Riv. Mat. Univ. Parma, Vol. **9** (2018), 191-226

Giuseppe Buffoni

On the structure of matrices with positive inverse

Abstract. This paper focuses on how monotone+ matrices, i.e., real nonsingular matrices with positive inverse, can either be perturbed or decomposed in such a way that the inverse-positivity is preserved and proved. Let a real matrix A be split into its components: diagonal entries D, nonpositive $-B$ and nonnegative C off-diagonal entries: $A = D - B + C$. Monotone+ matrices with only two components and their perturbations are identified by investigating the properties of the splittings $D - B$, $D + C$ and $D - B + C$. Monotone+ matrices characterized by three components are identified by means of more involved decompositions of A or suitable transformations of A, preserving the inverse-positivity, that emphasize the basic properties leading to inverse positivity. Special complex monotone+ matrices are described. The analysis is strongly based on some monotonicity properties of nonpositive and nonnegative perturbations of a monotone+ matrix preserving the inverse-positivity. The results are illustrated by numerical examples.

Keywords. Monotone matrices, nonnegative and nonpositive perturbations, monotonicity properties of the inverse, preserving monotonicity.

Mathematics Subject Classification (2010): 15B48, 65F30.

Contents

Received: July 5, 2018; accepted in revised form: January 24, 2019.

1 - Introduction

The concept of monotone matrices dates back to Ostrowski [**32**] and Collatz [**10**]: a matrix A is monotone if $det\ A \neq 0$ and $A^{-1} \geq 0$. Monotone matrices investigated in the literature have structures characterized by some properties of sign and size of their entries.

For structure of a monotone matrix A we mean any form in which A can be put, and such that the nonnegativitty of A^{-1} is preserved and can be proved. Different forms will be used: splitting of A, partition of A into blocks, representation of A as product of suitable matrices, transformation of A, such as $\Pi_1 A \Pi_2$, with Π_1 and Π_2 permutation matrices (it may be either $\Pi_1 = I$ or $\Pi_2 = I$, with I the identity matrix).

Let an $n \times n$ real matrix $A = (a_{ij})$ be split as

$$A = D - B + C, \quad \text{with} \;\; B,\ C \geq 0, \tag{1}$$

where D is a diagonal matrix, $B = (b_{ij})$ and $C = (c_{ij})$ are matrices with zero diagonal entries and off-diagonal entries given by

$$b_{ij} = c_{ij} = 0 \quad \text{when} \quad a_{ij} = 0,$$

$$b_{ij} = |a_{ij}| \quad \text{when} \quad a_{ij} < 0, \quad c_{ij} = a_{ij} \quad \text{when} \quad a_{ij} > 0, \quad \forall\, i \neq j.$$

Monotone matrices with nonpositive off-diagonal entries, the well known and extensively studied M-matrices, have the following structure

$$A = D - B, \quad \text{with} \quad diag\ D > 0,\ det\ A > 0,\ A^{-1} \geq 0.$$

The nonnegativity of A^{-1} implies that $\rho(D^{-1}B) < 1$, where $\rho(\cdot)$ is the spectral radius. Moreover, recall that if in addition A is irreducible, then $A^{-1} > 0$. This class is defined here as in the Varga's book ([**36**], p. 85). Note that the nonsingularity condition of A is not included in the definition of M-matrices adopted in the book by Berman and Plemmons ([**1**], p. 133). Properties of these matrices can be found, for instance, in Fan ([**14**], Theorems $3, 5, 5'$), Fiedler and Ptak [**16**], Varga ([**36**], p. 85), and Berman and Plemmons ([**1**], p. 134), where fifty different but equivalent characterizations of nonsingular M-matrices are given.

The concept of a regular splitting of a matrix was introduced by Varga ([**35**]; [**36**], p. 88). If A has a convergent regular splitting, defined by

$$A = M - N, \quad \text{with} \quad det\ M \neq 0,\ M^{-1} \geq 0,\ N \geq 0,\ \rho(M^{-1}N) < 1,$$

then $A^{-1} \geq 0$. Regular splittings have been studied by many authors (see [**2**]). We mention here the work by Ortega and Rheinboldt [**31**], where a convergent weak regular splitting $M - N$ of A is defined by $M^{-1} \geq 0$, $M^{-1}N \geq 0$, and $\rho(M^{-1}N) < 1$.

