1
Sign-solvability

1.1 A problem in economics

Qualitative economics is usually considered to have originated with the work of Samuelson [11, Chapter III] who discussed the possibility of determining unambiguously the qualitative behavior of solution values of a system of equations. In his pioneering paper [6] (see also [4] and [7, 8]) Lancaster put it this way: *Economists believed for a very long time, and most economists would still hope it to be so, that a considerable body of sensible economic propositions could be expressed in a qualitative way, that is, in a form in which the algebraic sign of some effect is predicted from a knowledge of the signs, only, of the relevant structural parameters of the system.*

Consider the following example, similar to one discussed in Samuelson [11], of a market for a product, say bananas, where the price and quantity are determined by the intersection of its supply and demand curves. We introduce a *shift* parameter α into the demand curve, and assume that an increase in α shifts the demand curve upward and to the right. For instance, α might represent people's taste for bananas, and as people's taste for bananas increases so does their demand for bananas. Let $S(p)$ denote the number x of bananas that farmers will produce if the price per banana is p. Simple economic principles tell us that as the price p increases farmers will supply more bananas. This gives a supply curve as indicated in Figure 1.1.

Let $D(p, \alpha)$ denote the number x of bananas that consumers will demand if the price per banana is p and people's taste for bananas is α. Again simple economic principles tell us that for α fixed, the demand for bananas decreases as the price p increases. For p fixed, as people's taste α for bananas increases so does their demand for bananas. This gives a family of demand curves as indicated in Figure 1.2.

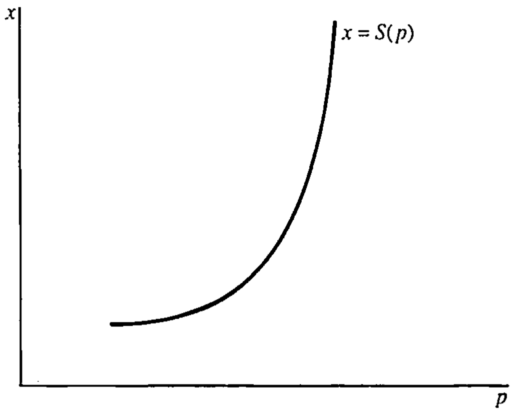

Fig. 1.1.

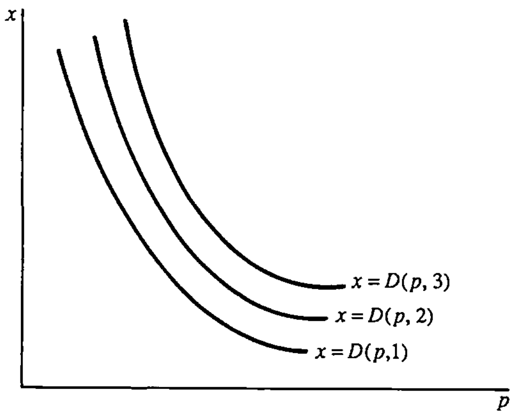

Fig. 1.2.

The equilibrium equations, where supply equals demand, are

$$\begin{aligned} S(p) - x &= 0 \\ D(p, \alpha) - x &= 0. \end{aligned} \tag{1.1}$$

The equilibrium points (p_α, x_α) are pictured in Figure 1.3. This figure suggests that as α increases so do p_α and x_α. That this is indeed the case can be justified mathematically as follows.

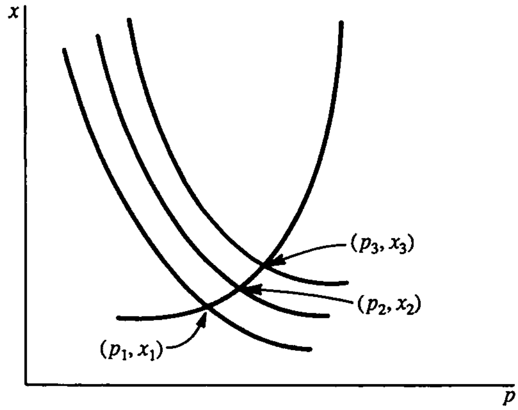

Fig. 1.3.

