Research News: Mathematics for Real Life Problems

Nonlinear Rescaling--a new efficient tool for solving real world constrained optimization.

by **Joshua Chamot**, **Office of Legislative & Public Affairs, National Science Foundation,**

edited by Colleen Kearney Rich, Mason University Relations

in collaboration with Dr. Roman Polyak.

Introduction

The real world is complex. When faced with a challenge, you can wade through a flood of options before arriving at the best outcome, if you arrive there at all.

Over the last 70 years, mathematicians, together with engineers, doctors, and economists, learned how to transform real-life problems into mathematical equations. By solving the mathematical problem, one finds the best possible answer to the real-world dilemma. This process is called optimization.

Over time, mathematicians have developed a number of tools for solving optimization problems. In this article, we will discuss one such tool, which happens to be suitable for solving a number of real-life nonlinear optimization problems (NLP) that were considered intractable just a few years ago.

The approach is based on the Nonlinear Rescaling (NR) principle originally introduced by Roman Polyak more then a quarter century ago back in the former Soviet Union. The NR approach is now used worldwide and has been critical for a range of applications, from processing image data to aiding the design of rigid towers to solving classification problems arising in statistical learning theory.

More recently, researchers began using nonlinear rescaling to tackle large-scale NLP facing doctors planning radiation therapy for their patients. For some patients, the method has meant the difference between recovery and suffering from a debilitating illness. A deeper look into the legacy of nonlinear rescaling reveals how we could never tackle some of the modern world's most complex problems without the aid of mathematics.

Optimization

"Nothing happens in the world without a law of optimization having a hand in it," said Leonard Euler a few centuries ago.

Companies rise and fall on their ability to optimize profits. Industries rise and fall on their ability to optimize their production lines. Countries rise and fall on their ability to optimize their economy in time of peace, and their military operations in time of war.

Despite its novel uses, optimization is not a novel concept. "People have been thinking about optimization since the ancient Greeks learned that a string with a given length encloses the most area when it is formed into the shape of a circle," said Roman.

Planning our budget, finding the shortest path to our friend for a party, or finding risk-free investment with the best possible return—all these are everyday optimization problems.

Linear Optimization

For much of the last century, one of the most effective optimization techniques has been linear programming (LP). If a business, such as an automobile manufacturer, wants to streamline production of a new car, linear optimization can reveal how to manage constraints such as limited energy resources, materials and labor to maximize profit.

Finding the cheapest way to deliver goods from a number of supply points with given supplies to a set of destinations with given demands is a typical linear programming problem .It has been considered by Leonid V. Kantorovich before World War II and independently by Tjalling C. Koopmans during the war. In 1975, they shared the Nobel Prize in Economic Sciences "for their contributions to the theory of optimum allocation of limited resources."

Every optimization problem has its objective function reflecting our target, and constraints reflecting the fact that our resources are limited. We have to find the best solution within available means. Shown on a graph, the LP problem would plot the constraints as intersecting straight lines and the solution lies in a vertex of the resulting polygon.

In 1947, George Dantzig devised the Simplex method for solving LP. The method was one of 10 best algorithms in the 20th century. The Simplex method travels from one vertex to another, improving the objective function up to the point when there is no better vertex.

For decades, this was, and still is, one of the best tools for handling LP problems. Practitioners solved thousands of real-life problems and saved billions of dollars. Only the late 80s simplex method got a strong competitor--the interior point methods (IPMs). While linear problems have a specific meaning to mathematicians, they can be better understood when one thinks about two traits that help define them: proportionality and additivity.

With proportionality, both the objective function and constraints have a linear cause-and-effect relationship with the outcome. For example, LP formulations assume that revenue is proportional to the fixed price of a product, independent of how much of it has been produced. It might contradict in some cases to the basic economic law of supply and demand.

With additivity, the impact of any one factor adds to the impact of the others, regardless of how the various factors interact with each other. For example, let’s consider an auto company which produces several types of cars.

In the LP formulation the total revenue of the company is the sum of the revenue of each type of cars independent of how many of them have been produced. Reality does not necessarily conform to the direct cause-and-effect environment of the linear world. While LP methods are fast and work well, when one solves transportation, diet or assignment problem, most processes in real life cannot be modeled as an LP. In the real world, answers are riddled with complexity and do not satisfy the two fundamental LP assumptions: proportionality and additivity. Nonlinear factors play a critical role in yielding the best solutions of many real-world problems.

Nonlinear Optimization

Radiation Therapy Planning (RTP), truss topology design (TTD), finding an optimal portfolio, optimal distribution of electricity across a power grid, antenna design, medical diagnostics or drug discovery, just to mention a few, are all Nonlinear Programming (NLP) problems.

The first two critical steps in NLP were made in the 17th century.