Let us consider matrices with nonnegative off-diagonal entries: they are called Metzler, quasi-positive ([**15**], annotated bibliography, p. 170), and essentially nonnegative matrices ([**36**], p. 260). If in addition they are irreducible, then are called essentially positive matrices ([**36**], p. 257). Monotone matrices belonging to this class have the following structure

$$A = D + C, \quad \text{with either } diag\ D \geq 0 \text{ or } diag\ D \leq 0,\ det\ \ A \neq 0, \quad A^{-1} \geq 0.$$

When $diag\ D \geq 0$, A is a nonnegative matrix. It follows that each row and each column of $D + C$ contains exactly one positive entry, property mentioned in ([**28**], p. 40). This property also holds for A^{-1}, and the positive entries $\hat{a}_{ij}$ of A^{-1} are given by $\hat{a}_{ij} = 1/a_{ji},\ \forall\ a_{ji} > 0$. This class of matrices will be denoted by P-matrices. Set $p = a_{1j_1}a_{2j_2}...a_{nj_n}$ the product of all $a_{ij} > 0$. Then, $det\ A = (-1)^m p$, where m is the number of permutations of the sequence $(j_1, j_2, ..., j_n)$ with respect to $(1, 2, ..., n)$. When all the positive entries of A are equal to 1, then A is a permutation matrix, for which $A^{-1} = A^T$.

Let us assume $diag\ D \leq 0$. When A is an essentially positive matrix with positive inverse, then necessarily $diag\ D < 0$ [**7**]. Properties of this class of matrices can be found in Fan ([**14**], Theorem 6), Buffoni and Galati [**7**]. This class of matrices will be denoted by W-matrices (W is M reversed); indeed, many properties of W-matrices are the reverse of those of irreducible M-matrices (see Appendix A). The sign of $det\ A$ depends on the order n of A: $(-1)^{n-1} det\ A > 0$. When A is an essentially nonnegative or essentially positive matrix with nonnegative inverse, simple examples show that it may be either $diag\ D < 0$ or $diag\ D \leq 0 \wedge \not< 0$.

For irreducible M-matrices $D - B$ and W-matrices $-|D| + C$ the following implications hold:

$$(2) \qquad (m):\ D^{-1}B \text{ irreducible}, \rho(D^{-1}B) < 1 \iff (D-B)^{-1} > 0,$$

$$(3) \qquad (w):\ \Gamma = |D|^{-1}C \text{ irreducible}, \rho(\Gamma) > 1, \forall i:\ \rho(\Gamma_i) < 1 \\ \iff (-|D| + C)^{-1} > 0,$$

where $\Gamma_i,\ i = 1, 2, ..., n$, are the $(n-1) \times (n-1)$ principal square submatrices of Γ. Note that the $n+2$ conditions fulfilled by W-matrices are not redundant [**7**]. Conditions (m) and (w) show the levels of complexity of the two structures.

Obviously, products of monotone matrices are monotone matrices.

Works addressing monotonicity-preserving perturbations of monotone matrices can be found in the literature (see [**24**]). Many works are about nonnegative perturbations of M-matrices, in particular tridiagonal M-matrices, [**4**, **21**, **22**, **23**, **24**, **27**]. Some others concern matrices which are discrete approximations of continuous differential problems [**5**,**6**,**37**], positive-definite matrices [**20**], and generic matrices [**8**,**9**].

A matrix A satisfying

$$(4) \qquad detA \neq 0, \quad A^{-1} > 0,$$

is called a monotone+ matrix, and a monotone+ matrix characterized by only two components in the splitting (1) is called a basic monotone+ matrix (e.g.,irreducible M-matrices $D - B$ and W-matrices $-|D| + C$).