Taking partial derivatives with respect to α in (1.1), we obtain

$$\begin{aligned} \frac{\partial S}{\partial p}\frac{\partial p}{\partial \alpha} - \frac{\partial x}{\partial \alpha} &= 0 \\ \frac{\partial D}{\partial p}\frac{\partial p}{\partial \alpha} + \frac{\partial D}{\partial \alpha} - \frac{\partial x}{\partial \alpha} &= 0. \end{aligned} \tag{1.2}$$

Equivalently,

$$\begin{bmatrix} \frac{\partial S}{\partial p} & -1 \\ \frac{\partial D}{\partial p} & -1 \end{bmatrix} \begin{bmatrix} \frac{\partial p}{\partial \alpha} \\ \frac{\partial x}{\partial \alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{\partial D}{\partial \alpha} \end{bmatrix}. \tag{1.3}$$

The conclusions just given, which were based on simple economic principles, are equivalent to

$$\frac{\partial S}{\partial p} > 0, \quad \frac{\partial D}{\partial p} < 0, \text{ and } \frac{\partial D}{\partial \alpha} > 0. \tag{1.4}$$

In (1.3) we replace by their signs those quantities whose signs are determined by (1.4) and obtain

$$\begin{bmatrix} + & - \\ - & - \end{bmatrix} \begin{bmatrix} \frac{\partial p}{\partial \alpha} \\ \frac{\partial x}{\partial \alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ - \end{bmatrix}. \tag{1.5}$$

Every matrix with the same sign pattern as the 2 by 2 matrix in (1.5) has a negative determinant and hence is invertible. It follows by inspection (or use Cramer's rule) that

$$\frac{\partial p}{\partial \alpha} > 0 \text{ and } \frac{\partial x}{\partial \alpha} > 0$$

independent of the magnitudes of the quantities in (1.4). We conclude that p_α and x_α are increasing functions of α.

The preceding example is a special case of a more general situation. Let $x_1, x_2, \ldots, x_n$ and α be $n+1$ variables satisfying the n functional relationships

$$f_i(x_1, \ldots, x_n, \alpha) = 0 \quad (i = 1, 2, \ldots, n). \tag{1.6}$$

If the directions of the rates of change of the f_i with respect to the x_j and α are known, can we determine the direction of the rates of change of the x_j with respect to α? Taking partial derivatives of the f_i with respect to α we obtain the linear system

$$A \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \tag{1.7}$$

where $A = [a_{ij}]$ is the matrix of order n with

$$a_{ij} = \frac{\partial f_i}{\partial x_j} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n),$$

and

$$\begin{aligned} u_j &= \frac{\partial x_j}{\partial \alpha} \quad (j = 1, 2, \ldots, n) \\ c_i &= -\frac{\partial f_i}{\partial \alpha} \quad (i = 1, 2, \ldots, n). \end{aligned}$$

Our question is equivalent to: Can we solve for the signs of the u_j knowing only the signs of the a_{ij} and the c_i? This is the origin of the study of sign-solvable linear systems that we discuss in more detail in the next section.

1.2 Sign-solvable linear systems

We define the *sign* of a real number a by

$$\operatorname{sign} a = \begin{cases} +1 & \text{if } a > 0 \\ 0 & \text{if } a = 0 \\ -1 & \text{if } a < 0. \end{cases}$$

The *sign pattern* of a real matrix A is the $(0, 1, -1)$-matrix obtained from A by replacing each entry by its sign. A real matrix A determines a *qualitative class* $\mathcal{Q}(A)$ consisting of all matrices with the same sign pattern as A. The *zero pattern* of A is the (0,1)-matrix obtained from A by replacing each nonzero entry by 1. A *nonnegative matrix* is a matrix whose sign pattern is a (0, 1)-matrix. A *positive matrix* is a matrix whose sign pattern contains only 1's. We denote a matrix each of whose entries equals 1 by J. If the matrix is m by n, then we also write $J_{m,n}$, and this is abbreviated to J_n if $m = n$.

