The Origins of Fixed Point Methods

Herbert E. Scarf

In reflecting on my own involvement in the development of algorithms for approximating fixed points of a continuous mapping, I am struck by two recurring themes. The first of these is that chance meetings are terribly important; where I was at various times, and who my colleagues were, made a great deal of difference to me. I was very fortunate to find myself among extremely stimulating associates at the beginning of my professional career, people whose casual observations could shift my understanding of a problem in what were significant and even crucial ways. The second theme, which will be discerned very clearly in the following account, has to do with the indirect and circuitous nature of my own research and the vital importance, for me, of genuine perplexity.

The basic ideas of a numerical algorithm for approximating fixed points of a continuous mapping were assembled in the spring of 1965. This was the second year of my appointment as a Professor of Economics and as a member of the research staff of the Cowles Foundation for Research in Economics at Yale University. But my intellectual motivation for the project arose from a set of earlier concerns about economic theory which can best be described in a brief summary of my professional career up to that point.

I received my PhD in mathematics from Princeton University in 1954; my thesis was supervised by Salomon Bochner and was on the subject of diffusion processes on differentiable manifolds. During my years as a graduate student, there was a substantial amount of activity at Princeton in game theory, but I remained totally innocent of this topic, of linear programming and of the elements of mathematical economics. Ralph Gomory, Lloyd Shapley and Martin Shubik were fellow graduate students; we spent a good deal of time talking to each other, but not on those subjects which were to become our eventual professional concerns. I must admit that I found combinatorial topology very difficult to assimilate and probably would have been unable, at that time, to give a statement of Brouwer's fixed point theorem, let alone a proof.

In 1954 I took a position in the Department of Mathematics at the Rand Corporation in Santa Monica, California. At Rand, I learned the elements of game theory from Shapley, and we collaborated on a paper on dynamic games with incomplete information. George Dantzig had recently arrived at Rand, and the basic themes of linear programming were being developed and applied to a striking variety of problems. John von Neumann, Jimmie Savage and David Blackwell were frequent consultants to the organization. Richard Bellman was certain that every problem involving the allocation of scarce resources could be formulated as a dynamic programming problem, and Lester Ford and Ray Fulkerson were just beginning their influential work on optimal flows in networks. It was a marvelous location for a freshly minted, twenty-four year old PhD.

Two occasional visitors to Rand played an extremely important role in my subsequent career. Kenneth Arrow and Samuel Karlin were working on inventory problems and I myself had become interested in the same topic. Arrow and Karlin were kind enough to invite me to collaborate with them at Stanford University in 1956 and 1957 – a collaboration which resulted in several volumes of collected papers. I stayed on as a member of the Department of Statistics and the Institute for Applied Mathematics in the Behavioral Sciences at Stanford – with a break in 1959-60 as a visiting research scholar at the Cowles Foundation – until 1963 when I moved permanently to Yale.

The atmosphere at Stanford was genuinely exciting. There was a sense of the great potential utility of mathematical reasoning in a variety of novel areas: in mathematical biology, statistical decision theory, game theory and in mathematical economics. Arrow had completed his work on social choice, and had collaborated with Gerard Debreu on demonstrating the existence and welfare properties of a competitive equilibrium, with a generality and elegance made available by the theory of convex sets. Arrow was now working on the stability of equilibrium, which he was investigating jointly with Hirofumi Uzawa and Leo Hurwicz. It was this subject which I turned to as my first research topic in economic theory.

In order to describe the basic issue with which we were concerned, it may be useful to review the elements of the general equilibrium model – restricting my attention to the important special case of an exchange economy. Let there be n goods, so that a typical commodity bundle is represented by a vector $x = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$, and m consumers. The ith consumer is assumed to have a utility function $u_i(x)$, which specifies his preferences for the consumption of commodities, and, moreover, to have an initial supply of goods given by the vector w^i. Since there is no production in the model, the total supply of goods available for distribution among the consumers is $w = \Sigma w^i$.