In this paper the structures of monotone+ matrices are investigated, or, in other words: the structures of the inverse $P^{-1}(= A)$ of real nonsingular positive matrices $P(= A^{-1})$ are investigated. The performed analysis is strongly based on some monotonicity properties of the inverse of a monotone+ matrix A affected by nonpositive and nonnegative perturbations preserving the positivity of the inverse. These premises are described in detail in Section 2. Indeed, the results in Sections 3 and 4 are obtained by perturbing a monotone+ matrix A by $sB - tC$, with s, t nonnegative real parameters. Note that these perturbations of A depend on its own off-diagonal entries. Let $\sigma(A, B)$ and $1/\rho(A^{-1}C)$

be defined as follows

$$(A + sB)^{-1} > 0 \; for \; 0 \leq s < \sigma(A, B), \quad (A - tC)^{-1} > 0 \; for \; 0 \leq t < \frac{1}{\rho(A^{-1}C)}.$$

In Section 3 the structures of basic monotone+ matrices and of their perturbations are identified. The main results are expressed in terms of implications of type

$$(A^{-1} > 0) \wedge \text{ sign of (either } 1 - \sigma(A, B), \; or \; 1 - \rho(A^{-1}C)) \\ \Longleftrightarrow \text{properties of } D, \; B, \; C.$$

Different structures of A are identified, depending on the assumptions made on $\sigma(A, B)$ and $\rho(A^{-1}C)$. Section 4 is about monotone+ matrices with more complex structures. These structures can be identified by investigating more involved decompositions of A, or by means of transformations of A, suitable to emphasize the properties that cause the inverse-positivity. Special structures of monotone+ matrices are reported. Section 5 deals with the product of monotone matrices. Some concluding remarks can be found in Section 6. The results are illustrated by numerical examples in Appendix C.

2 - Pencils of matrices with positive inverse

Let

$$A + sP - tQ \tag{5}$$

be a pencil of $n \times n$ real matrices, where A is a monotone+ matrix, and

$$P, \; Q \geq 0 \wedge \neq 0, \quad (s, \; t) \in R_2^+. \tag{6}$$

The parameters s, t may be considered as a measure of the size of the nonnegative sP and nonpositive $-tQ$ perturbation of the matrix A. When $det(A + sP - tQ) \neq 0$, let

$$Z(s, t) = (A + sP - tQ)^{-1}.$$

For $s = t = 0$ it is $Z(0, 0) = A^{-1} > 0$. Thus, $det(A + sP - tQ) \neq 0$ and $Z(s, t) > 0$ for (s, t) in a sufficiently small neighborhood Ω of $(0, 0)$. The purpose of this section is to characterize this open and connected set $\Omega \subset R_2^+$.

2.1 - *Basic monotonicity properties of $Z(s, t)$*

Many of the results are based on the following.

Lemma 1. *Let a monotone+ matrix A be perturbed as in* (5), (6), *and let* $(s,t) \in \Omega$, *such that* $Z(s,t) > 0$. *Then,*

$$Z_s < 0,\ Z_{ss} > 0, \quad Z_t > 0,\ Z_{tt} > 0, \quad Z_{st} = Z_{ts} < 0, \qquad (s,t) \in \Omega, \tag{7}$$

where the subscripts indicate derivatives with respect to s and t.

Proof. For $(s,t) \in \Omega$ the entries of $Z(s,t)$ are positive, continuous and differentiable functions of s and t. From the identity $(A + sP - tQ)Z(s,t) = I$, where I is the $n \times n$ identity matrix, we obtain

$$Z_s = -ZPZ, \quad Z_{ss} = 2\,ZPZPZ, \quad Z_t = ZQZ, \quad Z_{tt} = 2\,ZQZQZ,$$

$$Z_{st} = Z_{ts} = -ZPZQZ - ZQZPZ.$$

From the assumptions $P, Q \geq 0 \wedge \neq 0$ and $Z > 0$ it follows that the matrices ZP and ZQ have at least one positive column, so that $ZPZ > 0$ and $ZQZ > 0$, which imply (7). □

2.2 - *Either nonnegative or nonpositive perturbations*

Let us consider the two limit situations where the perturbations of opposite sign act separately on A. When either $t = 0$ or $s = 0$ the set Ω reduces to an interval on the s-axis or on the t-axis, respectively.

Lemma 2. *Let a monotone+ matrix A be perturbed as in* (5), (6).

(i) *When $t = 0$, let s_0 and s_0^* be defined by*

$$det(A + sP) \neq 0 \ for\ s \in [0, s_0), \quad Z(s,0) > 0 \ for\ s \in [0, s_0^*).$$

Then,

$$either \quad s_0^* < s_0 \leq +\infty \qquad or \quad s_0^* = s_0 = +\infty. \tag{8}$$

Moreover, when $s_0^* < +\infty$, $Z(s_0^*, 0)$ exists and at least one of its entries is equal to 0.

