Technical Report 70-137
NGL-21-002-008

September 1970

Applications of Partial Orderings to the Study of Positive Definiteness, Monotonicity, and Convergence of Iterative Methods for Linear Systems

by

James S. Vandergraft

UNIVERSITY OF MARYLAND
COMPUTER SCIENCE CENTER
COLLEGE PARK, MARYLAND

Technical Report 70-137
NGL-21-002-008

September 1970

Applications of Partial Orderings to the Study of Positive Definiteness, Monotonicity, and Convergence of Iterative Methods for Linear Systems

by

James S. Vandergraft

The research was supported in part by Grant NGL-21-002-008 from the National Aeronautics and Space Administration to the Computer Science Center of the University of Maryland.

Applications of Partial Orderings to
the Study of Position Definiteness, Monotonicity,
and Convergence of Iterative Methods for Linear Systems

1. Introduction

Consider a system of linear equations

$$AX = b \tag{1.1}$$

where A is a real nonsingular $n \times n$ matrix, X and b are elements of real Euclidean n-space, E^n . Most of the theorems which guarantee that the sequence $\{X_k\}$, defined by an iteration such as

$$X_{k+1} = BX_k + c \quad , \quad k = 0,1,\ldots \tag{1.2}$$

converges to the solution of (1.1) require A to be positive definite, or else the inverse of A must be nonnegative:

$$A^{-1} \geqslant 0 \ . \tag{1.3}$$

(A matrix which satisfies this latter condition is said to be monotone.) For example, a theorem of Reich (1949) says that if A is symmetric then the Gauss-Seidel method converges if and only if A is positive definite. For nonsymmetric matrices, the theory of M-matrices (see, for example, Varga (1962)) shows that if

$$a_{ij} \leqslant 0 \quad \text{for} \quad i \neq j \tag{1.4}$$

and (1.3) holds, then both the Jacobi and Gauss-Seidel methods converge. Finally, the theory of regular splittings, also discussed by Varga (1962), provides a rather general technique for obtaining iterative methods which are known to converge when applied to monotone matrices.

The theory of monotone matrices has received much attention, independent of its connection to convergent iterations. Bramble and Hubbard (1964), Bramble, Hubbard, and Thomée (1969), Price (1968), and others, have used properties of monotone matrices to obtain error bounds for discrete approximations to partial differential equations. For applications such as these, it is important to find conditions which are readily verified and which imply monotonicity. In this context, the theory of Stieltjes matrices, and results of Fan (1958) and Fiedler and Pták (1966) are of interest. Fan showed that (1.4) together with

$$(1.5) \qquad AX > 0 \quad \text{for some} \quad X \geqslant 0$$

implies that A is monotone, and Fiedler and Pták studied monotone matrices using a somewhat strengthened form of (1.5).

The purpose of this paper is to introduce a new concept, called K-semi positivity, which provides an important link between convergence theory, monotonicity, and positive definiteness. A necessary tool for this discussion is the theory of partial orderings, which are discussed briefly in the next section. In Section 3, K-semi positivity is defined and several fundamental facts are proved. The connections to positive definiteness and monotonicity are developed in Section 4, and the final section contains applications of these results to Jacobi's method and the theory of regular splittings.

2. Partial Ordering in E^n

The notation used here is essentially that of Vandergraft (1968). In particular, a cone in E^n will be a closed subset K which has a nonempty interior and satisfies $\alpha K \subset K$, $\alpha \geq 0$, $K + K \subset K$, and $K \cap \{-K\} = \{0\}$. The boundary of a cone K is denoted by δK, the interior by K^o. The partial ordering induced by K is denoted by $\geq^K$; that is, $X \geq^K Y$ means $X - Y \in K$, and $X >^K Y$ means $X - Y \in K^o$. If A is an $n \times n$ matrix, then A is called K-nonnegative $(A \geq^K 0)$ if $AX \in K$ for any $X \in K$, and A is K-positive $(A >^K 0)$ if $AX \in K^o$ for all $X \in K$, $X \neq 0$. Finally, A is K-monotone if $AX \in K$ implies $X \in K$. It is simple to prove that, if A is nonsingular, then A is K-monotone if and only if $A^{-1} \geq^K 0$.

