Herbert E. Scarf

Fixed-Point Theorems and Economic Analysis

Mathematical theorems can be used to predict the probable effects of changes in economic policy

During the first decade of this century, the eminent Dutch mathematician L. E. J. Brouwer (1912) demonstrated a basic mathematical theorem that has found an extraordinary variety of important uses in both pure and applied mathematics. Brouwer's fixed-point theorem, as it has come to be known, is a generalization to higher dimensions of the elementary theorem of the calculus that a continuous function of a single variable that is positive at one end of an interval and negative at the other end must vanish at some point in between.

Brouwer, one of the founders of modern topology, discovered his theorem while investigating the geometry of sets of points in n-dimensional space. As he demonstrated, the behavior of continuous mappings of one set into another (or of a set into itself) reveals important topological properties of the underlying sets. Most of the subsequent applications, however, are only incidentally geometric. Brouwer's theorem and other fixed-point theorems are used, most frequently, as a tool for demonstrating that a system of highly nonlinear equations has a common zero.

Herbert E. Scarf is Sterling Professor of Economics and Director of the Cowles Foundation for Research in Economics at Yale University. He received a Ph.D. in mathematics from Princeton University in 1954. After two years at the RAND Corporation in Santa Monica, Dr. Scarf joined the Statistics Department at Stanford University, where he remained until moving to Yale in 1963. His research has been primarily in mathematical economics—in particular, n-person game theory, the stability of economic equilibria, the computation of equilibrium prices, and the role of large-scale indivisibilities in production. He is currently President of the Econometric Society. Address: Department of Economics, Yale University, New Haven, CT 06520.

So far as I am aware, the first use of a fixed-point theorem in economic theory appeared in John von Neumann's model of an expanding economy published in 1937. Although von Neumann was not himself a professional economist, his paper was extremely influential, and, by the early 1950s, fixed-point theorems were familiar to virtually all mathematical economists. These novel mathematical techniques became available for the solution of a problem of the greatest economic significance: the demonstration of the existence of prices that satisfy the simultaneous equations and inequalities of the general equilibrium model of an economy.

By the late 1960s a variety of numerical algorithms (e.g., Scarf 1967, 1973; Hansen 1968 diss.; Kuhn 1968) had been developed for calculating—rather than merely asserting the existence of—the fixed points implied by Brouwer's theorem. When applied to the general equilibrium model, these computational methods provide an extremely powerful tool for analyzing the probable economic consequences of changes in a variety of policies. In this article, I shall provide an introduction to these basic computational techniques and some indication of the ways in which they are used.

The general equilibrium model is the basic economic paradigm that describes the functioning of a purely competitive market economy. The product of nearly two centuries of conceptual innovation and continued intellectual refinement, it has its beginnings in Adam Smith's analysis of the way in which capitalists respond to profit-making opportunities, and it attains its mature form in the general mathematical model presented by Leon Walras in 1874. In Walras's formulation, the agents in the economy are divided into two broad classes: producers, engaged in the transformation of factors of production into desired commodities, are distinguished from consumers, whose goals are the consumption of goods and services. The stocks of commodities in the economy, which may be consumed directly or used as inputs into production, are assumed to be entirely owned by consumers either in their tangible form—such as land, labor, or durable goods—or by means of a variety of financial instruments including corporate stocks and government bonds. In other words, producers transform what the consumers own into other goods that the consumers desire.

If the prices of all the goods and services in the economy are known, each consumer's income (or wealth) is determined by the market value of his assets. Income and a knowledge of relative prices permit the consumer to express his demands for consumer goods—as well as his offerings of labor, land, raw materials, and other stocks of commodities, which then become available to the producers. In a formal mathematical treatment, market demands, which are obtained by adding up the demands of the individual consumers in the economy, will be specific functions of the relative prices of *all* goods and services. The market demand for coffee, for example, will depend not only on the price of coffee, but also on the price of other drinks such as tea, which are substitutes for coffee.

The demand for coffee will also depend on the wage rate and the price of capital (as well as the prices of other assets owned by consumers), since these prices affect the consumers' disposable income. The de-

The general equilibrium model

Let us illustrate the general equilibrium model by considering a very simple example involving two consumers and two commodities: apples and oranges. No production will take place, and our goal will be to find the equilibrium prices that equate the market demand for each of the two goods to their supply. The following table describes the quantities of the two goods owned by the two consumers prior to trade:

	Apples	*Oranges*
Consumer 1	8	0
Consumer 2	4	8
	12	8

If the prices of the goods are π_1 and π_2, respectively, then the first consumer's income, obtained by selling his assets at the market prices, will be $8\pi_1$, and the second consumer's income will be $4\pi_1 + 8\pi_2$. Each consumer will have his own demand function (reflecting his private preferences for apples and oranges), which describes the quantities of each good demanded as a function of the two prices and that consumer's income. For simplicity let us assume that the first consumer spends one-quarter of his income on apples and three-quarters on oranges. The second consumer will be assumed to spend half of his income on each of the two goods.

If $\pi_1 = \$1$ and $\pi_2 = \$2$, the first consumer's income is \$8 and the second consumer's is \$20. The demands for apples and oranges will be given by:

	Apples	*Oranges*
Consumer 1	2	3
Consumer 2	10	5
	12	8

These prices will be the equilibrium prices, since the demand for each commodity is equal to its supply.

mand may depend, as well, on the price of a variety of other goods not immediately perceived as substitutes for coffee: for example, a high price for clothing may decrease the income available for the purchase of beverages.

On the production side of the economy, each producer is assumed to have complete knowledge of the different ways in which specific inputs can be transformed into specific outputs. Relative prices are taken to be independent of each producer's scale and composition of productive activity; when these prices are known, each producer selects—from among the alternatives that are technically possible for him—the production plan that maximizes his profit. Aggregation of these individual decisions results in the economy-wide supply functions that describe the levels of outputs and use of materials as mathematical functions of all relative prices.