(ii) *When $s = 0$, let t_0 and t_0^* be defined by*

$$det(A - tQ) \neq 0 \ for\ t \in [0, t_0), \quad Z(0,t) > 0 \ for\ t \in [0, t_0^*).$$

Then,

$$t_0^* = t_0 = \frac{1}{\rho(A^{-1}Q)} < +\infty. \tag{9}$$

Moreover, when $t \to t_0^*$ at least one entry of $Z(0,t)$ becomes $+\infty$.

Proof. Part (i). Obviously $s_0^* \le s_0$. The value s_0^* is the smallest positive real solution, if it exists, to the n^2 equations $Z(s,0) = 0$; otherwise $s_0^* = +\infty$.

Let us assume $s_0^* < +\infty$. From Lemma 1 it follows that

$$0 < Z(s,0) \le A^{-1}, \quad Z_s(s,0) < 0, \quad 0 \le s \le \hat{s} < s_0^*.$$

The value s_0^* is the least upper bound of the numbers $\hat{s}$. On the other hand, since at least one entry of $Z(s,0)$ must become infinite for $s \to s_0$ (Proposition b3 in Appendix B), it follows that $s_0^* < s_0 \le +\infty$. Therefore, $Z(s_0^*, 0)$ exists and at least one of its entries is equal to zero.

Under the assumption $s_0^* = +\infty$, also $s_0 = +\infty$.

Part (ii). Statement (9) follows from [**36**], p. 83 (or [**25**], p. 69), together with Proposition b1 in Appendix B. Moreover, from Lemma 1, which implies that the entries of $Z(0,t)$ are increasing with t for $t \in [0, t_0^*)$, and $t_0 = t_0^*$, it follows that at least one entry of $Z(0,t)$ becomes $+\infty$ as $t \to t_0$ (Proposition b3 in Appendix B). □

Remark. (r1) Let ν_j, $j = 1, 2, ..., n$, be the eigenvalues of the nonnegative matrix $A^{-1}P$, with $\nu_1 = \rho(A^{-1}P) > 0$. Since $A + sP = A(I + sA^{-1}P)$, it follows that $s_0 < +\infty$ if and only if there exist negative real eigenvalues ν_j, and it is given by

$$s_0 = min_{\nu_j < 0} \frac{1}{|\nu_j|} > \frac{1}{\nu_1}.$$

Otherwise, $s_0 = +\infty$. On the other side, simple examples show that either $s_0^* \le 1/\nu_1$ or $s_0^* > 1/\nu_1$.

(r2) Let $||\cdot||$ be any norm of the matrix in argument. Since $s_0^* < s_0$, none of the eigenvalues of $A + s_0^* P$ are zero and then $0 < \rho(Z(s_0^*, 0)) \le ||Z(s_0^*)|| < +\infty$. Differently, $||Z(0,t)|| \to +\infty$ as $t \to t_0^*$ (Proposition b3 in Appendix B).

(r3) The bound s_0^* is the smallest positive real solution, if it exists, to the n^2 equations $Z(s,0) = 0$. The value s_0^* is a function of the two matrices A and P, and it will be denoted by

$$s_0^* = \sigma(A, P), \tag{10}$$

while $t_0^* = 1/\rho(A^{-1}Q)$ depends only on the matrix $A^{-1}Q$. Given the monotonicity properties (7) of $Z(s,0)$, the computation of s_0^*, when $s_0^* < +\infty$, may be performed by means of the Newton approximation of the equations $Z(s,0) = 0$ [**8**]. Each iteration of this process needs the computation of the inverse of an $n \times n$ matrix, which requires $O(n^3)$ operations. It is possible to show [**8**] the quadratic convergence ([**33**], p. 260) of the process. When

$s_0^* = +\infty$, the successive approximations form a sequence diverging monotonically to $+\infty$. Results of numerical experiments, in particular for a diffusion matrix A of order $n \leq 50$ and different nonnegative perturbations P, can be found in [**8**, **9**, **21**, **24**].

(r4) The computation of the spectral radius $\rho(A^{-1}Q)$ may be performed through the power method ([**29**], p. 25). Each iteration of this process requires $O(n^2)$ operations.