Throughout this paper, results concerning K-nonnegative matrices will be used. Most of these results are direct extensions of the classical Perron-Frobenius theory of nonnegative matrices (see Gantmacher (1960)), and will not be restated here. There are, however, two rather special results, concerning K-nonnegative matrices, which will be of some use.

Lemma 2.1 If A is a nonsingular matrix, then $A \geq^K 0$ if and only if $X \in K^o$ implies $AX \in K^o$.

Proof Suppose $AX \in K^o$ for any $X \in K^o$. It suffices to show $AY \in K$ for any $Y \in \delta K$. But, if $AY \notin K$ for some $Y \in \delta K$, then since K is closed, there is a neighborhood S of Y with $A(S) \cap K = \emptyset$. But S contains points in K^o

so $A(S) \cap K \neq \emptyset$. This contradiction implies $AY \in K$. Conversely, suppose $A \geq^K 0$ but $AX \in \delta K$ for some $X \in K^o$. Using Lemma 2.1 of Vandergraft (1968) it follows that the set $S = \{Y: 0 \leq^K Y \leq X\}$ has the property that $A(S) \subset H$ where H is a subspace of dimension less than n . But using the fact that $X \in K^o$, it follows that for any $Z \in K$, $\alpha Z \in S$ for some $\alpha > 0$. Thus $\alpha AZ = A(\alpha Z) \in H$, and hence $AZ \in H$. But any $Y \in E^n$ can be written as $Y = Z_1 - Z_2$ where $Z_1, Z_2 \in K$. The above analysis shows that $AY \in H$, and hence A is singular.

The next result follows easily from Theorem 3.1 of Vandergraft (1968).

Lemma 2.2 If A is symmetric and positive definite, then there is a cone K with $A \geq^K 0$.

3. K-Semi Positive Matrices

Throughout this section, K will denote some fixed cone in E^n, and A is an $n \times n$ matrix. We begin with our basic definition, which is an obvious generalization of (1.5).

Definition The matrix A is called K-semi positive if $A(K^o) \cap K^o \neq \emptyset$.

If K is the usual cone of vectors with nonnegative components, then the class of K-semi positive matrices is identical with the class S defined by Fiedler and Pták (1966). The justification for introducing new terminology is two-fold. First, it is convenient to show explicitly the dependence on the cone K, and secondly, Lemma 2.1 shows that, for nonsingular matrices, K-nonnegativity is equivalent to $A(K^o) \subset K^o$. The above definition is merely a weakening of this condition. It is important to note, however, that unlike K-nonnegativity, the concept of K-semi positivity does not induce a partial ordering on the space of $n \times n$ matrices. For example, if $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, and K is the cone of vectors with nonnegative components, then both A and $-A$ are K-semi positive. Finally, it is clear that a condition which is equivalent to that of the definition is $A(K) \cap K^o \neq \emptyset$.

In the next lemma, we summarize some useful facts about nonsingular K-semi positive matrices.

Lemma 3.1 If A is nonsingular, then

i) A is K-semi positive if and only if A^{-1} is K-semi positive.

ii) If A is K-monotone, then A is K-semi positive.

The proof is a trivial application of the definition and will be omitted. Simple examples show that the converse of part ii) is not true.

We next prove a fundamental result connecting K-semi-positivity and convergence. Recall that a matrix A is <u>convergent</u> if the spectral radius $\rho(A)$ is less than 1; or equivalently, $\sum_{k=0}^{\infty} A^k$ converges.

<u>Theorem 3.1</u> If $A \leqslant^K I$ then A is K-semi-positive if and only if $I - A$ is convergent. (Equivalently, if $B \geqslant^K 0$ then B is convergent if and only if $I - B$ is K-semi-positive.)