If the general equilibrium model is constructed in a mathematical form—as it was by Walras—we are led to two explicit sets of functions of all relative prices: the market-demand functions and the market-supply functions. For arbitrary prices the values of these functions need not be consistent with each other. Producers cover the costs of production by the revenue obtained from sales. If the price of a specific commodity is too low, consumers may demand larger quantities than producers are willing to supply. Only very special prices—the equilibrium prices—will equate demand and supply in all markets simultaneously.

Fixed-point theorems can be used to demonstrate the existence of solutions to complex systems of equations. But are they necessary in order to demonstrate that there will be a set of equilibrium prices? Walras was, of course, unaware of Brouwer's theorem and the two elementary arguments for the existence of equilibrium prices that he presented seemed quite adequate to him (Walras 1874). One of the arguments was based on the assertion that a system of equations with the same number of equations as unknowns would always have a solution. But the most elementary examples (for instance, $x^2 = -1$) show that this need not be the case, and the argument is not conclusive unless there is something quite special about the equations arising from the general equilibrium model.

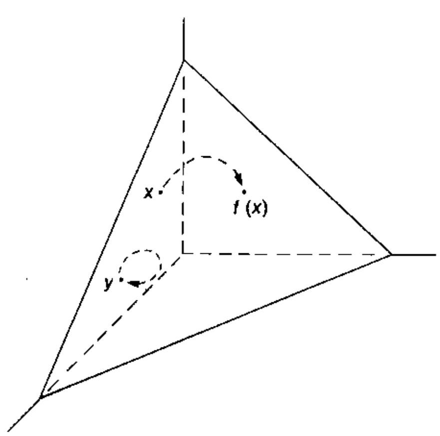

Figure 1. The general equilibrium model of the economy states that there are prices that equate supply and demand for all goods; these prices are fixed points of a particular mapping. In the sample mapping shown here, x is mapped to $f(x)$. However, y is mapped to itself and is therefore a fixed point.

The everyday observation that the price of a commodity will rise if its demand exceeds its supply and will otherwise fall was translated mathematically by Walras into a system of differential equations stating that the rate of change of the price of each commodity is proportional to the difference between its demand and supply. His second argument for the existence of equilibrium prices is that the system of differential equations can indeed be solved, at least conceptually, and if the solution path converges it will certainly converge to an equilibrium position. But again, differential equations need not have stable solutions, and it is quite easy to construct a system arising from a general equilibrium model whose solution wanders forever without converging. The equations for adjusting prices, therefore, represent neither an argument for the existence of equilibrium prices nor an effective computational procedure for their determination.

Equilibrium prices were finally proved to exist during a remarkable burst of intellectual activity in mathematical economics that took place in the early 1950s. Based on the work of Arrow, Debreu, Gale, Kuhn, McKenzie, Nikaido, and others (e.g., Arrow and Debreu 1954), the general equilibrium model was formulated with great generality, and fixed-point theorems were used to demonstrate the existence of prices that simultaneously equate demand and supply in all markets. A fundamental question in economic theory—the consistency of the general equilibrium model—was given a definitive answer.

Remarkable as this achievement certainly was, it seemed to be deficient in at least one major respect.

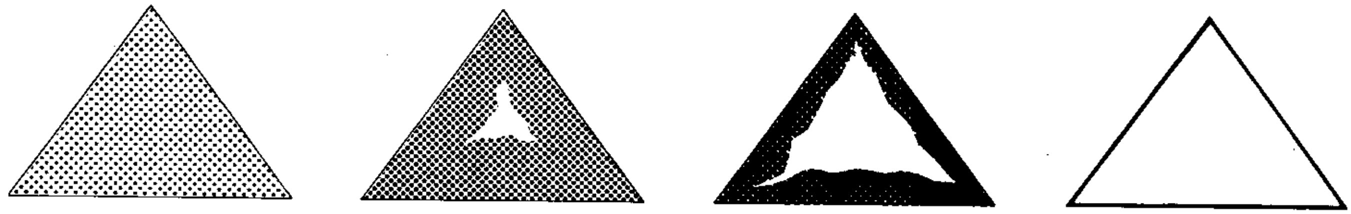

Figure 2. One way to demonstrate the existence of fixed points is to use the nonretraction theorem. The set of points represented by the colored dots is being mapped continuously into itself, with each point not on the boundary going to the boundary, and each boundary point remaining fixed. This mapping, called a retraction, cannot exist, however, because it tears the set apart and thus becomes discontinuous. Figure 3 shows why this theorem proves that fixed points must exist.

The original proof of Brouwer's theorem was nonconstructive, in the sense that while it asserted the existence of a solution to a system of equations, it gave no indication of an effective numerical procedure for calculating the solution.

The general equilibrium model, as is widely realized, is far from a perfect representation of the way in which the economy functions. It relies more heavily than is warranted on the assumption of purely competitive markets, not only for current transactions, but also for those expected to arise in the future. Not only are apples and oranges exchanged today, but the model assumes the existence of markets that permit the exchange of apples today for oranges to be delivered five years in the future. Moreover, the model does not allow for the possibility of economies of scale in production, and the resulting tendency to monopolistic behavior. A good deal of modern economic theorizing may, in fact, be seen as an attempt to replace the abstract simplifications of the general equilibrium model with more realistic alternatives.

