Who Invented the Interior-Point Method?

David Shanno

2010 Mathematics Subject Classification: 90C51, 90C05, 90C30
Keywords and Phrases: Interior-point methods, linear programming, nonlinear programming

The controversy

Thomas Edison is regarded by many as the greatest inventor in American history. While most people know that he invented the first long-burning incandescent light bulb and the phonograph, the claim is based more generally on the 1093 patents he was granted. The assumption is that the person receiving a patent is legally certified as the inventor of the device which is the subject of the patent.

The invention of the stored program computer during and in the period immediately following World War II vastly expanded the range of practical mathematical problems which could be solved numerically. A particular form of problem which received great interest is the linear programming problem, which allocates resources optimally subject to constraints. George Dantzig's development of the simplex method [5], provided the computational tool still prominent in the field today for the solution of these problems. Continuous development of variants of the simplex method has led to contemporary codes that are quite efficient for many very large problems. However, as the simplex method proceeds from one vertex of the feasible region defined by the constraints to a neighboring vertex, the combinatorial analysis indicates it can be quite inefficient for some problems. In [14], Klee and Minty showed that, in the worst case, the method has exponential complexity in the size of the problem.

The question that then presented itself is whether there is another algorithm for linear programming which has polynomial complexity. This question was first answered positively in 1979 by Khachian [13], who adapted the ellipsoid method of Shor [18] and showed that the complexity of the resulting algorithm was polynomial of order $\left(mn^3 + n^4\right) L$, where n represents the number of rows in A, m the number of columns, and L the length of the data. This result was an extremely important theoretical advance. It also created intense interest as a possible computational technique, including a wildly misinformed article in the New York Times claiming it solved the traveling salesman problem.

However, despite numerous attempts by many in the broad math programming community to implement a viable algorithm, it quickly became apparent that it was an extremely inefficient algorithm for computational work.

One interpretation of the simplex method is to consider what is purported to be the Norbert Wiener method of negotiating the halls of the massive main building at MIT. Not wishing to be distracted from thinking by watching where he was going, he simply dragged his hand along the wall, never removing it until he reached his destination. This algorithm clearly would eventually get him to where he was going, provided he began on the correct floor (an initial feasible point). I am not sure how he decided he had arrived, but in general this is akin to the simplex algorithm. A better method is to pay attention to where you are and take the best route. Interior-point algorithms attempt to emulate this strategy.

In a 1984 paper, Karmarkar [11] considered the linear programming problem in the form

$$\begin{aligned} \text{minimize } & c^T x \\ \text{subject to } & Ax = 0, \\ & e^T x = 1, \\ & x \geq 0. \end{aligned}$$

He began with an initial point x^0 that satisfied the constraints and used the projective transformation

$$T(x) = \frac{X_0^{-1} x}{e^T X_0^{-1} x}$$

where X_0 is the diagonal matrix $x_{jj} = x_j^0$. The current point x_0 is transformed to the point $\frac{1}{n}e$, which is the central point of the constraints $e^T x = 1, \; x_0 \geq 0$. Then, any vector in the null space of the matrix

$$\begin{bmatrix} AX_0 \\ e^T \end{bmatrix}$$

in particular

$$\delta = -\gamma \left[I - B^T (BB^T)^{-1} B \right] X_0 c,$$

can be used to reduce the objective function while remaining in the interior of the feasible region. Here, γ is a step length parameter to keep the step in the interior of the feasible region, which is accomplished by letting

$$\xi = \frac{1}{n} e + \delta$$

and the new estimate to the solution is

$$x^1 = \frac{X_0 \xi}{e^T X_0 \xi}.$$

Karmarkar demonstrated the complexity of this method is of order $(mn^2+n^3)L$, but the proof required that $c^T x^* = 0$, where x^* denotes the optimal solution. Todd and Burrell [19] dealt with this restriction by noting that if v^* is the optimal value of the objective function then

$$c^T x = (c - v^* e)^T x$$

is 0 at the optimal point. They then use duality theory to obtain a convergent sequence of estimates to v^*. Note that doing so adds a parameter to the sequence of estimates that will emerge in a different context shortly.

The originality of the use of projective transformations and the much stronger complexity results justifiably created a great deal of interest in the method. This interest, however, was mild compared to the interest created by a sequence of claims by Karmarkar and supported by Bell Labs, Karmarkar's employer, that an algorithm implementing the method was vastly superior to the simplex method.