<u>Proof</u> If A is K-semi-positive, then $AY \in K^o$ for some $Y \in K^o$. Let X be an eigenvector in K of $I - A$ corresponding to the eigenvalue $\rho = \rho(I-A)$, and let

$$t_0 = \sup\{t > 0 \;:\; tX \leqslant^K Y\}.$$

Since $Y \in K^o$, such a number t_0 exists, is positive and finite. Furthermore,

$$\rho\, t_0\, X = t_0(I-A)X \leqslant^K (I-A)Y = Y - AY <^K Y$$

hence $\rho t_0 < t_0$ and thus $\rho < 1$ which says that $I - A$ is convergent. Conversely, if $I - A$ is convergent, then the series $\sum_{k=0}^{\infty} (I-A)^k$ converges to A^{-1}, and since each term is K-nonnegative, it follows that the sum is also K-nonnegative. Thus $A^{-1} \geqslant^K 0$, so A is K-monotone, and by Lemma 3.1, A is K-semi-positive.

This proof actually shows that if $A \leqslant^K I$ and $I - A$ is convergent, then $A^{-1} \geqslant^K 0$. For K the cone of vectors with non-negative components, this was proven by Kuttler (1970).

The auxillary condition $A \leqslant^K I$, which appears in this theorem, is related to condition (1.4) $(a_{ij} \leqslant 0 \quad , \quad i \neq j)$. In fact, a direct generalization of (1.4) is

$$\alpha A \leqslant^K I \quad \text{for some} \quad \alpha > 0 \tag{3.1}$$

In Section 4, this condition will be discussed further.

We next give a spectral characterization of K-semi positivity.

<u>Theorem 3.2</u> If $A \leqslant^K I$ then the following statements are equivalent.

i) A is K-semi positive

ii) All eigenvalues of A have positive real part

iii) All real eigenvalues of A are positive.

<u>Proof</u> If $I - A \geqslant^K 0$ then $\rho(I-A)$ is an eigenvalue of $I - A$. Thus, if A has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ then $\rho(I-A) = 1 - \lambda_r$ where λ_r is real, and

$$1 - \lambda_r \geqslant |1-\lambda_i| \quad , \quad i = 1,2,\ldots,n \ . \tag{3.2}$$

This shows that the eigenvalues of A lie inside a circle, with center at 1, which passes through λ_r, where λ_r satisfies

$$\lambda_r \leqslant \lambda_i \quad , \quad \lambda_i \quad \text{real} \tag{3.3}$$

Now, if i) holds, then by Theorem 3.1, $I - A$ is convergent so $1 - \lambda_r < 1$. Hence $\lambda_r > 0$ and (3.2) implies ii) while (3.3) implies iii). Conversely, if either ii) or iii) holds, then $\lambda_r > 0$ and

$\rho(I-A) = 1 - \lambda_r < 1$ so $I - A$ is convergent. Again invoking Theorem 3.1, we conclude that i) is true.

Fan (1958) showed that if A satisfies (1.4) and is nonsingular, then $A^{-1} \geqslant 0$ if and only if all eigenvalues of A have positive real part. Hence the above theorem is an extension of Fan's result.

4. Monotonicity and Positive Definiteness

In this section we will investigate further the relationship between K-monotone, positive definite, and K-semi positive matrices. Observe, first, that Theorem 3.2 shows that, if A is symmetric, K-semi positive, and $A \leqslant^K I$, then A is positive definite. The converse of this is contained in our next theorem.

Theorem 4.1 If A is symmetric, then A is positive definite if and only if for some cone K, A is K-semi positive and $\alpha A \leqslant^K I$ for some $\alpha > 0$.

Proof If A is K-semi positive and $\alpha A \leqslant^K I$ then Theorem 3.2, applied to αA, shows that αA, hence A, is positive definite. Conversely, if A is positive definite, with $\rho(A) = \rho$, then for any $\alpha > 0$ such that $\alpha < 1/\rho$, $I - \alpha A$ is also positive definite. By Lemma 2.2, $I - \alpha A \geqslant^K 0$, for some cone K. Moreover, all eigenvalues of A have positive real part, so by Theorem 3.2, A is K-semi positive.