A price system $p = (p_1, \ldots, p_n)$ is given by a nonnegative vector in $\mathbb{R}^n$, normalized so that $\Sigma p_i = 1$. If such a vector is announced, the ith consumer has an income given by $p \cdot w^i$, and his demands for potential commodities are assumed to be obtained by the maximization of his utility function $u_i(x)$, subject to the budget constraint $p \cdot x \leqslant p \cdot w^i$. Under plausible conditions, this results in a demand function $x_i(p)$ which is continuous and which satisfies the identity $p \cdot x_i(p) = p \cdot w^i$. The market demand functions $x(p)$ are obtained by adding up the individual demand functions over all consumers; they are also continuous and satisfy $p \cdot x(p) = p \cdot w$. And, finally, the market *excess* demand functions $f(p) = x(p) - w$ satisfy the identity $p \cdot f(p) = 0$, for all prices, which is known as the Walras law.

A competitive equilibrium is a price vector p^* for which $f(p^*) \leqslant 0$, i.e., a price vector for which the demand for each good in the economy is less than or equal to its supply (from the Walras law, $f_i(p^*) = 0$ if $p_i^* > 0$). The existence of such an equilibrium price vector had already been established by an appeal to Brouwer's fixed point theorem. But it was entirely conceivable, based on what was known in the 1950's, that the market excess demand functions satisfied a variety of properties other than continuity and the Walras law, which might allow for an existence proof more appealing economically and less mathematically demanding than one based on Brouwer's theorem.

If p is not an equilibrium price vector, then some of the goods will have a demand greater than their supply and the others an excess supply. Our intuitive notion of the response of prices to a discrepancy between demand and supply is that the price of a good in excess demand will rise, and the price of a good in excess supply will fall. The scheme can be formalized by a system of differential equations: $dp_i/dt = f_i(p)$, and the major question is whether the solution, starting at an arbitrary disequilibrium price, will converge to equilibrium.

We are all familiar with methods of this sort in mathematical programming. The interpretation of dual variables as prices permits the simplex method to be viewed as a type of price adjustment mechanism. Moreover, the price adjustment mechanism will always converge for strictly convex programming problems: maximize $\{f_0(x) \mid f_i(x) \leqslant b_i\}$, in the following sense. Let p be a price vector for the constraints and let $x(p)$ maximize $f_0(x) - \Sigma p_i f_i(x)$, without regard to the constraints. Then a solution to the differential equation $dp/dt = f(x(p)) - b$, with some adjustments for the boundary of the nonnegative orthant, will tend to the true vector of Kuhn-

Tucker prices for the nonlinear programming problem. The basic issue explored by stability analysis of the general equilibrium model is whether the introduction of consumers, with preferences and initial holdings, requires the use of mathematical techniques more sophisticated than those customarily used in mathematical programming.

An affirmative answer to the general question of stability would yield an alternative proof, independent of fixed point theorems, for the existence of a competitive equilibrium. But, even more significantly, I was intrigued by the possibility that, if the general equilibrium model were always globally stable, then the price adjustment mechanism would provide a numerical algorithm, of a familiar sort, for calculating an equilibrium price vector. The answer to the basic question is, however, negative; in 1959, I constructed a very simple set of examples for which the price adjustment mechanism was globally unstable. If the initial price was different from the unique equilibrium price, the solution of the differential equation would cycle forever, without converging. At the present time, this result is no longer surprising, since we now know that market excess demand functions are essentially arbitrary aside from continuity and the Walras law. We can, therefore, construct general equilibrium models in which the price adjustment mechanism follows virtually any prescribed path.

During my visit to the Cowles Foundation in 1959-60, I renewed my earlier friendships with Gerard Debreu and Martin Shubik. Shubik was in the audience when I gave a lecture at Columbia University on the example of instability, and we took a long walk to his apartment in the East 50's of New York City. During the walk, Martin described a problem that he had been concerned with, relating the *core* of an exchange economy to the set of its competitive equilibria. Using our earlier notation, the core of an economy is a distribution of society's assets $w = \Sigma w^i$, with the property that no coalition can find a position of higher utility for all of its members by a redistribution of the assets which they had originally owned themselves. Martin posed two questions, the first of which was whether each competitive equilibrium for the model of exchange gave rise to a distribution which was in the core, and, secondly, whether the core would converge to the set of competitive equilibria as the number of consumers tended to infinity. The first of these questions was answered immediately. Lloyd Shapley was spending the evening in Martin's apartment, and provided a positive proof within seconds of hearing the conjecture. The second question was a good deal more subtle and I took it back with me when I returned to Stanford in the summer of 1960.