A simpler transformation of the current point into the interior of the feasible region is the basis of the affine scaling method where instead of a projective transformation, the simple linear transformation was proposed by Barnes [2] and Vanderbei et al. [20]. Here, the standard form of the linear programming problem defined by

$$\begin{aligned} \text{minimize } & c^T x \\ \text{subject to } & Ax = b, \\ & x \geq 0 \end{aligned}$$

is used and the transformation becomes

$$\xi = X_0^{-1} x.$$

Here, the sequence of iterates is defined by

$$x^1 = x^0 + \gamma \Delta x,$$

where again γ is chosen to assure that the iterates do not touch the boundary of the feasible region and

$$\Delta x = \left[D - DA^T (ADA^T)^{-1} AD\right] c,$$

where

$$D = X_0^2.$$

It was later discovered that this work was originally published in 1967 by Dikin [6] who in 1974 proved convergence of the method [7]. No strong complexity bound equivalent to Karmarkar's is known for this algorithm.

Both of the above algorithms create room to move entirely in the interior of the feasible region by transforming the space. A more general method for

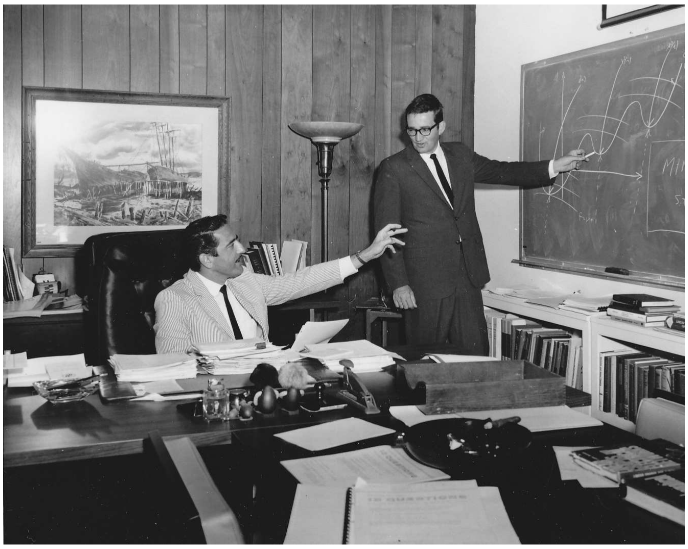

Figure 1: Anthony V. Fiacco (left) and Garth McCormick in 1967 in Fiacco's office at Research Analysis Corporation (RAC) in McLean, VA (Photo printed with the permission of John McCormick).

remaining in the interior was studied prior to either of these methods. An alternative method for remaining interior to the feasible region is to add a component to the objective function which penalizes close approaches to the boundary. This method was first suggested in 1955 in an unpublished manuscript by Frisch [9] and developed in both theoretical and computational detail by Fiacco and McCormick [8] in 1968. Applied to the linear programming problem in standard form, the problem is transformed to

$$\text{minimize } c^T x - \mu \sum_{i=1}^{n} ?\ln(x_i\ ),$$
$$\text{subject to } Ax = b.$$

Here, the method is akin to the invisible fence that is used to keep dogs in an unfenced yard. The closer the dog gets to the boundary, the more he feels shock. Here the amount of shock is determined by the parameter μ, and as μ tends to 0, the boundary, in this case where the solution lies, is approached.

The above reformulation is a nonlinear programming problem, and the first-order conditions may be derived by forming the Lagrangian and differentiating. The resulting step directions are

$$\Delta x = -\frac{1}{\mu_0} X_0 P X_0 c + X_0 P e,$$

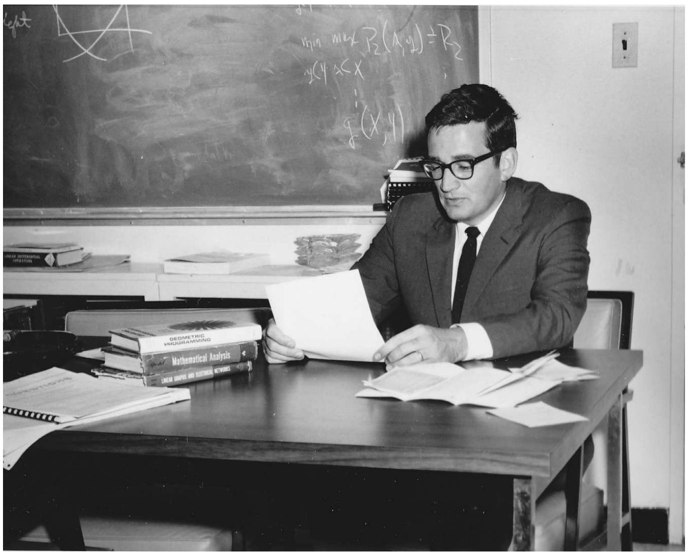

Figure 2: Garth McCormick at the desk in his office (Photo printed with the permission of John McCormick).

