J. Res. Natl. Bur. Stand.* 49 (1952) 409-436.

[47] HESTENES, M.R., *Calculus of Variational and Optimal Control Theory*, Wiley, New York, 1966.

[48] IOFFE, A.D. AND TIKHOMIROV, V.M., *Theory of Extremum Problems*, Nauka, Moscow, 1974.

[49] JACOBI, C. G. J., *Vorlesungen über analytische Mechanik : Berlin 1847/48*, (Nach einer Mitschr. von Wilhelm Scheibner), Hrsg. von Helmut Pulte, Deutsche Mathematiker-Vereinigung, Braunschweig, Vieweg, 1996.

[50] JOHN, F., Extremum problems with inequalities as subsidiary conditions, in: *Studies and Essays, Presented to R. Courant on his 60th Birthday, Jan. 1948*, Interscience, New York, 187-204.

[51] KALL, P., *Stochastic Linear Programming*, Springer Verlag, Berlin, 1976.

[52] KANTOROVICH. L.V., *Mathematical Methods for Production Organization and Planning*, Leningrad, 1939.

[53] KANTOROVICH. L.V., On an efficient method for solving some classes of extremum problems, *Doklady AN SSSR*, 28 (1940) 212-215.

[54] KANTOROVICH. L.V., Fuctional analysis and applied mathematics, *Uspekhi Matem. Nauk*, 6 (1948) 89-185.

[55] KANTOROVICH. L.V., *Functional Analysis*, Nauka, Moscow, 1959.

[56] KANTOROVICH. L.V., *Economical Calculation of the Best Utilization of Ressources*, Moscow, Academy of Sciences, 1960.

[57] KANTOROVICH. L.V., *The Best Use of Economic Resources*, Harvard U. Press, Cambridge, 1965.

[58] KARMARKAR, N., A new polynomial time algorithm for linear programming, *Combinatorica*, 4 (1984) 373-395.

[59] KARUSH, W., *Minima of functions of several variables with inequalities as side conditions*, *MSc Thesis*, Univ. of Chicago, 1939.

[60] KHACHIYAN, L., A polynomial algorithm in linear programming, *Doklady AN SSSR*, 244 (1979) 1093-1096.

[61] KINDERLEHRER, D. AND STAMPACCHIA, G., *An Introduction to Variational Inequalities and their Applications*, Acad. Press, New York - London, 1980.

[62] KIWIEL, K., Proximal level bundle methods for convex nondifferentiable optimization, saddle point problems and variational inequalities, *Math. Programming*, 69 B, (1995), 89-109.

[63] KLEE, V. AND MINTY, G., How good is the simplex algorithm? in: *Inequalities III*, O. Shisha, ed., Academic Press, New York (1972), 159-175.

[64] KOOPMANS, T.C., *Exchange Ratios between Cargoes on Various Routes (Non-Refrigerated Dry Cargoes)*, Memorandum for the Combined Shipping Adjustment Board, Washington, D.C., 1942.

[65] KOOPMANS, T.C., *Activity Analysis of Production and Allocation*, Wiley, New York, 1951.

[66] KOOPMANS, T.C. AND MONTIAS, M., On the Description and Comparison of Economic Systems, in: *Comparison of Economic Systems*, Univ. of California Press, 1971, pp. 27-78.

[67] KREIN, M.G. AND NUDELMAN, A.A., *Markov's Problem of Moments and Extremum Problems*, Nauka, Moscow, 1973.

[68] KUHN, H.W. AND TUCKER, A.W., Nonlinear Programming, *Proc. of the Second Berkeley Symp. in Math., Statistics and Probability*, Univ. of California Press, Berkeley 1951, 481-492.

[69] KUHN, H.W. AND TUCKER, A.W., Linear and nonlinear programming, *Per. Res.*, 5 (1957) 244-257.

[70] LAGRANGE, J.L., Théorie des Fonctions Analytiques, *Journal de l'École Polytechnique*, (1797).

[71] LAGRANGE, J.L., Mécanique Analytique, *Journal de l'École Polytechnique*, (1788).

[72] LANCASTER, L.M., The history of the application of mathematical programming to menu planning, *Europian Journal of Operation Research 57* (1992), 339–347.

[73] LEIFMAN, L.J., *Functional Analysis, Optimization and Mathematical Economics: A Collection of Papers Dedicated to the Memory of L.V. Kantorovich*, Oxford Univ. Press, 1990.

[74] LEMARECHAL, C., Lagrangian decomposition and nonsmooth optimization: Bundle algorithm, prox iterations, augmented Lagrangian, in *Nonsmooth Optimization: Methods and Applications (Erice, 1991)* Gordon and Breach, Montreux, 1992, 201-206.

[75] LENSTRA, J.K., RINNOOY KAN, A.H.G. AND SCHRIJVER, A. *History of Mathematikal Programming*, CWI North-Holland, 1991.

[76] LEVENBERG, K., A method for the solution of certain nonlinear problems in least squares, *Quarterly of Applied Mathem.* 2 (1944) 164-168.

[77] LEVITIN, E.S. AND POLYAK, B.T., Minimization methods under constraints, *Zhurn. Vychisl. Matem. i Matem. Fiz.* 6 (1966) 787-823.

[78] LIONS, J.L. AND STAMPACCHIA, G., Variational inequalities, *Comm. Pure Appl. Math.*, 20, (1967), 493-519.