In spite of its deficiencies, however, the general equilibrium model, as illustrated in the gray box on the facing page, provides an extremely useful technique for analyzing the way in which the economy might respond to modifications in economic policy or to changes in the economic environment. For example, the consequences of a substantial increase in the price of imported oil may be examined by constructing a general equilibrium model of the domestic economy whose solution, prior to the price increase, is consistent with prices, levels of output, and the distribution of income previously observed. After modifying the parameter of the model that represents the price of oil, a new equilibrium is calculated in order to assess the consequences of the price increase for whatever variables are considered to be of significance. In a similar fashion, the consequences of a tariff on particular imports, or of a change in the personal or corporate income tax schedules, may be examined by the numerical solution of a general equilibrium model before and after the change is imposed.

Exercises like these comprise what is known in economic theorizing as "comparative statics." Traditionally, such exercises have been carried out in analytic form when the changes in question are extremely small, or by means of elementary geometric diagrams when the number of variables is quite limited. But if either of these conditions is not satisfied, the analysis can be done only by the explicit numerical solution of the equilibrium model. Given the generality of the supply and demand equations underlying the model, this requires the development of numerical algorithms for computing the fixed points implied by Brouwer's theorem, rather than the mere assertion of their existence.

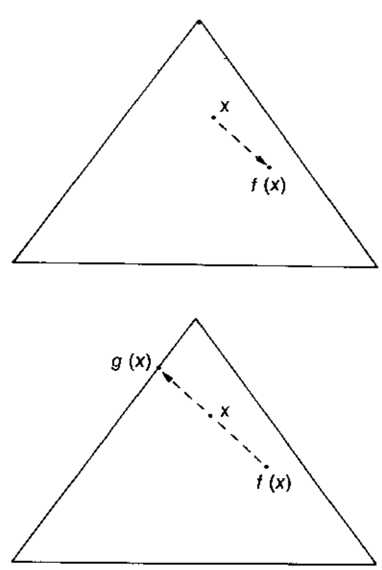

Figure 3. If there are no fixed points, then each *x* must move to an *f*(*x*) that is different from *x* (*top*). This means that a second continuous mapping can exist that moves the point to *g*(*x*) on the boundary (*bottom*). If *x* is on the boundary, *g*(*x*) will be the same point as *x*. But this second mapping is a retraction, which cannot exist, and our original premise must be false — thus, fixed points do exist.

Brouwer's theorem

In order to set the stage for Brouwer's theorem in a form that leads to an effective computational procedure, let us consider the particular n-dimensional set called the simplex, whose points are given by $x = (x_1, x_2, \ldots, x_{n+1})$, with coordinates ≥ 0 and that sum to unity. By a mapping of this set into itself we simply mean a function $f(x)$ that associates with each such x another point in the set, as shown in Figure 1.

The mapping is called continuous if each coordinate of the point to which x is mapped—the image of x—varies continuously with x: informally, small changes in x yield small changes in each coordinate of the image of x. By a fixed point of the mapping we mean a particular point x' that is mapped into itself—i.e., a point for which $f(x') = x'$. Brouwer's theorem asserts the existence of at least one such fixed point.

Before discussing a proof of Brouwer's theorem that is oriented toward calculating fixed points, let us make a brief digression and exhibit a preliminary argument based on what is known as the nonretraction theorem. A retraction is a continuous mapping of the simplex into itself with the following two additional properties: every point in the simplex is mapped onto the boundary of the simplex, and every boundary point of the simplex is mapped to itself.

Figure 2 will help the reader imagine a retraction. Any continuous mapping of the simplex into itself can be thought of as the end result of a continuous deformation that starts with the original simplex and continuously transforms it until each point reaches its image. But if the mapping is a retraction—carrying

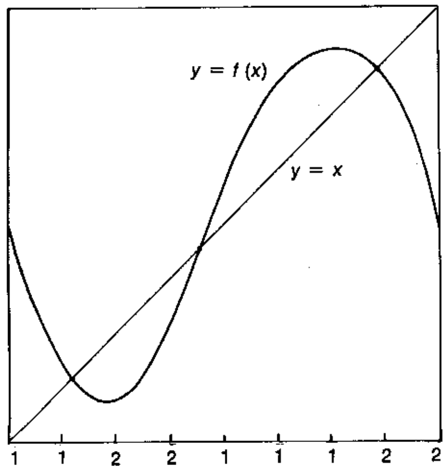

Figure 4. The mapping of the set of those points x between 0 and 1 into itself is represented by $y = f(x)$. The x axis represents the set before mapping; this mapping moves each point on the x axis up until it hits $y = f(x)$ and then over to the y axis, which represents the set after mapping. In this mapping there are three points (the colored dots) that are mapped to themselves. Dividing the set of points into subintervals with labeled ends provides a way of locating the fixed points: if the label is 1 when $f(x) \geq x$ and 2 when $f(x) < x$, each subinterval whose ends have different labels will contain a fixed point.

every point in the simplex to the boundary, and leaving the boundary unchanged—such a process would have to tear the simplex apart at some point, and the mapping would thus be discontinuous. This intuitive argument is the basic idea behind the nonretraction theorem, which states that there can be no retraction of the simplex.

To see that Brouwer's theorem follows from the nonretraction theorem, let us reason by contradiction:

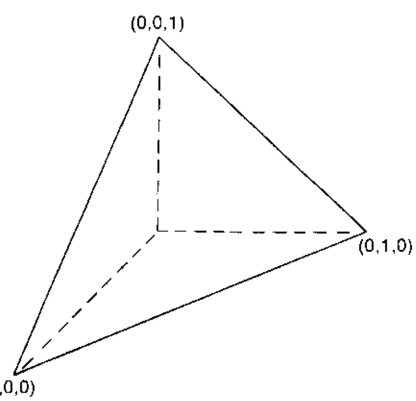

Figure 5. The location of fixed points in a simplex – or the particular set of points with coördinates ≥ 0 and that sum to unity – can be determined by subdividing the simplex. This two-dimensional simplex is represented here in three-dimensional space.