where

$$P = \left[I - X_0 A^T \left(A X_0^2 A^T\right)^{-1} A X_0\right],$$

and as before

$$x^1 = x^0 + \gamma \Delta x.$$

Fiacco and McCormick actually developed this method for the much harder general nonlinear programming problem. They showed that for a sequence of μ's which decreases monotonically to 0, the sequence of solutions for each value of μ converges to the solution of the problem. Their book noted that it applied as well to the linear programming problem, but did not further study this particular line of development as at the time they developed this work they felt the algorithm would not be competitive with the simplex method.

In 1985 at the Boston ISMP meeting, Karmarkar gave a plenary lecture in which he claimed his algorithm would be 50 or 100 times faster than the best simplex codes of that period. This was greeted with a great deal of skepticism and more that a little annoyance by many in the audience.

At the same meeting, Margaret Wright presented the results in Gill et al. [8] that showed there existed values for μ and v^* that make Karmarkar's algorithm a special case of the logarithmic barrier method of Fiacco and McCormick. This observation led to a major outpouring of theoretical papers proving order n^3L complexity for a wide variety of choices for the sequence of μ's and the search parameter γ. It also led to implementation work on numerical algorithms. An early example of this was the implementation of a dual-affine scaling algorithm (derived by applying the affine variant to the dual problem) of Adler et al. [1]. I was personally involved, first with Roy Marsten, in creating a dual-affine scaling implementation. We later joined with Irv Lustig to create an implementation of the primal-dual interior-point code [17] based on an algorithm published by Kojima et al. [15] which assumed the knowledge of an initial feasible point. We addressed initial feasibility using the analysis of Lustig [16]. We later discovered that the implemented algorithm can be derived directly by applying the Fiacco and McCormick logarithmic barrier method to the dual of the problem in standard form and applying Newton's method to the first order conditions.

Meanwhile, AT&T had begun development of the KORBX commercial package which included an eight processor supercomputer and an interior point code to be marketed at a multimillion dollar price. AT&T continued to claim (but not publish) strong computational results for their product. In 1988, they announced that they had obtained a patent on Karmarkar's method to protect their investment [11]. This patent in and of itself created quite a stir in the mathematics community, as up until that time mathematics was considered not patentable. However, the value of mathematical algorithms in the workplace was changing this view, and continues to do so today.

Irv, Roy and I meanwhile completed our first implementation of the primal-dual method [17], and in the fall of 1989 presented a computational comparison of our code with KORBX on a set of results which had finally appeared in publication [4]. The comparison was not favorable to KORBX. We distributed free of charge source of our OB1 code to researchers, but were marketing it to industry through XMP Software, a company Roy had started. Shortly after the presentation of the comparative results, we received a letter from AT&T informing us that, while they encouraged our promoting research in this area, we were not to market our code as they owned the patent on all such algorithms. This led us to carefully study the patent. The abstract of the patent follows.

> A method and apparatus for optimizing resource allocations is disclosed which proceeds in the interior of the solution space polytope instead of on the surface (as does the simplex method), and instead of exterior to the polytope (as does the ellipsoid method). Each successive approximation of the solution point, and the polytope, are normalized such that the solution point is at the center of the normalized polytope. The objective function is then projected into the normalized space and the next step is taken in the interior of the polytope, in the direction of steepest-descent of the objective function gradient and of such a magnitude as to remain within the interior of the polytope. The process is repeated until the optimum solution is closely approximated. The optimization method is sufficiently fast to be useful in real time control systems requiring more or less continual allocation optimization in a changing environment, and in allocation systems heretofore too large for practical

implementation by linear programming methods.