2.3 - *Perturbations with both positive and negative entries*

The monotonicity properties of $Z(s,t)$ imply that $Z(s,t) > 0$ for $(s,t) \in [0, s_0^*) \times [0, t_0^*)$. In the (s,t) plane, this set may be either a rectangle when $s_0^* < +\infty$ or an infinite strip when $s_0^* = +\infty$. When $s_0^* < +\infty$ an extension $\Omega_0 \subset \Omega$ of this rectangle can be constructed. The set Ω_0 (Figure 1) is the union of the two sets

$$(s,t) \in [0, s^*(t)) \times [0, t_0^*) \quad \text{and} \quad (s,t) \in [0, s_0^*) \times [0, t^*(s)),$$

where the boundaries $s^*(t)$ and $t^*(s)$ are precisely specified in the following Lemmas 3 and 4, respectively.

Lemma 3. *Let a monotone+ matrix A be perturbed as in* (5), (6) *and $s_0^* < +\infty$. For fixed $t \in [0, t_0^*)$ let $s^*(t)$ be defined by*

$$det(A + sP - tQ) \neq 0, \quad Z(s,t) > 0 \quad for \quad s \in [0, s^*(t)), \quad t \in [0, t_0^*).$$

Then,

(i) *$s^*(t)$ is a continuous non decreasing function for $t \in [0, t_0^*)$, with $s^*(0) = s_0^*$; moreover, $Z(s^*(t), t)$ exists and at least one of its entries is equal to zero;*

(ii) *$s^*(t)$ is piecewise derivable.*

Proof. Part (i). The assumption $s_0^* < +\infty$ implies that $s^*(t) < +\infty$ for $0 \leq t < \tilde{t} \leq t_0^*$ Thus, for $t \in [0, \tilde{t})$ there exist finite positive real solutions to the n^2 equations $Z(s,t) = 0$; $s^*(t)$ is the smallest of them. From Lemma 2 part (i), it follows that $det(A + sP - tQ) \neq 0$ for $s \in [0, s^*(t))$. Therefore $s^*(t)$ is a solution to the n^2 polynomial equations $Z(s,t)\ det(A + sP - tQ) = 0$; thus, it is a continuous function of t [**34**], and $s^*(0) = s_0^*$. Moreover, from the monotonicity properties (7) of $Z(s,t)$, it follows that $s^*(t)$ is non decreasing with t, and $Z(s^*(t), t)$ exists and at least one of its entries is equal to zero.

Part (ii). Let $Z(s,t) = (z_{ij}(s,t))$; as previously stated, $s^*(t)$ is the smallest solution, if it exists, to the n^2 equations $z_{ij}(s,t) = 0,\ i,j = 1,2,...,n$. Let $s^*(t)$ be a solution to $\zeta(s^*(t),t) = z_{i_0 j_0}(s^*(t),t) = 0$ for t in a finite interval. From Lemma 1 $\zeta_s(s,t) < 0$ and $\zeta_t(s,t) > 0$ when $\zeta(s,t) > 0$; thus, when $s^*(t)$ is not a constant, ds^*/dt exists and it is given by

$$\frac{ds^*}{dt} = -\frac{\zeta_t(s^*,t)}{\zeta_s(s^*,t)} > 0.$$

Obviously, $ds^*/dt = 0$ when $s^*(t) = constant$.

If there exists $\hat{t}$ such that

$$z_{i_1\, j_1}(s^*(t),t) = 0 \quad \text{for} \quad t \le \hat{t} \quad \text{and} \quad z_{i_2\, j_2}(s^*(t),t) = 0 \quad \text{for} \quad t \ge \hat{t},$$

then the derivative of $s^*(t)$ may be discontinuous at $t = \hat{t}$. □

Lemma 4. *Let a monotone+ matrix A be perturbed as in (5), (6) and $s_0^* \le +\infty$. For fixed $s \in [0, s_0^*)$ let $t^*(s)$ be defined by*

$$det(A + sP - tQ) \neq 0, \quad Z(s,t) > 0 \quad for \quad s \in [0, s_0^*), \quad t \in [0, t^*(s)).$$

Then

$$t^*(s) = \frac{1}{\rho(Z(s,0)Q)}, \quad s \in [0, s_0^*),$$

is a continuous strictly increasing function of s for $s \in [0, s_0^*)$, with $t^*(0) = t_0^*$, and at least one entry of $Z(s,t)$ becomes $+\infty$ as $t \to t^*(s)$. Moreover, when $s_0^* = +\infty$, it follows that $\Omega_0 = \Omega$.