There was an immediate conceptual difficulty: how to model an ever larger number of consumers without allowing for extremely diverse preferences? I took a clue from Shubik — and from Edgeworth, who had provided an analysis for the case of two goods in 1881 — and assumed a fixed number of types of consumers, each of whom was replicated precisely the same number of times in the passage to infinity. I was, after considerable difficulty, able to produce a proof of convergence to the set of competitive equilibria, under the rather restrictive assumption that the distributions being considered gave the same commodity bundle to all consumers of the same type.

I lectured on the result at a conference at Princeton University in 1962, where I first met Robert Aumann, who subsequently made a dramatic extension by formulating a model in which the number of consumers was of the power of the continuum, thereby avoiding my previous difficulty with types of consumers. Another dramatic simplification was unexpectedly announced to me by Gerard Debreu during a ride from the San Francisco airport to Stanford in the spring of 1962. Debreu presented an extremely simple argument showing that under mild assumptions on preferences, an allocation in the core would assign precisely the same commodity bundle to each consumer of the same type. He then provided an elegant and geometrically appealing proof for the main theorem to replace my earlier, convoluted reasoning.

A reader might very well ask what all of this has to do with the computation of fixed points. After returning to Yale in 1963, and spending the first year adjusting to a new subject – economics – in which I had no formal training, it occurred to me that, if I could establish the non-emptiness of the core, under classical assumptions, then the convergence proof would provide an alternative argument for the existence of a competitive equilibrium. I had, moreover, worked out a constructive algorithm for finding a point in the core of an economy with three consumers whose preferences satisfied the customary convexity assumptions. With perhaps more optimism than was warranted, I attempted to extend this constructive argument to the general case of n consumers, with the ultimate goal of obtaining an algorithm for the computation of equilibrium prices.

The project was far more difficult than I had expected. I did formulate a general theorem stating that a *balanced* n-person game, without transferable utility, had a non-empty core, but the proof of the theorem appealed to precisely those fixed point theorems whose use I was trying to avoid. I was aware of another line of attack, based on a combinatorial lemma involving mathematical objects which I had termed *primitive sets*, but I could find no algorithmic argument for the lemma, even in the special case of four consumers.

I was, however, extremely fortunate. Bob Aumann was visiting at Yale during the academic year 1964-65, and one day when I complained to him about my frustrations, he suggested that I might wish to take a look at a recent paper by Lemke and Howson, which provided an algorithm for computing Nash equilibrium points for a general two-person non-zero sum game. Prior to Lemke's work, the existence of a Nash equilibrium for an n-person game had been demonstrated by an appeal to Kakutani's fixed point theorem, and, as such, there was no effective algorithm for its calculation. Kuhn had shown that a vector representing a Nash equilibrium – for a two-person non-zero sum game – would lie on a vertex of a certain convex polyhedron, but his argument required the examination of all vertices, and Kakutani's theorem was still necessary to demonstrate existence. Lemke's ingenious path following argument not only demonstrated existence of Nash equilibria for two-person games, but also provided an effective algorithm which typically traversed a small subset of all possible vertices. The algorithm is also available for the more general linear complementarity problem, and was fundamental in initiating this vital field.

The essence of Lemke's algorithm can be found in the following charming little tale. Imagine a house with many rooms, each of which has precisely two doors. The number of rooms is finite and one of the rooms has a door opening to the outside of the house. Then there must be at least one other door leading out of the house, which can be found by the following algorithm. Enter the house through the known door to the outside, and then move from room to room, always exiting through the door not used to enter the room. It is easy to argue that the algorithm never revisits a room previously encountered; since the number of rooms is finite, the algorithm must terminate, and termination is only possible by encountering a second door to the outside.

One look at the Lemke-Howson paper was enough to convince me that the path following argument could be translated directly to the combinatorial setting I had been struggling with for so long. I showed Aumann the algorithm that same evening, spent the next several weeks learning how to write Fortran programs, and calculated my first example of a point in the core of a four-person exchange economy.