While the patent is for the Karmarkar algorithm, consequent discussions with AT&T patent lawyers made it clear that they were claiming that Karmarkar had invented interior point methods and they held the patent more broadly. The claim was obviously ridiculous, as there is a full chapter entitled *Interior Point Algorithms* in the Fiacco and McCormick book, which was published and won the Lancester prize in 1968. The people we were dealing with at AT&T seemed totally unaware of the existence of this book, despite its prominence in the mathematical programming community. The AT&T patent was granted in 1988, and there is a rule that nothing can be patented that has been in the public domain for a year or more prior to filing an application for the patent. Thus by the Edison criterion, Karmarkar invented the interior point method, but in fact he was well behind the true pioneers.

Meanwhile AT&T continued to claim to Roy, Irv and me that their patent applied to our code. After we consulted our own patent lawyer and were told what of the great expense of challenging the patent, we accepted a licensing agreement with AT&T. For a variety of reasons, the agreement proved to be unworkable, and we shut down XMP Optimization. We then joined with CPlex to create the CPlex barrier code. This code was derived by applying Newton's method to the log-barrier method of Fiacco and Mccormick applied to the dual problem. It is equivalent to an interior-point method, but using the term barrier rather than interior-point did not fall within the linguistic purview of the AT&T patent.It eventually became clear that AT&T had finally understood that the idea of interior-point methods did not originate with Karmarkar, and to the best of my knowledge they have never again tried to enforce the patent.

There is a further irony in AT&T receiving the Karmarkar patent. That patent is specifically for the projective transformation algorithm. Yet Bob Vanderbei, who was a member of the AT&T KORBX team, has told me that the method implemented in KORBX was the affine scaling method, which was also not eligible to be patented as Dikin's paper was published in 1967. AT&T did patent several techniques involved in the implementation of the affine scaling method [21], [22], such as how to incorporate bounds and ranges, but not the affine scaling interior point itself. Thus the only patent granted specifically for an interior point method was granted to the one algorithm that to the best of my knowledge has never been successfully implemented.

Who did invent interior-point methods?

With any invention that has proved highly successful, there is never a simple single answer to this question. A case can be made that Orville and Wilbur Wright invented the airplane. It is impossible to credit them alone with the creation of the Boeing 787. Further, in building the plane that made the first powered flight, they undoubtedly learned a great deal from others whose attempts had failed.

In a letter to Robert Hooke on February 15, 1676, Isaac Newton said "If I have seen further it is by standing on ye sholders of Giants." Personally, I fully credit Fiacco and McCormick with the invention of interior point methods, and as the result of many discussions with them over the years, I know that they fully agreed with Newton. Indeed a prominent giant in the development of interior point methods is clearly Newton himself, for all of the complexity results for linear programming depend on using Newton's method to solve the first order equations, and current nonlinear programming algorithms depend on Newton's method to find a search direction. Another such giant is Lagrange. Both are easy choices, as most methods for solving continuous math programming problems are highly reliant on their work.

On more recent work, both Frisch [9] and Carrol [3] must be credited with suggesting two different penalty functions to keep the iterates within the feasible region. Fiacco and McCormick certainly credited them. However, only Fiacco and McCormick developed a whole complete theory of interior point methods, including convergence results and a wealth of ideas for numerical implementation. They did not, however, analyze computational complexity. This field was really just beginning at the time of their work. The book contains many hidden gems, and as Hande Benson, a young colleague of mine has recently discovered, is still totally relevant today.

In addition, Fiacco and McCormick also developed the SUMT code to implement the general nonlinear programming algorithm documented in the book. Unfortunately, this was not the success that their theoretical work was. The difficulties encountered in attempting to solve many applications led some people to dismiss the practical value of interior point methods. The problem was simply that the theory was well in advance of computational tools developed later.

One particular difficulty was devising a good method to compute the decreasing sequence of μ's. This was greatly improved by the analysis done when applying the algorithm to linear programming. A good sequence is dependent on the measure of complementarity.

Another difficulty was nonconvexity of the objective function in nonlinear programming. The vast later research in trust region methods greatly improved the algorithms, and research on this continues today.

The algorithm of SUMT was a pure primal algorithm. The use of the interior point theory to derive primal-dual algorithms produced much better estimates of the Lagrange multipliers.

Central to applying the method to very large linear programming problems was the development of efficient sparse Cholesky decompositions to solve the linear equations. The computers at the time this research was done had such limited memories that this work had not yet been undertaken. At that time, it was believed that only iterative methods could be used to solve very large linear systems. The development of almost unlimited computer memories and the development of sparsity preserving ordering algorithms has allowed for very rapid solution of large sparse linear systems. These advances have also been applied to the solution of large sparse nonlinear programming problems.