Assume that there is a continuous mapping $x \rightarrow f(x)$ of the simplex into itself with no fixed points. Then for every x, the two points x and $f(x)$ are distinct from each other. This assumption will permit us to construct a new mapping, g, which will be a continuous retraction. We begin by drawing the line segment shown in Figure 3, which starts at $f(x)$, passes through x, and terminates at a boundary point of the simplex that we will call $g(x)$. This new mapping $x \rightarrow g(x)$ is obviously a retraction, and, assuming the nonretraction theorem to be true, we have arrived at a contradiction.

We shall see the relevance of this digression in a bit, but our goal now is to find an argument for Brouwer's theorem that stands by itself rather than depending on an alternative theorem, such as the nonretraction theorem, whose proof is also by no means obvious. Let us turn to a different source of possible ideas by examining the elementary proof of Brouwer's theorem when $n = 1$. In this case the simplex is one-dimensional, and can be identified with the interval consisting of those points x with $0 \leq x \leq 1$. A continuous mapping of the interval into itself is described by a continuous function $y = f(x)$, such that for all x in the interval, $f(x)$ is also between 0 and 1. A fixed point of such a mapping is given by the intersection of this curve and the 45° line representing $y = x$: in Figure 4 there are three such fixed points.

Let us be very precise about an argument for the obviously true statement that a continuous curve defined for all x between 0 and 1 and contained in the square must intersect the 45° line at least once. Begin by dividing the interval into a large number of nonoverlapping subintervals. As in Figure 4, let us associate with each end point of a subinterval x a label that takes on the value of 1 or 2 according to the following rule: The label associated with x is 1 if $f(x) \geq x$ and is 2 if $f(x) < x$. Notice that the label associated with the lower end point of the interval is certainly 1 and, unless the mapping has a fixed point at the upper end of the interval, the label associated with the upper end is 2. The labels for the end points of the subintervals depend on the particular mapping, but regardless of their values there must be at least one subinterval whose two end points are differently labeled. If the subintervals are sufficiently small, then any point x in a subinterval whose end points are differently labeled will serve as an approximate fixed point of the mapping, in the sense that $|f(x) - x|$ will be small.

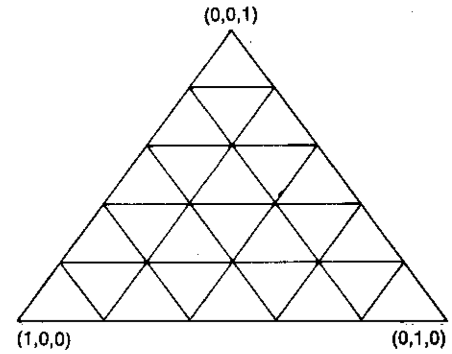

Figure 6. To find the fixed points in the simplex, we follow the same procedure used with the interval in Figure 4. The first step is to divide the set of points into subsimplices.

Demonstrating the existence of a point for which $f(x)$ is actually equal to x requires us to take a sequence of finer and finer divisions, and to find an approximate fixed point for each such division. We then select a convergent subsequence of approximate fixed points, which, because of the continuity of the mapping, tend to a true fixed point. In practice, however, such a construction is never required, and it is sufficient to find a point x that is close to its own image, or $f(x)$.

The basic ideas involved in generalizing this argument to higher values of n can be illustrated most clearly when $n = 2$ and the simplex

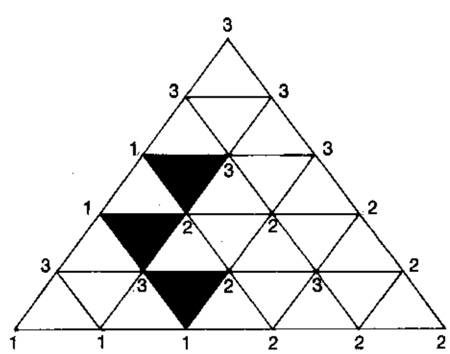

Figure 7. The vertices of the subsimplices are labeled arbitrarily, except that the labels of the vertices on the boundary of the simplex depend on the vertices' coordinates, as explained in the text. The colored subsimplices are those with different labels for each vertex, which means that if the labels are based on the mapping they will contain good approximations of fixed points.

consists of those points $x = (x_1, x_2, x_3)$ in three-dimensional space with $x_1 + x_2 + x_3 = 1$, and $x_1, x_2, x_3 \geq 0$. As shown in Figure 5, the simplex has three vertices with coordinates (1,0,0), (0,1,0), and (0,0,1), and three boundary faces, each of which is opposite one of the vertices. Our first step is to divide the simplex into a large number of small subsimplices, as illustrated in Figure 6.

Now let us imagine that each vertex of each subsimplex has associated with it a label that is either 1, 2, or 3. When we apply this approach to a proof of Brouwer's theorem, the labels associated with the vertices will depend on the particular mapping of the simplex into itself under consideration. For the moment, however, the labels will be quite arbitrary except that: A vertex on the boundary of the simplex will receive a label i only if the i^{th} coordinate of the vertex is positive.

According to this rule, the vertices (1,0,0), (0,1,0), and (0,0,1) must receive the labels 1, 2, and 3 respectively. Any vertex on the face on which $x_1 = 0$ must receive a label that is either 2 or 3, a vertex on the adjoining face on which $x_2 = 0$ must be labeled either 1 or 3, and similarly for the remaining face. Figure 7 illustrates an assignment of labels that is consistent with this requirement.

Sperner's lemma

The remarkable combinatorial theorem first enunciated by Emanuel Sperner in his doctoral thesis written in 1928 states that at least one subsimplex must have three distinct labels associated with its three vertices. In Figure 7 there are three such "completely labeled" triangles. (The force of Sperner's lemma can best be understood by experimenting with different assignments of labels in an attempt to avoid a completely labeled triangle.)