First, Fermat (1601–1665) introduced the optimality criteria for unconstrained optimization, allowing us to distinguish the minimum or maximum of a nonlinear function.

Second, Newton (1642–1727) devised a method for finding an extremum of nonlinear function or solving systems of nonlinear equations. Since then, the Newton method (1669) has been one of the most important tools in mathematics in general and in optimization in particular.

It is sufficient to say that Newton method played the key role in recent developments in the Interior Point Methods, which were the mainstream in optimization over the last 20 years.

The Newton method starts in the neighborhood of the extremum. Each step requires solving a linear system of equations doubling the number of exact digits at the solution. In very few steps, the method produces an approximation with up to ten digits of accuracy. The main limitation, however, is the local nature of the method — it converges to the solution only from a starting point in a small neighborhood of the solution.

The main NLP tool was and still is the Lagrangean, a function introduced by Lagrange in 1797. The Lagrangean combines the objective function and the constraints with weights-Lagrange multipliers, which one can view as prices for constraint violation. At the solution Lagrangean balances the gains of the objective function with the losses related to the constraint violation. Therefore, having the optimal Lagrange multipliers (prices), sometimes one can solve an NLP by finding one unconstrained optimizer of the Lagrangean, which is much easier than solving the original NLP.

One can view the classical Lagrangean as the bridge between constrained and unconstrained optimization. The bad news, however, is that the bridge, unfortunately, doesn’t always work, because the unconstrained Lagrange optimizer often doesn’t exist at all. We will see later how it can be fixed in the framework of the NR approach.

Over the last 50 years, several methods for solving NLP have been introduced. The most important requirements to such a method are: numerical stability (the method shouldn't fail), speed, and accuracy. The NR principle leads to such NLP methods, which met, to a large degree, all three requirements.

In short, the NR principle consists of replacing the original problem by an equivalent one, i.e., a problem which has the same solution. The Lagrangean for the equivalent problem is the basic tool in the NR theory.

The equivalent problem one obtains by nonlinear rescaling the objective function and/or the constraints. For rescaling, one uses a nonlinear scalar function of a scalar argument with particular properties. The NR method requires a positive scaling parameter, one for all constraints, or a vector of positive scaling parameters, one for each constraint.

The NR method alternates finding an approximation for the unconstrained optimizer of the Lagrangean for the equivalent problem with updating Lagrange multipliers (prices). The NR process can be viewed as a pricing mechanism. If the unconstrained optimizer violates a particular constraint, it means that we overused the corresponding resource. This leads to the price increase for this resource. If the optimizer strongly satisfies a particular constraint, then the corresponding resource is only partially used at this stage of the game, and the price for this resource will be reduced.

The NR methods offer explicit simple formulas for prices update. The rescaling drastically sharpens the price reaction to the constraint violation. It has strong impact on the computational process. In very few steps (price updates) one finds a good approximation for the optimal prices or optimal Lagrange multipliers.

What remains is finding the unconstrained optimizer of the Lagrangean for the equivalent problem. We have to keep in mind two important points: First, the unconstrained optimizer of the Lagrangean for the equivalent problem always exists. Second, after very few updates, the current optimizer is always in the neighborhood of the next one and can be used as a starting point in the Newton method for finding the next optimizer.

This makes the Newton method very efficient in the neighborhood of the solution. Each subsequent price update requires less Newton steps. At some point the NR method reaches the so-called “hot” start, when only one Newton step is required for the Lagrange multiplier update. Each update, in turn, often produces an extra digit of accuracy. It reduces substantially the number of Newton steps and allows finding approximations with very high accuracy. The main secret lies in the fact that NR method uses the Lagrangean for the rescaled problem. Such Lagrangean has several important advantages over the classical Lagrangean for the original problem.

Over the last 15 years, the NR methods were tested on a number of real-life applications, as well as on special sets of problems designed to fail the NLP solvers. The NR methods proved to be numerically stable and, in many instances, produced solutions with unprecedented accuracy in very reasonable time.

Let us concentrate on one important application: saving lives.

Saving Lives

Each year about 500,000 people receive radiation therapy treatment in the United States alone. The fundamental predicament of radiation therapy treatment is that it does not only affect ill but also healthy tissue. Therefore, for each individual patient, a treatment plan has to be established such that the radiation effects are sufficient to target the tumors while providing acceptably small radiation levels for neighboring healthy organs.

The technique of using beams of radiation to kill cancer cells is known as Intensity Modulated Radiotherapy Treatment (IMRT). Mathematicians have realized that planning radiation therapy can be formulated as an NLP problem.