Interior point algorithms require an initial feasible point x_0. Finding such a point for pure primal methods such as SUMT is often as difficult as solving the optimization problem. Development of primal-dual algorithms led to reformulation of the problem in such a way that a feasible initial point is easily found for the reformulated problems [16], [17]. The resulting algorithm approach feasibility and optimality simultaneously. This approach is now the standard approach in modern interior-point linear programming codes. It has also proved particularly important in improving interior-point algorithms for nonlinear programming, the problem that originally interested Fiacco and McCormick.

The salient point is that any great piece of original work is never close to a finished product, but rather a starting point from which improvements can be made continuously. It can also be extended to new areas of application. Certainly the work of Fiacco and McCormick meets that test of time. I know of no even vaguely comparable work on this topic.

References

[1] Adler, I., Karmarkar, N., Resende, M. and Veiga, G. (1989), An implementation of Karmarkar's algorithm for linear programming, *Mathematical Programming* 44, 297–335.

[2] Barnes, E. (1986), A variation on Karmarkar's algorithm for solving linear programming problems, *Mathematical Programming* 36, 174–182.

[3] Carrol, C. (1961), The created response surface technique for optimizing restrained systems, *Operations Research* 9, 169–184.

[4] Cheng Y., Houck D., Liu J., Meketon M., Slutsman L., Vanderbei R. and Wang P. (1989), The AT&T KORBX system. AT&T Tech. Journal, 68:7–19.

[5] Dantzig, G.(1963), *Linear Programming and Extensions,* Princeton University Press, Princeton, NJ.

[6] Dikin, I. (1967), Iterative solution of problems of linear and quadratic programming, *Soviet Mathematics Doklady* 8, 674–675.

[7] Dikin, I. (1974), On the speed of an iterative process, *Upravlyaemye Sistemi* 12, 54–60.

[8] Fiacco, A. and McCormick, G. (1968), *Nonlinear programming: Sequential Unconstrained Minimization Techniques,* John Wiley and Sons, New York.

[9] Frisch, K. (1955), The logarithmic potential method of convex programming, Memorandum, University Institute of Economics, Oslo, Norway.

[10] Gill, P., Murray, W., Saunders, M., Tomlin, J, and Wright, M. (1986), On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method, *Mathematical Programming* 36, 183–209.

[11] Karmarkar, N. (1984), A new polynomial time algorithm for linear programming, *Combinatorica* 4, 373–395.

[12] Karmarkar, N. (1988), Methods and apparatus for efficient resource allocation, United States Patent Number 4744028.

[13] Khachian, L. (1979), A polynomial time algorithm in linear programming, *Soviet Mathematics Doklady* 20, 191–194.

[14] Klee, V. and Minty, G. (1972), How good is the simplex algorithm? in O. Shisha, ed. *Inequalities – III,* Academic Press, New York, 159–175.

[15] Kojima, M., Mizuno, S. and Yoshise, A. (1989), A primal-dual interior point method for linear programming, in N. Megiddo, ed. *Progress in Mathematical Programming: Interior Points and Related Methods,* Springer Verlag, New York, 29–47.

[16] Lustig, I. (1990), Feasibility issues in a primal-dual interior-point method for linear programming, *Mathematical Programming* 49, 145–162.

[17] Lustig, I. Marsten, R. and Shanno, D. (1991), Computational experience with a primal-dual interior point method for linear programming, *Linear Algebra and its Applications* 152, 191–222.

[18] Shor, N. (1964), On the structure of algorithms for the numerical solution of optimal planning and design problems, Ph. D. Thesis, Cybernetic Institute, Academy of Sciences of the Ukrainian SSR, Kiev.

[19] Todd, M. and Burrell, B. (1986), An extension of Karmarkar's algorithm for linear programming using dual variables, *Algorithmica* 1:4, 409–424.

[20] Vanderbei, R., Meketon, M. and Freedman, B. (1986), A modification on Karmarkar's linear programming algorithm, *Algorithmica* 1:4, 395–407.

[21] Vanderbei, R. (1988), Methods and apparatus for efficient resource allocation, United States Patent Number 4744026.

[22] Vanderbei, R. (1989), Methods and apparatus for efficient resource allocation, United States Patent Number 4885686.

David Shanno, Professor Emeritus
RUTCOR – Rutgers Center of
Operations Research
Rutgers University
New Brunswick, NJ 08903-5062
USA
shanno@rutcor.rutgers.edu