Assuming for the moment that Sperner's lemma is correct, let us now consider an arbitrary continuous mapping $(x_1, x_2, x_3) \rightarrow [f_1(x), f_2(x), f_3(x)]$ of the simplex into itself, and see how to apply the lemma in order to demonstrate the existence of a fixed point. Since the mapping carries the simplex into itself, any x on the simplex that is not a fixed point must have at least one of its coordinates decreasing under the mapping. For any x that is a vertex of a subsimplex and is not a fixed point, let the label associated with x be a subscript i for which $f_i(x) < x_i$—i.e., a coordinate that is decreasing under the mapping.

This labeling is certainly consistent with the requirement of Sperner's lemma that each vertex on the boundary receives a label i only if the i^{th} coordinate of the vertex is positive. Sperner's lemma can therefore be applied, and we conclude that there is at least one subsimplex all of whose vertices have different labels.

If the subsimplices are very small, the three vertices of such a completely labeled triangle will be very close to each other. At each vertex a different coordinate will be decreasing under the mapping. It is quite elementary to argue, using the continuity of the mapping, that any point x in such a completely labeled triangle will be an approximate fixed point in the sense that x will be close to its image, $f(x)$. As before, a nonconstructive limiting argument is required to verify the existence of an actual fixed point rather than an approximate one, but this is never required in practice.

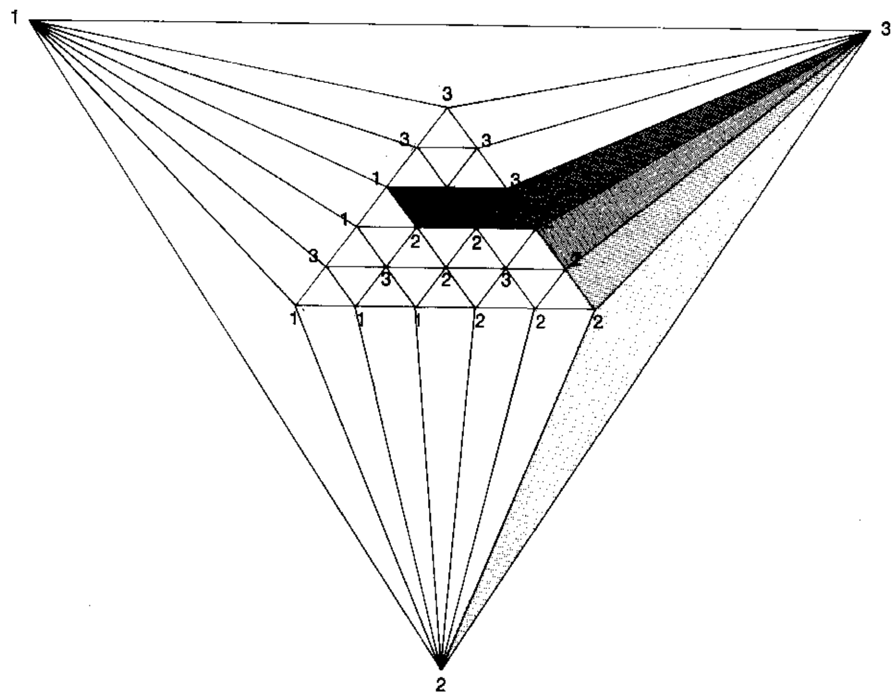

Figure 8. Completely labeled subsimplices, or those with a different label for each vertex, can be found by embedding the divided simplex of Figure 7 in a larger simplex. We start with the lightest colored triangle, which has two rather than three different labels. The path through adjoining triangles that also have the labels 2 and 3 must end at a completely labeled triangle for reasons explained in the gray box on this page.

Lemke's house

Let us imagine a house with a large—but finite—number of rooms all located on a single floor. Each room is assumed to have precisely two doors that lead either to another room or to the outside of the house. If there is one room with a door to the outside, there must be at least one other door to the outside.

To understand why, imagine that we enter the house through the known door to the outside and continue from room to room, always departing from a room through the door not used in entering that room. The trip may take us through many different rooms, but we will never enter the same room twice. Since the house has a finite number of rooms we must stop at some point, and this can occur only when we leave the house.

This elementary observation is the basis of Lemke's argument that our algorithm for Sperner's lemma terminates with a completely labeled triangle. In Figure 8 the rooms consist of those triangles whose vertices bear the labels 2 and 3 with label 1 missing. Each such room has two doors—the sides bearing the labels 2 and 3. The initial room with a door to the outside is the initial triangle of the algorithm; the path leaves the house when it arrives at a completely labeled triangle.