Consider a system of m equations in n unknowns given by

$$Ax = b \tag{1.8}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

and

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

are real matrices. The linear system (1.8) is sign-solvable provided we can solve for the signs of the entries of x knowing only the signs of the entries of A and of b. More precisely, (1.8) is *sign-solvable* provided that for each matrix $\widetilde{A}$ in the qualitative class $Q(A)$ and for each matrix $\tilde{b}$ in the qualitative class $Q(b)$,

$$\widetilde{A}x = \tilde{b}$$

is solvable and

$$\{\tilde{x} : \text{ there exists } \widetilde{A} \in Q(A) \text{ and } \tilde{b} \in Q(b) \text{ with } \widetilde{A}\tilde{x} = \tilde{b}\} \tag{1.9}$$

is entirely contained in one qualitative class. If $Ax = b$ is sign-solvable, then (1.9) is called the *qualitative solution class* of $Ax = b$ and is denoted by $Q(Ax = b)$. Suppose z satisfies $Az = b$ and w is in $Q(z)$. There exists a nonnegative invertible diagonal matrix D such that $w = Dz$. Then w satisfies $(AD^{-1})w = b$ and AD^{-1} is in $Q(A)$. It follows that if $Ax = b$ is sign-solvable, then $Q(Ax = b)$ is the entire qualitative class $Q(z)$. We also note that if $Ax = b$ is sign-solvable and E is an invertible diagonal matrix of order n, then $(AE)x = b$ is sign-solvable and $Q((AE)x = b) = Q(E^{-1}z)$. In particular, by taking $E = -I_n$ we see that if $Ax = b$ is sign-solvable, then so is $Ax = -b$. The linear system

$$\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

from (1.5) is an example of a sign-solvable linear system.

Let A be an m by n matrix and let B be an m by p matrix. Then $AX = B$ is *sign-solvable* provided that for each $\widetilde{A}$ in $Q(A)$ and each $\widetilde{B}$ in $Q(B)$ there is an n by p matrix $\widetilde{X}$ such that $\widetilde{A}\widetilde{X} = \widetilde{B}$ and

$$\{\widetilde{X} : \text{ there exists } \widetilde{A} \in Q(A) \text{ and } \widetilde{B} \in Q(B) \text{ with } \widetilde{A}\widetilde{X} = \widetilde{B}\}$$

is contained in one qualitative class. Thus the sign-solvability of $AX = B$ is equivalent to the sign-solvability of $Ax = b$ for every column b of B.

Theorem 1.2.1 *The homogeneous linear system $Ax = 0$ is sign-solvable if and only if every matrix in the qualitative class $Q(A)$ has linearly independent columns.*

Proof First suppose that every matrix $\widetilde{A}$ in $Q(A)$ has linearly independent columns. Then the only solution to $\widetilde{A}x = 0$ is the trivial solution. Thus $Ax = 0$ is sign-solvable and $Q(Ax = 0) = \{0\}$. Now suppose that $Ax = 0$ is sign-solvable. Since 0 is a solution of $Ax = 0$, we conclude that $Q(Ax = 0) = \{0\}$ and hence that every matrix $\widetilde{A}$ in $Q(A)$ has linearly independent columns. □

A matrix is an *L-matrix* provided every matrix in its qualitative class has linearly independent rows. The number of rows in an L-matrix does not exceed the number of its columns. By Theorem 1.2.1, *the homogeneous linear system $Ax = 0$ is sign-solvable if and only if the transpose A^T of A is an L-matrix.*[1]