The number α in this theorem cannot, in general, be replaced by 1. To verify this, consider the matrix

$$A = \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix} \tag{4.1}$$

whose eigenvalues are $\frac{1}{2}$, $\frac{1}{2}$, 2 . Clearly, A is positive definite, but if there were a cone K with $A \leqslant^K I$, then $\rho(I-A)$ would have to be an eigenvalue of I - A , which certainly is not true. Thus, the condition " A is K-semi positive and $A \leqslant^K I$ " is somewhat stronger then positive definiteness.

We next consider K-monotone matrices and show how they are related to K-semi positive matrices.

<u>Theorem 4.2</u> If A is K-semi positive, and $\alpha A \leqslant^K I$ for some $\alpha > 0$, then A is nonsingular and K-monotone.

<u>Proof</u> Since $\alpha > 0$, αA is also K-semi positive, and Theorem 3.1 shows that $I - \alpha A$ is convergent. Furthermore, $0 \leqslant^K \Sigma(I-\alpha A)^K = (\alpha A)^{-1} = \alpha^{-1}A^{-1}$ hence $A^{-1} \geqslant^K 0$ and A is K-monotone.

The condition $\alpha A \leqslant^K I$, $\alpha > 0$, is a special form of

(4.2) $$BA \leqslant^K I \quad \text{for} \quad B \geqslant^K 0$$

Using this more general condition, we obtain:

<u>Theorem 4.3</u> Let A be nonsingular. Then A is K-monotone if and only if A is K-semi positive, and there exists a nonsingular $B \geqslant^K 0$ with $BA \leqslant^K I$.

<u>Proof</u> If A is K-monotone then A is K-semi positive, and $B = A^{-1}$ satisfies the conditions of the theorem. Conversely, if $B \geqslant^K 0$ and $BA \leqslant^K I$ then, by Theorem 4.2, $A^{-1}B^{-1} = (BA)^{-1} \geqslant^K 0$. Since $B \geqslant^K 0$, this implies $A^{-1} \geqslant^K 0$ which shows that A is K-monotone.

A simple rephrasing of a theorem of Price (1968) shows that A is monotone if and only if there is a nonsingular matrix B with $B \geqslant 0$, $BA \leqslant I$ and $I - BA$ convergent. This result can also be obtained from Theorem 4.3 together with Theorem 3.1.

Condition (4.2) has been used by Ortega and Rheinboldt (1967) in the study of iterative methods for nonlinear equations. In keeping with their terminology, we will call a matrix B which satisfies (4.2) a <u>K-positive left subinverse</u> of A . (Obviously the proof also holds if B is a right subinverse, $AB \overset{K}{\leqslant} I$.)

We conclude this section with the following summary of several of our results.

$$\boxed{\begin{array}{c} A \text{ is} \\ \text{positive} \\ \text{definite} \end{array}} \xleftrightarrow[A = A^T]{} \boxed{\begin{array}{c} A \text{ is K-semi positive,} \\ \text{and has a K-positive} \\ \text{subinverse} \end{array}} \longleftrightarrow \boxed{\begin{array}{c} A \text{ is} \\ \text{K-monotone} \end{array}}$$

5. Convergence Theorems

The results of preceding sections will now be used to prove some useful convergence theorems. We begin with a simple application to Jacobi's method.

Theorem 5.1 Let A be a matrix which has unit diagonal. If, for some cone K, A is K-semi positive and $A \leqslant^K I$, then the Jacobi method converges.

Proof The Jacobi iteration matrix is $J = I - A$. The hypotheses say that $J \geqslant^K 0$ and $I - J = A$ is K-semi positive, so by Theorem 3.1, J is convergent.