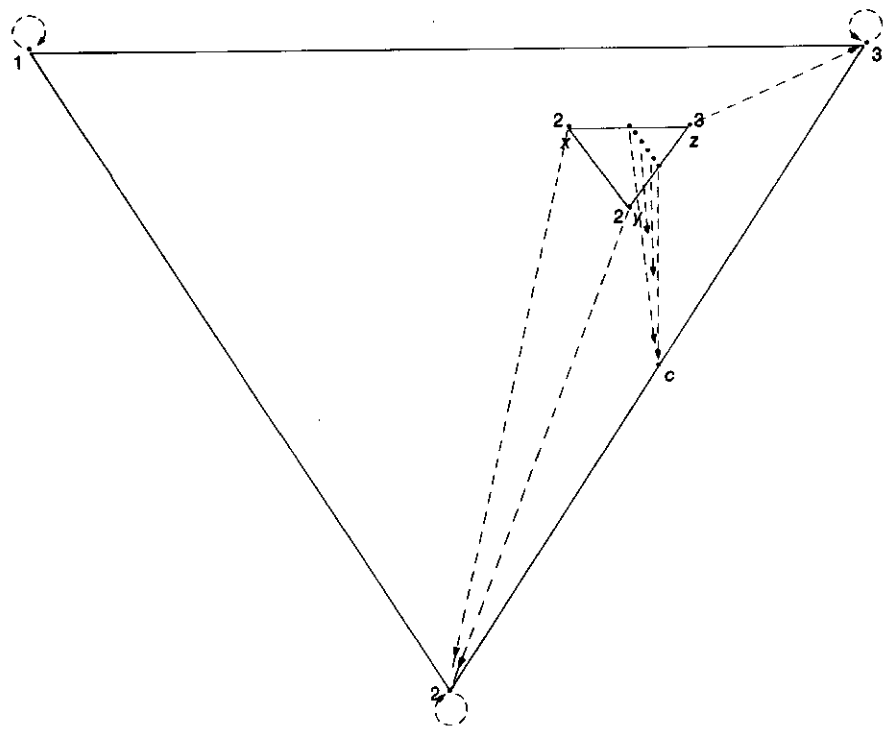

Figure 9. Our arbitrary assignment of labels can be used to construct a mapping that would contradict the nonretraction theorem if there were no completely labeled triangles.

Computing fixed points

Sperner's lemma is valid in higher dimensions as well, and can be used to provide a general proof for Brouwer's theorem. Sperner's original argument asserting the existence of a completely labeled subsimplex was an inductive demonstration based on the dimension of the simplex, and was not oriented toward finding a completely labeled subsimplex by an explicit computational procedure. But by the 1960s various aspects of economic theory had become strongly influenced by the development of linear programming, input-output analysis, and related topics in which the ability to solve concrete problems in an explicit numerical fashion was of considerable significance. Sperner's lemma was reexamined, and algorithms—of which the following was the first example—were provided that converted these nonconstructive arguments into computational techniques of practical significance (Scarf 1967).

We begin by embedding the simplex, and its subsimplices, in a larger simplex as shown in Figure 8. The larger simplex is divided by joining its three new vertices to the vertices lying on the boundary of the original simplex, and each of the new vertices is given a specific label 1, 2, or 3. Because of the labeling requirement on the boundary of the simplex assumed in Sperner's lemma, these new labels can be selected so as to create no additional completely labeled simplices.

The new construction shown in Figure 8 makes it very easy to find a triangle whose three vertices bear two of the three desired labels—for example, the rightmost triangle in the figure. Beginning with this triangle, we construct a path of adjacent triangles, each of which has vertices labeled 2 and 3. The path is uniquely determined by the initial triangle. When we enter a new triangle on the path, it is through an edge whose two vertices bear the labels 2 and 3; if the triangle is not completely labeled, there will be a unique other edge in the triangle whose vertices are also labeled 2 and 3, and we use this edge in order to exit into a new triangle.

The remarkable aspect of this algorithm, first pointed out by Lemke (1965) in a different context, is that it never returns to a triangle it has previously encountered, as explained in the gray box on page 293. The algorithm can, therefore, never be forced to exit from the large simplex, since this would require a return to the initial triangle. But more important, the algorithm must terminate—since there are a finite number of triangles—and termination can occur only when we reach a triangle whose three vertices bear the labels 1, 2, and 3. This argument demonstrates Sperner's lemma, and, subject to our ability to move through the sequence of triangles efficiently, the algorithm provides a numerical procedure for approximating fixed points of a continuous mapping.

This elementary algorithm has been tried on a large number of specific examples since its introduction some 15 years ago, and it performs rather well on problems of moderate size (say, up to 20 variables) in spite of some of its obvious drawbacks. One clear deficiency is that it is necessary to specify the particular division of the simplex before determining an approximate fixed point of the mapping. If the accuracy is poor because the subsimplices are not sufficiently small, the only available recourse is to refine the division and carry out the algorithm again—starting at the boundary and discarding the previous estimate. As we shall see, recent modifications of the basic ideas permit us to initiate the algorithm at any convenient estimate of the fixed point.

A second flaw in this first version of the algorithm is that the only information used when the mapping is evaluated at a particular point x is which coordinate is decreasing. Newer variants incorporate information about the magnitude of the changes as well as their signs, with a considerable improvement in performance (e.g., Scarf 1967; Merrill 1971; Eaves 1972).

In order to see the form these modifications of the algorithm might take, let us try to put Sperner's lemma in a more general setting by showing its relation to the nonretraction theorem. Our previous argument began with a mapping of the simplex into itself and then assigned to each vertex of the divided simplex a label that depended on the particular mapping. Let us now turn the argument completely on its head by beginning with an arbitrary assignment of labels to the vertices—consistent with the requirement of Sperner's lemma—and using these labels to construct a continuous mapping of the simplex into itself. As we shall see, this new mapping turns out to be a retraction of the simplex into itself if there are no completely labeled triangles. Sperner's lemma, then, is an elementary consequence of the nonretraction theorem.

More efficient algorithms

What is of greater consequence for us is the fact that the solution of a certain system of equations based on the new mapping will lead to a path of points on the simplex traversing precisely those triangles appearing in our algorithm for Sperner's lemma. When this observation is generalized, it will lead directly to a class of fixed-point algorithms that are vastly more efficient than the one previously described.

Consider the division of the triangle shown in Figure 9, which is special in the sense that the only vertices on the boundary are the three vertices of the triangle themselves, which are given the labels 1, 2, and 3. Our purpose is to construct a mapping $x \rightarrow h(x)$ of this simplex into itself, based on the labels that are associated with the vertices of the subdivision. We begin by mapping each vertex of the subdivision into that one of the three boundary vertices that has the same label. In Figure 9, x and y are mapped to the boundary vertex labeled 2, and z to the boundary vertex labeled 3. We then complete the mapping by making it linear in each simplex: for example, the point halfway between y and z, the point halfway between x and z, and all the points between these two points are mapped to the same boundary point c.