Corollary 1.2.2 *If the linear system $Ax = b$ is sign-solvable, then A^T is an L-matrix.*

Proof Suppose that $Ax = b$ is sign-solvable but A^T is not an L-matrix. By Theorem 1.2.1, there exist $\widetilde{A}$ in $Q(A)$ and $z \neq 0$ such that $\widetilde{A}z = 0$. Let $\tilde{x}$ be a solution of $\widetilde{A}x = b$. Then $\widetilde{A}(\tilde{x} + cz) = b$ for all real numbers c. We may choose c so that the sign patterns of $\tilde{x} + cz$ and $\tilde{x}$ are different and thus contradict the sign-solvability of $Ax = b$. □

As an example we show that the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix} \qquad (1.10)$$

[1] In the literature an L-matrix is often defined to be a matrix for which every matrix in its qualitative class has linearly independent *columns*. This is convenient for discussing the connections between L-matrices and sign-solvable systems, but we have found it more convenient in the general study of L-matrices to require, as we have, linearly independent rows.

is an L-matrix. Let $\widetilde{A}$ be a matrix in $\mathcal{Q}(A)$. The sign pattern of $\widetilde{A}$ implies that no row of $\widetilde{A}$ is a multiple of another row. Every 3 by 1 sign pattern of 1's and -1's is the sign pattern of some column of $\widetilde{A}$ or its negative. It follows that no nontrivial linear combination of the rows of $\widetilde{A}$ equals zero, and hence the rows of $\widetilde{A}$ are linearly independent.

If A is an invertible diagonal matrix, then $Ax = b$ is a sign-solvable linear system. We now show that the sign-solvability of $Ax = b$ for all b implies that the rows of A can be permuted to obtain an invertible diagonal matrix.

Theorem 1.2.3 *The linear system $Ax = b$ is sign-solvable for all b if and only if A is a square matrix and there exists a permutation matrix P such that PA is an invertible diagonal matrix.*

Proof Suppose that $Ax = b$ is sign-solvable for all b. Then the linear system $Ax = b$ has a unique solution for all b. Hence A is an invertible square matrix and $A^{-1}b$ is the unique solution. Assume that some row of A^{-1}, say the first row, contains nonzero entries c and d in columns j and k, respectively, where $j \neq k$. Choose b so that its entry in row j has the same sign as c, its entry in row k has the same sign as $-d$, and its remaining entries are 0. Then there exist $\tilde{b}$ and $\hat{b}$ in $\mathcal{Q}(b)$ such that the first entry of $A^{-1}\tilde{b}$ is positive and that of $A^{-1}\hat{b}$ is negative, contradicting the sign-solvability of $Ax = b$. Hence each row of A^{-1} contains a unique nonzero entry, implying that $A^{-1}Q$ is a diagonal matrix for some permutation matrix Q. Therefore $Q^{-1}A$ is an invertible diagonal matrix. The converse is obvious. □

If A is an L-matrix, then A^T is an L-matrix if and only if A is square. A square L-matrix is also called a *sign-nonsingular matrix*, abbreviated SNS-matrix. In the next theorem we use the determinant in order to obtain characterizations of SNS-matrices. These characterizations were already evident in the early papers [11, 6, 1]. A square matrix A has a *signed determinant* provided the determinants of the matrices in $\mathcal{Q}(A)$ all have the same sign. We recall that the *standard determinant expansion* of a matrix $A = [a_{ij}]$ of order n is

$$\det A = \sum_{\sigma} \operatorname{sgn}(\sigma) a_{1i_1} a_{2i_2} \cdots a_{ni_n} \tag{1.11}$$

where the summation extends over all permutations $\sigma = (i_1, i_2, \ldots, i_n)$ of $\{1, 2, \ldots, n\}$ and $\operatorname{sgn}(\sigma)$ denotes the sign of the permutation σ.