German researchers Rembert Reemtsen of Brandenburg Technical University in Cottbus and Markus Alber of the University Hospital of Tübingen used the NR approach for solving the IMRT problem. In the case of IMRT, the correspondent nonlinear optimization problem maximizes radiation to the tumor under constraints that limit the exposure to healthy cells. The problem incorporates thousands of variables, such as beam angles and tungsten block positions, radiation intensity, and tens of thousands constraints, which are limiting the negative effect of radiation therapy on the healthy organs. The IMRT became a perfect candidate for using the NR approach due to its extreme stability and very high accuracy.

Combining the NR methodology with their own ideas and other modifications appeared in the last decade, Alber and Reemtsen were able to generate solutions with accuracy unprecedented so far. Their findings were published in the June 2007 issue of *Optimization Methods and Software.*

Tackling a model developed at Tübingen, the end result is an improved way to plan radiation therapy—one that has successfully treated real-world cases. Currently, the method is integrated into a full treatment system that is finding its way into various hospitals around the world. The treatments are helping doctors kill more tumor cells and fewer healthy ones in thousands of patients undergoing radiation therapy.

Turning Theory into Practice

One of the basic ideas in modern optimization is replacing a constrained optimization problem by a sequence of unconstrained optimization problems, which are much easier to solve. As we already saw using the classical Lagrangian, this does not always work.

In the 1950s and 1960s, the so-called barrier functions were introduced. For constrained optimization, the barrier function infinitely grows when the approximation approaches the boundary. Therefore the minimizer of the barrier function practically always exists, and it is an interior point of the feasible set. By increasing the barrier parameter, the minimizer gets closer to the optimal solution. This was the main idea of barrier methods for many years.

In the late 1960s, Fiacco and McCormick developed the corresponding theory called the Sequential Unconstrained Minimization Technique (SUMT), which became a classic in

Applied Mathematics. The main problem with the barrier function is that when the approximation gets closer to the solution it becomes more and more difficult to find the unconstrained minimizer of the barrier function. In a sense, the barrier function separates the optimal solution from the interior feasible set by an "infinite" wall: the barrier. Due to the singularity of the barrier function at the solution, the Newton method loses its efficiency near the solution, exactly where one can expect the method to be most efficient. The calculations become unstable; therefore, it is difficult if not impossible to find solution with high accuracy, which is absolutely critical in case like IMRT.

It was another 20 years before Yu.Nesterov and A. Nemirovsky discovered the remarkable Self-Concordance (SC) property of the log barrier function. The SC property became the foundation for the Interior Point Methods. Due to the SC properties, it became possible to improve substantially both the complexity bounds of the IPMs and their numerical efficiency. In spite of very impressive results, the SC theory has not removed the intrinsic difficulties associated with the singularity of the log barrier function at the solution and the need of unbounded increase of the barrier parameter to guarantee convergence.

The irony: The main difficulties associated with both classical Lagrangians and classical barrier functions had been resolved by Roman Polyak years before the IPM era even started. Unfortunately, it would be ten years before the optimization community learned about his discovery.

The first realization of the NR principle was the Modified Barrier Functions (MBFs) theory and methods developed by Polyak in the early 1980s. The MBF theory and methods eliminate not only the basic drawbacks of both the classical Lagrangean and the classical barrier functions, but it also combines the advantages of the two classical optimization tools.

First, MBFs removes the “infinite wall,” allowing the approximation to be outside the feasible set. In this regard, MBF is an exterior point method.

Second, along with the barrier parameter logarithmic MBF has an extra tool--the vector of Lagrange multipliers that characterize the prices for constraints violation. The extra tool eliminates the need of unbounded increase of the barrier parameter to guarantee convergence. At the same time, it allows improving the convergence rate and reducing the computational effort per step as compared to the classical barrier methods.

Third and most important, it restores the efficiency of the Newton method near the solution. In fact, Newton MBF method is much more efficient in the neighborhood of the solution then far from it. This allows MBF method and its modifications to find solutions with very high accuracy, which is critical in cases like IMRT.

Polyak and his colleagues have been refining NR concept for the last 25 years, adding new techniques and modifications that have proven to be valuable for a wide range of real-world applications.

In an important early effort, Aharon Ben Tal and Michael Zibulevsky made a slight modification of the MBF. Their penalty/barrier method is a particular realization of the

NR principle. The penalty/barrier method was developed at the Technion in Israel and had the most direct impact on the design of large-scale trusses.

A truss is a set of pin-jointed straight bars. The bars are subjected to axial tension and compression when the truss is loaded at the joints. With a given load and a given set of joints at which the truss is fixed, one has to find such bar volumes that the equilibrium conditions are satisfied. While one can see numerous examples of trusses on a trip down any major highway, the structures are far more complicated to build than they seem and the interlocking triangles composing a truss can be assembled in almost countless ways. One has to find a way to compose them that the construction not only can withstand given load, but it is as light as possible.