This new mapping $x \rightarrow h(x)$ obviously has the property that every boundary point of the large triangle is mapped into itself. Moreover, a point contained in a subsimplex whose vertices are missing at least one label will certainly be mapped to the boundary of the simplex. If there are no completely labeled simplices, then every point is mapped to the boundary and $h(x)$ is a retraction. The validity of Sperner's lemma is therefore an immediate consequence of the nonretraction theorem.

This mapping can be put to another use. Let us return to the particular labeling in our previous example of Sperner's lemma (see Fig. 8) and examine the set of all points in the larger simplex that are mapped into the particular point c. As Figure 10 indicates, the set of points for which $h(x) = c$ includes those points on a path traversing precisely the same subsimplices as does our algorithm and terminating with a subsimplex all of whose labels are distinct.

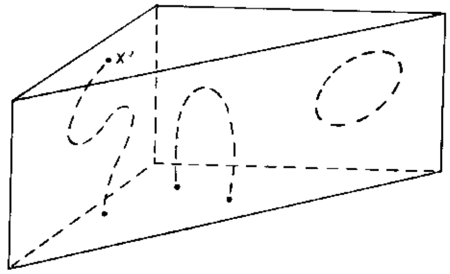

Figure 11. If all points on the top of the wedge are mapped into x' and the set of points on the bottom is mapped into itself, the whole wedge can be mapped into a two-dimensional simplex by interpolating between the first two mappings. A fixed point of the mapping on the bottom can be found by tracing a path of fixed points of the interpolation. However, if we started our algorithm in the middle of the wedge, we might encounter an area that seemed to have fixed points, represented by the closed loop, and be temporarily sidetracked.

But the inverse image of c has other parts as well. There is a path for which $h(x) = c$ joining the two other completely labeled subsimplices, and a loop of almost completely labeled simplices sitting off by itself. These observations can be put in a general context that is extremely useful in the construction of more effective algorithms than those based on Sperner's lemma. Consider the two-dimensional set obtained by eliminating the three completely labeled triangles. The mapping h takes this entire two-dimensional set into a one-dimensional set: the boundary of the simplex. It follows that the collection of points mapping into a particular boundary point should be one-dimensional. In general, this is correct and the inverse image of a single boundary point (with a few exceptional boundary points) can be shown to consist of a finite number of paths and loops (e.g., Eaves and Scarf 1976; Milnor 1965).

To see how this observation can be used, consider the wedge of points in Figure 11. Let us be given a continuous mapping $x \rightarrow f(x)$ on the set $t = 0$, whose fixed points we are interested in approximating. Select the point x' to be an arbitrary estimate of a fixed point $f(x)$, and define a new mapping, $g(x)$, that takes every point on the top of the wedge $t = 1$ into the same point x'. Then complete the mapping of the entire three-dimensional wedge into a two-dimensional simplex by interpolating linearly between these two mappings—i.e., by defining $F(t,x) = tg(x) + (1 - t)f(x)$, for t between 0 and 1.

By analogy with our previous argument, we would expect the collection of fixed points (i.e., points for which $F(t,x) = x$) to be a finite set of paths and loops. From the nature of our construction the set of fixed points can touch the boundary of the wedge only at the top or bottom face.

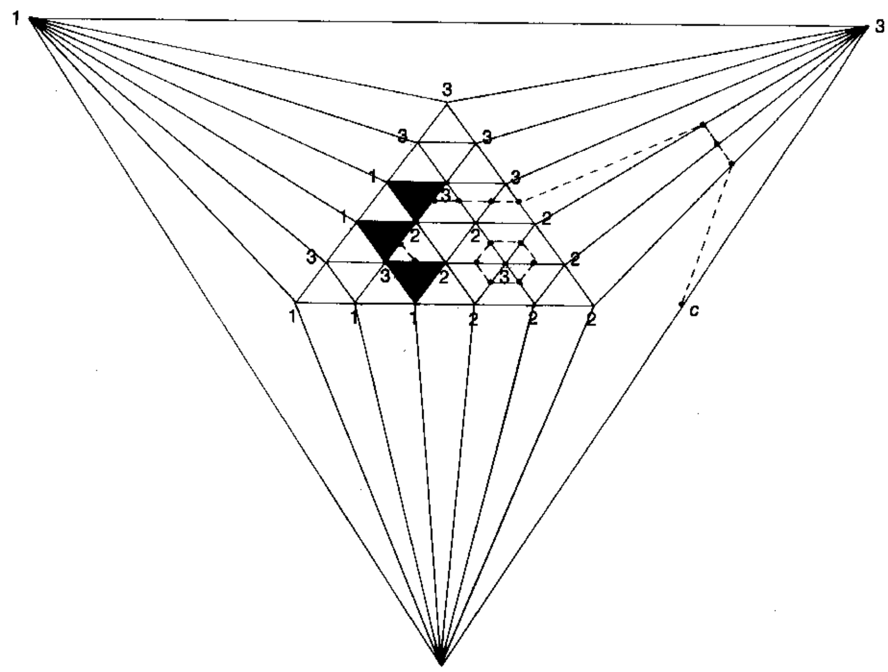

Figure 10. The points mapped into c, using the divided simplex in Figure 8, include those points on the path from the starting triangle to the completely labeled subsimplex. Also included are a path between the other two completely labeled subsimplices, and a path contained within a group of subsimplices that have only the labels 2 and 3.

Only the path starting at x' can intersect the top face, and if this particular path is followed to the bottom it must terminate in a fixed point of the mapping under consideration.