Lemma 1.2.4 *Let $A = [a_{ij}]$ be a matrix of order n. Then A has a signed determinant if and only if one of the following holds:*

(i) *Every term in the standard determinant expansion of A is zero.*

(ii) *There is a nonzero term in the standard determinant expansion of A and every such term has the same sign.*

Proof If either (i) or (ii) holds, then clearly A has a signed determinant. Now assume that A has a signed determinant and that (i) does not hold. Let $\sigma = (i_1, i_2, \ldots, i_n)$ be a permutation for which the corresponding term

$$t_\sigma = \operatorname{sgn}(\sigma) a_{1i_1} a_{2i_2} \cdots a_{ni_n}$$

in the determinant expansion of A is not zero. The sign of the determinant of the matrix in $\mathcal{Q}(A)$ obtained by multiplying each entry of A not occurring in t_σ by a positive number ϵ is the same as the sign of t_σ for ϵ sufficiently small. Since A has a signed determinant, (ii) holds. □

Theorem 1.2.5 *Let $A = [a_{ij}]$ be a matrix of order n. Then the following are equivalent:*

(i) *A is an SNS-matrix.*
(ii) $\det A \neq 0$ *and A has a signed determinant.*
(iii) *There is a nonzero term in the standard determinant expansion of A and every nonzero term has the same sign.*

Proof The equivalence of (i) and (ii) is a consequence of the facts that $\mathcal{Q}(A)$ is a connected set and the determinant is a continuous function. The equivalence of (ii) and (iii) is a consequence of Lemma 1.2.4. □

Let n be an integer with $n \geq 2$, and let

$$H_n = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ -1 & -1 & 1 & \cdots & 0 & 0 \\ -1 & -1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & -1 & 1 \\ -1 & -1 & -1 & \cdots & -1 & -1 \end{bmatrix} \qquad (1.12)$$

be the lower Hessenberg matrix with -1's on and below the main diagonal and 1's on the superdiagonal. It follows by induction on n and the Laplace expansion of the determinant by the first row that each nonzero term in the standard determinant expansion of H_n is $(-1)^n$. Hence by Theorem 1.2.5, H_n is an SNS-matrix for all $n \geq 2$. □

Let $A = [a_{ij}]$ be a matrix of order n. Then A has an *identically zero determinant* provided each of the $n!$ terms in the standard determinant expansion is 0.

It follows from Lemma 1.2.4 that A has an identically zero determinant if and only if the determinant of each matrix in $\mathcal{Q}(A)$ is zero. The Frobenius–König theorem [2] asserts that there is a nonzero term in the standard determinant expansion of A if and only if A does not have a p by q zero submatrix for any positive integers p and q with $p + q = n + 1$. Thus A has an identically zero determinant if and only if there exist positive integers p and q with $p+q = n+1$ such that A has a p by q zero submatrix. Theorem 1.2.5 implies that any matrix which is obtained from an SNS-matrix by replacing some of its nonzero entries with zeros is either an SNS-matrix or has an identically zero determinant. It also implies that if A is an SNS-matrix and $a_{ij} \neq 0$, then the matrix of order $n - 1$ obtained from A by deleting row i and column j is either an SNS-matrix or has an identically zero determinant. More generally, if A is an SNS-matrix and B is a square submatrix of A which does not have an identically zero determinant, then the matrix obtained from A by deleting the rows and columns which intersect B either is an SNS-matrix or has an identically zero determinant.

If C is an m by n matrix and u is an m by 1 column vector, then $C(i \leftarrow u)$ denotes the matrix obtained from C by replacing its ith column by u ($i = 1, 2, \ldots, n$).