Proof. The result follows from Lemma 2 part (ii), together with Proposition b2 in Appendix B. Moreover, when $s_0^* = +\infty$, the set Ω_0, given by $(s,t) \in [0,+\infty) \times [0, t^*(s))$ cannot be further extended. □

The positivity of $Z(s,t)$ is not proved in the following boundary points of Ω_0:

$$Z(s_0^*, t) \ for \ t \ge t_0^*, \qquad Z(s, t_0^*) \ for \ s \ge s_0^*.$$

A dichotomy arises: either some of the entries of $Z(s_0^*, t_0^*)$ are equal to zero, or all its entries are positive. These two possible outcomes lead to different scenarios in the (s,t) plane.

(i) Assume that at least one entry of $Z(s_0^*, t_0^*)$ is zero. Thus, $s^*(t) = s_0^*$, and from Proposition b4 in Appendix B it follows that $s^*(t) = s_0^*$ for any $t \ge t_0^*$. The curve $t^*(s)$, defined for $s \in [0, s_0^*)$, is increasing with s and may intersect the line $s = s_0^*$ at a finite point, or not.

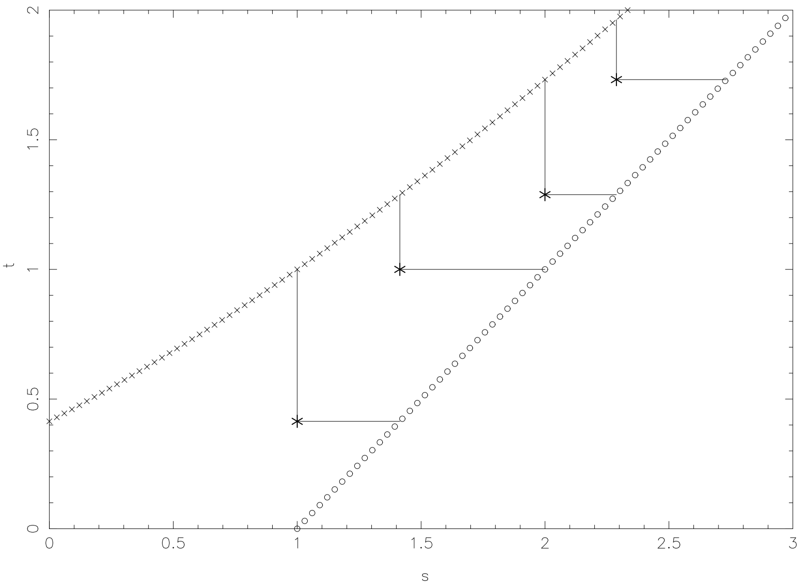

Fig. 1. Successive approximations Ω_k of the set Ω under the assumptions $Z(s_k^*, t_k^*) > 0$; points $(s_k^*, t_k^*), k = 0, 1, 2, 3$, are indicated by "$*$"; boundary points $Z \geq 0 \wedge \ngtr 0$, i.e. $s^*(t)$, "$\circ$"; $||Z|| = +\infty$, i.e. $t^*(s)$, "$\times$"; $Z > 0$, continuous line.

(ii) Assume now that $Z(s_0^*, t_0^*) > 0$. Then, from Lemma 2 there exist s_1^* and t_1^* such that $Z(s, t_0^*) > 0$ for $s \in [s_0^*, s_1^*)$, and $Z(s_0^*, t)$ for $t \in [t_0^*, t_1^*)$. The continuity and monotonicity properties of $Z(s, t)$ in Ω_0, lead to

$$s_1^* = \lim_{t \to t_0^*} s^*(t) = s^*(t_0^*), \quad t_1^* = \lim_{t \to s_0^*} t^*(s) = t^*(s_0^*).$$

Moreover, from Lemmas 3 and 4, $s^*(t)$ can be defined for $t \in [t_0^*, t_1^*)$, and $t^*(s)$ for $s \in [s_0^*, s_1^*)$; moreover, they are continuous, non decreasing and increasing, respectively.