There are several numerical methods that allow us to follow such a path. The wedge can be divided into subsimplices (tetrahedra in this example), and the mapping $(t,x) \rightarrow F(t,x)$ replaced by a continuous approximation that is linear in each subsimplex. Calculating which subsimplices are on the path from $t = 1$ to $t = 0$ depends on the particular way in which the wedge is divided (see Merrill 1971, Eaves 1972, and van der Laan and Talman 1979 for extremely useful suggestions about appropriate divisions), but the mathematical operations are essentially those we have become familiar with in linear programming. And the resulting algorithms, which have been tried on thousands of examples ranging up to 60 variables, work extremely well in practice; in addition to being global in character, they are competitive in time of execution with Newton's method—the classical local method for solving systems of equations.

A second path-following method requires the underlying mapping on $t = 0$ to be differentiable. The path from $t = 1$ to $t = 0$ may then be shown, in general, to be a curve satisfying a differential equation virtually identical to that arising in a continuous variant of Newton's method. The differentiable approach, introduced by Kellogg, Li, and Yorke (1976) and by Smale (1976), leads to a global Newton's method with no requirement that the algorithm be initiated in the vicinity of the answer.

The availability of these numerical techniques has led, during the last decade, to the construction and solution of general equilibrium models designed to illustrate a variety of economic issues. Several authors have been concerned with the impact on the economies of the United States and other countries of changes in domestic taxes (e.g., Shoven and Whalley 1972), international negotiations to reduce tariffs and other barriers (e.g., Brown and Whalley 1980), and the ways in which individual taxes and subsidies compound and affect each other (e.g., Piggott and Whalley, in prep.). The gains to Britain in joining the Common Market have been examined (Miller and Spencer 1977), as well as the consequences for the United States of a variety of policies in response to increased energy prices (Hudson and Jorgenson 1974). Problems of international trade have been studied by means of general equilibrium models, one of which involved over 200 commodities (Ginsburg and Waelbroeck 1981).

Fixed-point methods of the sort we have discussed are guaranteed to provide the solution in these and other examples of applied general-equilibrium analysis. It is true that simpler methods may be successful, either by accident or by the intelligent use of some particular feature of the problem. But the existence of global methods that work in all cases has transformed the general equilibrium model from an abstract argument for the consistency of economic reasoning to an effective tool for the analysis of economic policy.

References

Arrow, K. J., and G. Debreu. 1954. Existence of an equilibrium for a competitive economy. *Econometrica* 22:265–90.

Brouwer, L. E. J. 1912. Über Abbildungen von Mannigfaltigkeiten. *Mathematische Annalen* 71:97–115.

Brown, F., and J. Whalley. 1980. General equilibrium evaluations of tariff cutting proposals in the Tokyo round and comparisons with more extensive liberalization of world trade. *Econ. J.* 90:838–66.

Eaves, B. C. 1972. Homotopies for the computation of fixed points. *Math. Programming* 3:1–22.

Eaves, B. C., and H. Scarf. 1976. The solution of systems of piecewise linear equations. *Math. of Operations Res.* 1:1–27.

Ginsburg, V. A., and J. L. Waelbroeck. 1981. *Activity Analysis and General Equilibrium Modelling*. North-Holland.

Hansen, T. 1968. On the approximation of a competitive equilibrium. Ph.D. diss., Yale Univ.

Hudson, E. A., and D. W. Jorgenson. 1974. U.S. energy policy and economic growth. *Bell J. Econ. and Mgt. Sci.* 5:461–514.

Kellogg, R. B., T. Y. Li, and J. Yorke. 1976. A constructive proof of the Brouwer fixed point theorem and computational results. *SIAM J. Numerical Anal.* 13:473–83.

Kuhn, H. W. 1968. Simplicial approximation of fixed points. *PNAS* 61:1238–42.

Lemke, C. E. 1965. Bimatrix equilibrium points and mathematical programming. *Mgt. Sci.* 11:681–89.

Merrill, O. H. 1971. Applications and extensions of an algorithm that computes fixed points of certain non-empty convex upper semicontinuous point to set mappings. Technical Report 71-7, Univ. of Michigan.

Miller, M. H., and J. E. Spencer. 1977. The static economic effects of the U.K. joining the E.E.C.: A general equilibrium approach. *Rev. Econ. Stud.* 44:71–94.

Milnor, J. 1965. *Topology from the Differentiable Viewpoint*. Charlottesville: Univ. of Virginia Press.

Piggott, J., and J. Whalley. In prep. Economic effects of U.K. tax-subsidy policies: A general equilibrium appraisal.

Scarf, H. E. 1967. The approximation of fixed points of a continuous mapping. *SIAM J. App. Math.* 15:1328–43.

Scarf, H. E., with the collaboration of T. Hansen. 1973. *The Computation of Economic Equilibria*. Yale Univ. Press.

Shoven, J. B., and J. Whalley. 1972. A general equilibrium calculation of the effects of differential taxation of income from capital in the U.S. *J. Public Econ.* 1:281–321.

Smale, S. 1976. A convergent process of price adjustment and global Newton methods. *J. Math. Econ.* 3:107–20.

Sperner, E. 1928. Neur Beweis für die Invarianz der Dimensionszahl und des Gebietes. *Abh. a.d. Math. Sem. d. Univ. Hamburg* 6: 265–72.

van der Laan, G., and A. J. J. Talman. 1979. A restart algorithm for computing fixed points without an extra dimension. *Math. Programming* 17:74–84.

von Neumann, J. 1937. Über ein ökonomisches Gleichungsystem und eine Verallgemeinerung des Brouwersche Fixpunktsatzes. *Ergebnisse eines mathematischen Kolloquiums* 8:73–83.

Walras, L. 1874. *Elḗments d'Economie Politique Pure*. Lausanne: Corbaz.