Let A be a nonsingular matrix of order n and let b be an n by 1 column vector. Then Cramer's rule asserts that the unique solution $x = (x_1, x_2, \ldots, x_n)^T$ of $Ax = b$ satisfies

$$x_i = \frac{\det A(i \leftarrow b)}{\det A} \quad (i = 1, 2, \ldots, n).$$

In particular, the (i, j)-entry of A^{-1} equals

$$\frac{\det A(i \leftarrow e_j)}{\det A} = (-1)^{i+j}\frac{\det A_{j,i}}{\det A} \quad (i, j = 1, 2, \ldots, n)$$

where e_j is the n by 1 column vector whose only nonzero entry is a 1 in the jth position and $A_{j,i}$ is the matrix obtained by deleting row j and column i of A. We now obtain a Cramer-type rule for sign-solvable systems.

Theorem 1.2.6 *Let A be a matrix of order n and let b be an n by 1 column vector. Then $Ax = b$ is sign-solvable if and only if A is an SNS-matrix and each of the matrices $A(i \leftarrow b)$ is either an SNS-matrix or has an identically zero determinant.*

Proof First assume that $Ax = b$ is sign-solvable. By Corollary 1.2.2, A is an SNS-matrix. Let $\widetilde{A}$ and $\widetilde{b}$ be arbitrary matrices in $\mathcal{Q}(A)$ and $\mathcal{Q}(b)$, respectively.

By Cramer's rule, $\widetilde{A}x = \tilde{b}$ has a unique solution $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^T$ where

$$\tilde{x}_i = \frac{\det \widetilde{A}(i \leftarrow \tilde{b})}{\det \widetilde{A}}.$$

By Theorem 1.2.5, $\det \widetilde{A}$ and $\det A$ have the same sign. Hence, for each $i = 1, 2, \ldots, n$, the sign of $\det \widetilde{A}(i \leftarrow \tilde{b})$ is the same as the sign of $\det A(i \leftarrow b)$ for all $\widetilde{A}$ in $\mathcal{Q}(A)$ and all $\tilde{b}$ in $\mathcal{Q}(b)$. It follows that for each i, either $A(i \leftarrow b)$ has an identically zero determinant, or by Theorem 1.2.5 that $A(i \leftarrow b)$ is an SNS-matrix. The converse is an immediate consequence of Cramer's rule. □

Corollary 1.2.7 *If $Ax = b$ is a sign-solvable linear system where A is a square matrix and c is obtained from b by replacing some of its nonzero entries with zeros, then $Ax = c$ is sign-solvable.*

Recall that a square matrix A is an SNS-matrix if and only if $\widetilde{A}^{-1}$ exists for all $\widetilde{A}$ in $\mathcal{Q}(A)$. If A is an SNS-matrix, then the matrices $\widetilde{A}^{-1}$ need not have the same sign pattern. For example, the matrices

$$\begin{bmatrix} -1 & 1 & 0 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{bmatrix} \text{ and } \begin{bmatrix} -1 & 1 & 0 \\ -1 & -1 & 1 \\ -2 & -1 & -1 \end{bmatrix}$$

are SNS-matrices with the same sign pattern, but the signs of the entries in the (3,1) position of their inverses are different. An SNS-matrix A such that the matrices in $\{\widetilde{A}^{-1} : \widetilde{A} \in \mathcal{Q}(A)\}$ have the same sign pattern is called a *strong SNS-matrix*, which we abbreviate to S^2NS-matrix. It follows that a matrix A of order n is an S^2NS-matrix if and only if $AX = I_n$ is sign-solvable where I_n is the identity matrix of order n. Any matrix of the form $A = PD$ where P is a permutation matrix and D is an invertible diagonal matrix is an example of an S^2NS-matrix.

The following corollary is an immediate consequence of Cramer's rule.

Corollary 1.2.8 *Let $A = [a_{ij}]$ be a matrix of order n. Then A is an S^2NS-matrix if and only if*

(i) *A is an SNS-matrix, and*

(ii) *for each i and j with $a_{ij} = 0$, the submatrix $A_{i,j}$ of A of order $n - 1$ obtained by deleting row i and column j is either an SNS-matrix or has an identically zero determinant.*