If A is symmetric, with unit diagonal, then the convergence of the Jacobi method implies that all eigenvalues of $I - A$ are less than 1, in modulus. Clearly, this implies that A is positive definite. Using Theorem 4.1 we conclude that, if the Jacobi method converges, then A is K-semi positive, and for some $\alpha > 0$, $\alpha A \leqslant^K I$. Note that this statement is nearly the converse of Theorem 5.1. The matrix

$$A = \begin{pmatrix} 1 & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & 1 & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & 1 \end{pmatrix}$$

whose eigenvalues are $\frac{3}{4}, \frac{3}{4}, \frac{3}{2}$, shows that we may not take $\alpha = 1$ in this statement. On the other hand, the matrix (4.1) shows that $\alpha A \leqslant^K I$ for some $\alpha > 0$, is not sufficient to prove Theorem 5.1.

If Reich's theorem, which was stated in the introduction, is reworded, using Theorem 4.1, we obtain a theorem for the Gauss-Seidel method which is quite similar to the above results for Jacobi's method.

Theorem (Reich) If A is symmetric, then Gauss-Seidel converges if and only if $\alpha A \leqslant^K I$ and A is K-semi positive, for some cone K, and some $\alpha > 0$.

We next turn to the theory of regular splittings. Following Varga (1962), we will say that $A = M - N$ is a K-regular splitting if $N \geqslant^K 0$, M is nonsingular, and $M^{-1} \geqslant^K 0$. For the case where K is the cone of nonnegative vectors, Varga proves that the iteration

$$X_{k+1} = M^{-1}NX_k + M^{-1}b \tag{5.1}$$

converges to the solution of $AX = b$, whenever $A^{-1} \geqslant 0$. The next theorem improves and extends this result.

Theorem 5.2 Let $A = M - N$ be a K-regular splitting. Then (5.1) converges if and only if A is K-semi positive.

Proof By definition, $I - M^{-1}A = M^{-1}N \geqslant 0$, hence if A is K-semi positive, then so is $M^{-1}A$, and Theorem 3.1 shows that $M^{-1}N$ is convergent. Conversely, if $M^{-1}N$ is convergent, then $0 \leqslant^K \Sigma(M^{-1}N)^k = (M^{-1}A)^{-1} = A^{-1}M$. But $M^{-1} \geqslant^K 0$ so $A^{-1} \geqslant^K 0$ and hence A is K-semi positive.

Using Theorem 4.2, we can replace the hypothesis $M^{-1} \geqslant^K 0$ by a somewhat simpler condition.

Corollary Let $A = M - N$ where $N \geqslant^K 0$, $\alpha M \leqslant^K I$ for some $\alpha < 0$, and M is K-semi positive. Then (5.1) converges if and only if A is K-semi positive.

We conclude by applying these results to a matrix derived by Bramble and Hubbard (1964). If a certain fourth order discretization is applied to a linear two-point boundary value problem, the resulting matrix, after dividing by diagonal elements, is

$$A = \begin{matrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \frac{-1}{2+h^2q(h)} & 1 & \frac{-1}{2+h^2q(h)} & 0 & 0 & \cdots \\ \frac{1}{24+12h^2q(2h)} & \frac{-4}{6+3h^2q(2h)} & 1 & \frac{-4}{6+3h^2q(2h)} & \frac{1}{24+12h^2q(2h)} & \\ 0 & \frac{1}{24+12h^2q(3h)} & \frac{-4}{6+3h^2q(3h)} & 1 & \cdots & \\ & & & \ddots & & \ddots \\ & & \ddots & & \ddots & \end{matrix}$$

where $h > 0$ is the mesh spacing, and $q(x) > 0$ is a coefficient in the differential equation. After a rather long and tedious analysis, Bramble and Hubbard show that A is monotone, and that the "backward-forward Gauss-Seidel method"

$$X_{k+1} = (I-U)^{-1}(I-L)^{-1}LUX_k + (I-U)^{-1}(I-L)^{-1}b \tag{5.2}$$

converges. (Here, we have written $A = I - L - U$ where L, U are lower and upper triangular, respectively.) Using the results of this section, we can simplify this analysis considerably. First of all, from the fact that, for small $h > 0$, the row sums for rows 3, 4, .., $h - 2$ are negative, one can construct a nonsingular matrix B with $B \geqslant 0$, $BA \leqslant I$. (Start with a matrix of all 1's. Modify the first few and last few rows, so that BA has nonpositive off-diagonal. Next modify further, so that B is nonsingular, and multiply by a small constant so that $BA \leqslant I$.) Next, if h is small enough, it is easily seen that $AX > 0$ where $X = (1,5,10,10^{10},10^{10^{10}}\ldots)^T$ so A is semi positive. Theorem 4.3 can now be applied to show that $A^{-1} \geqslant 0$, for small h.