Then, a set Ω_1, $\Omega_0 \subset \Omega_1 \subset \Omega$, is constructed. If some of the entries of $Z(s_1^*, t_1^*)$ are equal to zero, then $s^*(t) = s_1^*$ for $t > t_1^*$, and the process is stopped. On the contrary, if $Z(s_1^*, t_1^*) > 0$, then the process may be continued. Under the assumption that $Z(s_k^*, t_k^*) > 0$, $k = 0, 1, 2, ...$, a sequence of sets Ω_k, which are successive approximations of the set Ω, is obtained (figure 1).

R e m a r k . (r5) The function $t^*(s)$ is strictly increasing with s, thus the function $s = (t^*)^{-1}(t)$ exists, and $(t^*)^{-1}(t) < s^*(t)$ for t sufficiently small, until these two curves intersect, otherwise $(t^*)^{-1}(t) < s^*(t)$ for $t \to +\infty$. Indeed, $s^*(t)$ may be either a constant or strictly increasing with t, and the two curves $s^*(t)$ and $t^*(s)$ may either cut across at a finite point, or approach $+\infty$ without intersection. Let $(\hat{s}, \hat{t})$ be the possible intersection point of $s^*(t)$ and $t^*(s)$.

Thus, the entries of the limit matrix $\lim_{t\to\hat{t}} Z(\hat{s},t)$ are either finite and nonnegative or $+\infty$. The analytical expressions of $s^*(t)$ and $t^*(s)$ can be obtained only for n equal to the first few integers. The various expected scenarios can be illustrated by pencils of matrices of order $n = 2, 3$ (for instance, see Examples (e4), (e5), (e9), (e10) and Figure 2 in Appendix C).

(r6) Let us assume that $s = t = u$ and let $A + u(P - Q)$ be nonsingular with $Z(u,u) > 0$ for $u \in [0, u^*)$. Then, $u^* \geq min(s_0^*, t_0^*)$. The critical value u^* is determined by the intersection point, if it exists, between the bisecting line $t = s$ and one of the two curves $s^*(t)$ or $t^*(s)$. Otherwise, $u^* = +\infty$. An iterative process to compute u^* is described in [**9**], where results of numerical experiments, for a diffusion matrix A of order $n \leq 50$ and different perturbation P and Q, can be found. The algorithm is rather involved, nevertheless it produces reliable results.

3 - Structures of A when either $\rho(A^{-1}C) \leq 1$ or $\sigma(A,B) \geq 1$

Let $A = D - B + C$ be a monotone+ matrix. From (9) and $\rho(A^{-1}C) < 1$ it follows that the perturbed matrix $A - tC = D - B + (1-t)C$, $t > 0$, preserves the inverse-positivity also for $t > 1$; this means that C is a sufficiently "weak" component of A with respect to B. Analogously, from (8), (10) and $\sigma(A,B) > 1$ it follows that the perturbed matrix $A + sB = D - (1-s)B + C$, $s > 0$, preserves the inverse-positivity also for $s > 1$; this means that $-B$ is a sufficiently "weak" component of A with respect to C. Note that the matrices characterized by $\rho(A^{-1}C) \leq 1$ include the limit case $C = 0$ ($1/\rho(A^{-1}C) = +\infty$), i.e. the case of irreducible M-matrices, and those characterized by $\sigma(A,B) \geq 1$ include the limit case $B = 0$ ($\sigma(A,B) = +\infty$), i.e. the case of W-matrices.

Theorem 1. *Let $A = D - B + C$ be a monotone+ matrix and either $\rho(A^{-1}C) < 1$ or $\sigma(A,B) > 1$ (it may be $\sigma(A,B) = +\infty$). Then, the following implications hold*

$$\text{(11)}\quad (i)\ A^{-1} > 0,\ \rho(A^{-1}C) < 1 \iff (D-B)^{-1} > 0,\ \sigma(D-B,C) > 1,$$

$$\text{(12)}\quad (ii)\ A^{-1} > 0,\ \sigma(A,B) > 1 \iff (D+C)^{-1} > 0,\ \rho((D+C)^{-1}B) < 1.$$