I was still quite far from a general algorithm for computing fixed points of an arbitrary continuous mapping. My computational procedure did permit me to approximate a competitive equilibrium in the sense that I could obtain an approximate point in the core of an exchange economy, and, if the number of consumers was sufficiently large, this would approximate a competitive equilibrium. The required double limit was extremely unsatisfactory, however, and it was not until the fall of 1966 that I realized that my combinatorial lemma, and its related algorithm for finding a completely labeled primitive set, could be used to prove Brouwer's theorem directly and to calculate an approximate fixed point of a general continuous mapping. The paper, 'The Approximation of Fixed Points of a Continuous Mapping', was written in early 1967. I am astonished, as I reread the paper, that I was still totally unaware of the relationship between my combinatorial lemma and Sperner's lemma – the classical combinatorial approach to Brouwer's theorem. The two mathematical results are virtually identical, but, for some unknown reason, I could not rid myself of the prejudice that there simply was no constructive algorithm to find the completely labeled simplex whose existence is asserted by Sperner's lemma.

The computational procedure worked pretty well in practice, and I had solved quite a few numerical examples by the time I left for Israel, on a professional visit, in the spring of 1967. Just before leaving, Terje Hansen, who was at that time a graduate student in my class in mathematical economics at Yale, told me that he had discovered a great improvement in the algorithm, and asked me to provide him with some computer time to test his method. When I returned, some six weeks later, Hansen had programmed his algorithm and was solving fixed point problems whose dimensions were considerably larger than those I had previously dealt with.

In order to describe Hansen's improvement, it is useful to be explicit about the concept of primitive sets. Let $x(n+1),\ldots,x(N)$ be a large set of vectors on the simplex $S=\{x=(x_1,\ldots,x_n): x_i \geqslant 0, \Sigma x_i = 1\}$, more or less uniformly distributed throughout the simplex. Make the non-degeneracy assumption that no two vectors $x(j)$ and $x(k)$ have the same ith coordinate for any i. Then $\{x(j^1),\ldots,x(j^n)\}$ form a primitive set if no vector $x(k)$ is strictly larger, in all coordinates, than $\min\{x(j^1),\ldots,x(j^n)\}$. We also define n special vectors $x(1),\ldots,x(n)$ — which are not themselves on the simplex — by letting $x(i)$ be negative in its ith coordinate and greater than 1 in its remaining coordinates, and extend the definition of primitive sets to include some of these special vectors — called *slack vectors* — in addition to those lying on the simplex.

If the set of points $\{x(j)\}$ is fairly dense, then the vectors forming a primitive set will be close to each other, and close to the ith face of the simplex if the ith slack vector is contained in the primitive set. Moreover, if $\{x(j^1),\ldots,x(j^n)\}$ is a primitive set, then any particular one of its members can be removed, and there will be a unique replacement which forms a new primitive set — except for one special case: the primitive set consists of $n-1$ slack vectors and one vector from the original list, which we are attempting to remove. It is this replacement property which permits us to wander throughout the simplex — using Lemke as a guide — in search of a specific primitive set with desirable properties. For example, if each of the vectors $x(j)$ is assigned an integer label from the set $\{1,\ldots,n\}$, which is arbitrary aside from the stipulation that $x(i)=i$ for $i=1,\ldots,n$, a variant of Lemke's argument will yield a path of primitive sets starting at a vertex of the simplex and terminating with a completely labeled primitive set.

Now let $x \rightarrow g(x)$ be a continuous mapping of the simplex into itself. If we label a vector $\xi \in \{x(n+1),\ldots,x(N)\}$ with the first subscript i for which $g_i(\xi) \geqslant \xi_i$, it is easy to see that any non-slack vector in a completely labeled primitive set will yield an approximation to a fixed point of the mapping. This does yield a workable algorithm for approximating fixed points, but its applicability is compromised by the fact that the replacement step requires a search throughout the entire list of vectors $\{x(j)\}$. It was this crucial problem that Hansen solved by providing an alternative to my concept of primitive sets.