Before the NR tool emerged, the Israeli team could only handle relatively small problems (150 constraints or so). Using NR approach, the Technion researchers and others were able to tackle truss design problems with as many as 200,000 constraints and 5,000 variables and yield very accurate results much faster than other nonlinear optimization solvers.

One of the leading research groups studying truss design and other structural optimization problems has been guided for many years by Jochem Zowe at University of Erlangen-Nuremberg in Germany. Zowe and his team have used the software package that emerged from the Technion developments as their main optimization tool, and their work has affected structural designs around the globe.

The NR method and its progeny have been critical for applications ranging from the layout of the interlocking metal beams in tower trusses to strategies for distributing electricity across a power grid.

Recently the NR approach was used in statistical learning theory. For many years, the Support Vector Machine (SVM) technique developed by Vladimir Vapnik was and still is one of the basic tools used for data classification.

Roman Polyak and his former students Shen-Shyang Ho and Igor Griva developed the NR SVM model, which allows attaching to each vector of the data set specific waits (Lagrange multipliers), which characterize the contribution of each particular vector to the classification rule. The extra information allows, on the one hand, reduction of the number of support vectors practically without compromising the quality of classification and, on the other hand, improves the classification of the new cases due to the “double conformation” process. In other words, due to the extra information, the NR SVM along with classification of the new case provides conformation that the case belongs with high probability to a particular class and does not belong to the alternative class. In short, the NR approach substantially reduces the number of support vectors by removing vectors with small contribution into the classification rule and improves the quality of classification by using the “double conformation” procedure.

Lately the NR approach has been used for solving Semi-Definite Optimization, which has a wide area of real-world applications. The NR theory became the foundation for one of the best NLP solver, PENNON, developed by Michal Kocvara (Czech Republic) and Michael Stingl (Germany).

A Researcher's Story

Under normal circumstances, science is an international endeavor. But the circumstances under which Roman developed the NR theory were far from normal. In 1979, he was fired from his job in Kiev because of his desire to emigrate from the Soviet Union.

The authorities refused to grant him and his family an exit visa without any particular reason. He became a refusenik, which meant that he couldn't get a job according to his qualifications, publish papers, travel, or even participate in any scientific meetings. Roman continued his research in isolation from the scientific community.

He recognized the limitations of both SUMT and classical Lagrangean. At the same time, it was clear that both classical tools have important elements, which can complement each other. It was evident that the basic idea of the barrier method to keep the computation strictly inside the feasible set contradicts to the purpose of constrained optimization-finding the solution on the boundary.

Those days, removing the "barrier" was very much on his mind. He first shifted the barrier and then introduced the pricing mechanism for penalizing constrained violation.

The pricing mechanism replaced infinite penalty for constrained violation by a reasonable price increase, which depends on the size of the current deficit of a particular resource. That allowed the current approximation to stay outside of the feasible set, which makes MBF an exterior point method

The “freedom” of constraints violation substantially increased the efficiency of Newton method and allowed discovering the “hot” start phenomenon in constrained optimization.

Wide recognition of Roman’s results had to wait until 1988 when the Soviet government finally granted Polyak permission to leave the Soviet Union. Some of his colleagues were already familiar with his original work; therefore, at the end of 1988, he was offered a visiting position in the Mathematical Sciences Department, IBM T.J. Watson Research Center.

It was at IBM that his ideas were tested in the early 1990s and his paper “Modified Barrier Functions (Theory and Methods)” was published in 1992 in *Mathematical Programming*, the leading world journal on optimization.

While the human drama came to the happy end, the drama of ideas is still far from over. In the summer of 2006, Roman Polyak and his former PhD student Igor Griva announced a new optimization milestone: They had achieved a new level of speed and accuracy for solving real-life NLP problems by introducing a new Exterior Point Method, which is based on the NR theory. Their findings were published in *Mathematical Programming*.

.

The Newton method is the critical component of the NR success story. For hundred of years, the Newton method was local. Substantial efforts have been made for its “globalization,” i.e., to make sure that Newton method does not fail far from the solution.

Recently Roman developed the co-called Regularized Newton method, which allows finding the unconstrained optimum for any convex function from any starting point. Moreover, the Regularized Newton method retains under the standard conditions all of the best properties of the classical Newton method in the neighborhood of the solution, in particular, its quadratic convergence rate.

The corresponding paper will be published soon in *Mathematical Programming*.

Professor Roman Polyak together with his graduate students and colleagues advancing the NR theory and methods and finding new NLP real-life applications. Recently, he and his former PhD student Igor Griva were granted a U.S. patent for their NR optimization tools.

Roman is currently working on a book that will summarize the developments of the NR theory and methods for the last 25 years.