To prove convergence of (5.2) consider the cone

$$K = \{(x_1,\ldots,x_n)^T : \sum_{i=1}^{n} x_i \geqslant 0 \ , \ \sum_{i=1}^{n} (-1)^i x_i \leqslant 0\}$$

A sufficient condition for $B \geqslant^K 0$ is that

$$\sum_{i=1}^{n} b_{ij} \geqslant 0 \ , \quad j = 1,2,\ldots,n \tag{5.3}$$

and either

$$\operatorname{sgn}\Big(\sum_{i=1}^{n} (-1)^i b_{ij}\Big) = (-1)^j \tag{5.4}$$

or

$$(-1)^{i+j} b_{ij} \geqslant 0 \tag{5.5}$$

Now, let A_0 be the matrix A, with $h = 0$, and write $A_0 = I - L - U$. Then the iteration (5.2) is of the form (5.1), where

$$N = LU$$

$$M = (I-L)(I-U)$$

By examining L and U, one sees that $-L$ and $-U$ both satisfy (5.3) and (5.5). Hence $-L \geqslant^K 0$, $-U \geqslant^K 0$ and $N = L \cdot U \geqslant^K 0$. Next, we note that for small α, $I - \alpha M$ has positive diagonal elements, which approach 1 as α becomes small, and the off-diagonal elements tend to zero as $\alpha \to 0$. Hence, for small α, $I - M$ satisfies (5.3) and (5.4), so $\alpha M \leqslant^K I$ for some $\alpha > 0$. Finally, if $X = (1,0,\ldots,0)^T$ then $X \in K$ and $A_0 X = MX = (1,-1/2,1/24,0,\ldots,0)^T \in K^0$, so A_0 and M are K-semi positive. But then, the hypotheses of the corollary to Theorem 5.2 are satisfied, and hence (5.1) converges. By continuity, (5.1) also converges for all h sufficiently small.

References

1. Bramble, J. H. and Hubbard, B. E. (1964), On a Finite Difference Analogue of an Elliptic Boundary Problem Which is Neither Diagonally Dominant Nor of Nonnegative Type, J. Math. Phys., XLIII, 117-132.
2. Bramble, J. H., Hubbard, B. E. and Thomée, V. (1969), Convergence Estimates for Essentially Positive Type Discrete Dirichlet Problems, Math. Comp. 23, 695-709.
3. Fan, Ky (1958), Topological Proofs for Certain Theorems on Matrices with Nonnegative Elements, Monatsk. Math. 62, 219-237.
4. Fiedler, M. and Pták, V. (1966), Some Generalizations of Positive Definiteness and Monotonicity, Num. Math. 9, 163-172.
5. Gantmacher, F. R. (1960), The Theory of Matrices, Chelsea, New York, Vol. II.
6. Kuttler, J. R. (1970), Characterization of Monotone Matrices (to appear).
7. Ortega, J. and Rheinboldt, W. (1967), Monotone Iterations for Nonlinear Equations with Applications to Gauss-Seidel Methods, SIAM J. Num. Anal. 4, 171-190.
8. Price, H. S. (1968), Monotone and Oscillation Matrices Applied to Finite Difference Approximations, Math. Comp. 22, 489-516.
9. Reich, E. (1949), On the Convergence of the Classical Iterative Methods of Solving Linear Simultaneous Equations, Ann. Math. Statist. 20, 448-451.
10. Vandergraft, J. S. (1968), Spectral Properties of Matrices Which Have Invariant Cones, SIAM J. Appl. Math., 16, 1208-1222.
11. Varga, R. (1962), Matrix Iterative Analysis, Prentice Hall, Englewood Cliffs, N. J.