[79] MANCINO, O.G. AND STAMPACCHIA, G., Convex programming and variational inequalities, *JOTA*, 9, (1972), 3-23.

[80] MARTINET, B., Régularisation d'inéquations variationelles par approximations successives, *RIRO* , 4, (1970), 154-159.

[81] MINKOWSKI, H., Allgemeine Lehrsätze über konvexe Polyeder. *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, math.-physik. Klasse* (1897), 198-219.

[82] MONGE, G., Mémoire sur la théorie des déblais et de remblais. Histoire de l' Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, (1781), 666-704.

[83] MOREAU, J.J., Proximité et dualité dans un espace Hilbertien, *Bull. Soc. Math. France*, 93, (1965), 273-299.

[84] MOSCO, U., Convergence of convex sets and of solutions of variational inequalities, *Advances in Mathematics*, 3, (1969), 510-585.

[85] NEUMANN, J.V., Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100 (1928) 295-320.

[86] NEUMANN, J.V., AND MORGENSTERN, O., *The Theory of Games and Economic Behavior*. Springer, New York, 1944.

[87] NEUSTADT, L.W., The existence of the optimal control in the absence of convexity conditions, *J. Math. Anal. Appl.* 7 (1963) 110-117.

[88] PADBERG, M., *Linear Optimization and Extensions*, Springer, 1961.

[89] PANNE, C. AND POOP, W., Minimum-cost cattle feed under probabilistic problem constraint. *Management Sci.* 9 (1963), 405-430.

[90] POLYAK, B.T., History of mathematical programming in the USSR. *Math. Programming, Ser. B 91* (2002), 401-416.

[91] PONTRYAGIN, L.S. AND BOLTYANSKI, V.G. AND GAMKRELIDSE, R.V., *Mathematical Theory of Optimal Processes*, Nauka, Moscow, 1956.

[92] POWELL, M., A method for nonlinear constrainets in minimization problems, in: Fletcher, R. (ed.): *Optimization*, Acad. Press, New York, (1969) 283-298.

[93] PSHENICHNYJ, B.N., *Necessary Conditions for Extrema*, Nauka, Moscow, 1969.

[94] PSHENICHNYJ, B.N., *Convex Analysis and Extremum Problems*, Nauka, Moscow, 1980.

[95] POUSSIN, C.J., Sure la méthode de l'approximation minimum. *Anales de la Societe Scientifique de Bruxelles* 35 (1911), 1-16.

[96] REMEZ, E. YA. , *Foundations of Numerical Methods for Chebyshev Approximation*, Naukova Dumka, Kiew, 1969.

[97] ROCKAFELLAR, R.T., *Convex Analysis*. Princeton University Press, 1972.

[98] ROCKAFELLAR, R.T., Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, 14, (1976), 877-898.

[99] ROUCHE, N., HABETS, P. AND LALOY, M., *Stability Theory by Lyapunov's Direct Method*. Springer, 1977.

[100] ROXIN, E., The existence of optimal control, *Michigan Math. J.* 9 (1962) 109-119.

[101] SCHRAMM, H. AND ZOWE, J., A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results, *SIAM J. Optim.* 2 (1992), 121-152.

[102] SHOR, N., Application of the subgradient method for the solution of network transport problems, Notes of Sc.Seminar, Ukrainian Acad. of Sci., Kiew, 1962.

[103] SLATER, M., *Lagrange Multipliers Revisited*. Cowles Commission Discussion Paper, No. 403.

[104] STAMPACCHIA, G., Formes bilineares coercives sur les ensembles convexes, *Comptes Rendus Acad. Sciences Paris*, 258, (1964), 4413-4416.

[105] STIEGLER, G.J., *Journal Econometrica* (1949).

[106] TICHATSCHKE, R., *Proximal Point Methods for Variational Problems*, 2011, http://www.math.uni-trier.de/ tichatschke/monograph.de.html

[107] TIKHONOV, A.N., Improper problems of optimal planning and stable methods of their solution, *Soviet Math. Doklady*, 6, (1965), 1264-1267 (in Russian)

[108] TIKHONOV, A.N., Methods for the regularization of optimal control problems, *Soviet Math. Doklady*, 6, (1965), 761-763 (in Russian).

[109] TIKHONOV, A.N., On the stability of the functional optimization problems, *J. Num. Mat. i Mat. Fis.*, 6, (1966), 631-634 (in Russian).

[110] TINTNER, G., Stochastic linear programming with applications to agricultural economics, *2nd Symp. Linear programming* 2 (1955), 197-228.

[111] VENTZEL, E.S., *Elements of Game Theory*, Fizmatgis, Moscow, 1959.

[112] VOROBYEV, N.N., Finite non-coallition games, *Uspekhi Matem. Nauk*, 14 (1959), 21-56.

[113] YUDIN, D.B. AND GOLSTEIN, E.G., *Problems and Methods of Linear Programming*, Sov. Radio, Moscow, 1961.

[114] YUDIN, D.B. AND NEMIROVSKIJ, A.S., *Complexity of Problems and Efficiency of Optimization Methods*, Nauka, Moscow, 1979.

[115] YUSHKEVICH, A.P., *Histrory of Mathematics in Russia before 1917*. Nauka Moscow, 1968.

[116] ZUKHOVITSKIJ, S.I., On the approximation of real functions in the sense of P.L. Chebyshev, *Uspekhi Matem. Nauk*, 11 (1956), 125-159.