Let e^i be the n-vector $(0, \ldots, 0, 1, -1, \ldots, 0)$ with $+1$ in the ith position and -1 in position $i+1$ (if $i = n$, the first coordinate is -1 and the last coordinate 1). Hansen considered $n \times n$ integral matrices $[x^1, \ldots, x^n]$, with the property that $x^{j+1} = x^j + e^{\phi(j)}$, with $\phi(j)$ a permutation of the integers $1, \ldots, n$ (if $j = n$, $j+1$ is understood to be 1). It is easy to verify that there is a unique, easily computed replacement for any column in such a matrix, yielding another matrix of precisely the same form. If the columns are non-negative and D is the common column sum, then the vectors x^j / D lie on the unit simplex. When the labels associated with the vectors x^j are appropriately assigned, the columns of a Hansen matrix with distinct labels can be used to approximate fixed points as before, but the replacement operation is now trivial to carry out for problems of substantial size.

Let the columns of such a matrix, each divided by D, be the vertices of a subsimplex in the unit simplex S. Then the collection of subsimplices arising from the set of all Hansen matrices form a classical simplicial subdivision of the simplex known as the Freudenthal subdivision. But we did not know this elementary geometrical fact in 1967; to the best of my recollection neither Hansen nor I ever made a drawing which would have immediately revealed that we were working with a simplicial subdivision and that our combinatorial lemma was virtually identical with Sperner's lemma.

Our innocence was not to last much longer. In January of 1968, I received a letter from Bob Aumann alerting me to a paper by D. I. A. Cohen entitled 'On the Sperner Lemma', which had been published in the *Journal of Combinatorial Theory* in June of 1967. In the paper, Cohen uses an argument similar to that of Lemke to establish the existence of an odd number of completely labeled simplices, under the usual assumptions of Sperner's lemma. Cohen does not appear to have thought about converting his proof into an algorithm for calculating a completely labeled simplex; this would have required some reworking since his argument uses induction on the dimension of the simplex. The same issue of the journal also contains a paper by Ky Fan, which generalizes Cohen's approach from the simplex to an orientable pseudomanifold, and which presents the tale of the house each of whose rooms has precisely two doors. And, finally, in April of that same year, Harold Kuhn sent me a draft of an article, eventually published in the *Proceedings of the National Academy of Sciences*, in which he directly used the Lemke argument to find a completely labeled simplex implied by Sperner's lemma. During our discussions of the next several months, Harold and I became aware of the geometrical interpretation of Hansen's construction and of the basic identity of the two algorithms.

By the middle of 1967 we were in possession of an effective algorithm for the approximation of fixed points of a continuous mapping. It is not possible, in this essay, to summarize the many improvements in the basic algorithm realized during the next two decades, and the great variety of applications that have been made in economics and other fields. But I will mention one particular set of refinements which has greatly enhanced the performance of fixed point methods. The earliest algorithms had the drawback that the computation was initiated at a vertex of the unit simplex. If the accuracy obtained at the final completely labeled subsimplex was not sufficient for the problem at hand, we could either avail ourselves of some numerical technique like Newton's method — not guaranteed to converge — or take a finer subdivision, and reinitiate the algorithm at a vertex. If the latter approach was selected, the information obtained from previous calculations was discarded totally.

Kuhn's paper contains a nibble at this problem, in the sense that his algorithm can be started anywhere on the boundary of the simplex, rather than at one of the n vertices. But the decisive improvements were made by Eaves, Merrill, and Van der Laan and Talman, whose algorithms permit the computation to be initiated at an arbitrary point on the simplex and allow a continual refinement of the simplicial subdivision in the course of the procedure. These algorithms, which are essentially piece-wise linear homotopy methods, are used in virtually all applications. They are intimately related to the continuous path following algorithms of Kellogg, Li and Yorke, and Smale.

* * *

I finally met Professor Sperner at a conference on fixed point methods at the University of Southampton in July of 1979. The two of us alighted from the same train, and after a few moments of hesitation, we introduced ourselves. We talked about the consequences of the theorem he had demonstrated some fifty years earlier, and he seemed very pleased. Afterwards, reflecting on our talk, I thought about that mysterious feature of mathematics, always perplexing and enchanting to me: the way in which an act of pure imagination can lead to unexpected practical applications in areas that are totally removed from the originator's wholly abstract concerns.