Proof. Part (i). ($\Longrightarrow$) Let A be perturbed by the nonpositive matrix $-tC$, $t > 0$. Then, from (9) it follows that

$$\text{(13)}\qquad (A - tC)^{-1} > 0 \quad for \quad 0 \leq t < \frac{1}{\rho(A^{-1}C)}.$$

Under the assumption $\rho(A^{-1}C) < 1$, (13) holds for $t = 1$, so that

$$(A - C)^{-1} = (D - B)^{-1} > 0. \tag{14}$$

Furthermore, since $A^{-1} > 0$ and $(D - B)^{-1} > 0$, the inverse of $A + tC = D - B + (1+t)C$ is positive for either $0 \le t < \sigma(A, C)$ or $0 \le 1+t < \sigma(D-B, C)$. Thus,

$$\sigma(D - B, C) = 1 + \sigma(A, C) > 1.$$

($\Longleftarrow$) From $(D - B)^{-1} > 0$ and $\sigma(D - B, C) > 1$ it follows that the inverse of $A = D - B + C$ is positive. Moreover, the inverse of $A - tC = D - B - (t-1)C$ is positive for $0 \le t < 1/\rho(A^{-1}C)$ and $0 \le t - 1 < 1/\rho((D - B)^{-1}C)$. Thus,

$$\rho(A^{-1}C) = \frac{\rho((D - B)^{-1}C)}{1 + \rho((D - B)^{-1}C)} < 1.$$

Part (ii). ($\Longrightarrow$) Let A be perturbed by the nonnegative matrix sB, with $s > 0$. Then, from (8) and (10) it follows that

$$(A + sB)^{-1} > 0 \quad for \quad 0 \le s < \sigma(A, B). \tag{15}$$

Under the assumption $\sigma(A, B) > 1$, (15) holds for $s = 1$, so that

$$(A + B)^{-1} = (D + C)^{-1} > 0. \tag{16}$$

Furthermore, since $A^{-1} > 0$ and $(D + C)^{-1} > 0$, the inverse of $A - sB = D + C - (1 + s)B$ is positive for either $0 \le s < 1/\rho(A^{-1}B)$ or $1 \le 1 + s < 1/\rho((D + C)^{-1}B)$. Thus,

$$\rho((D + C)^{-1}B) = \frac{\rho(A^{-1}B)}{1 + \rho(A^{-1}B)} < 1.$$

($\Longleftarrow$) From $(D+C)^{-1} > 0$ and $\rho((D+C)^{-1}B) < 1$ it follows that the inverse of $A = D-B+C$ is positive. Moreover, the inverse of $A+sB = D+C+(s-1)B$ is positive for $0 \le s < \sigma(A, B)$ and $0 \le s - 1 < \sigma(D + C, B)$. Thus,

$$\sigma(A, B) = 1 + \sigma(D + C, B) > 1. \tag{17}$$

□

R e m a r k . (r7) When $\rho(A^{-1}C) < 1$, then $(D - B)^{-1} > 0$ which implies *diag* $D > 0$, B irreducible, $\rho(D^{-1}B) < 1$. Moreover

$$\rho((D - B + tC)^{-1}tC) = \frac{t}{t + 1/\rho((D - B)^{-1}C)} < 1, \qquad 0 \leq t < \sigma(D - B, C),$$

which confirms that C is a weak component of A with respect to B. The matrix A is composed by an irreducible M-matrix $D - B$ perturbed by a nonnegative matrix C preserving the inverse-positivity.

(r8) When $\sigma(A, B) > 1$, then $(D + C)^{-1} > 0$ which implies *diag* $D < 0$, C irreducible, and that the matrix $|D|^{-1}C$ must satisfy conditions (w) in (3). Moreover,

$$\sigma(D - sB + C, sB) = 1 + \frac{\sigma(D + C, B)}{s}, \qquad 0 \leq s < \frac{1}{\rho((D + C)^{-1}B)},$$

which confirms that B is a weak component of A with respect to C. The matrix A is composed by a W-matrix $D + C$ perturbed by a nonpositive matrix $-B$ preserving the inverse-positivity. Furthermore, it follows that A can be represented as the product of a W-matrix by an M-matrix

$$A = (D + C)\,(I - (D + C)^{-1}B).$$

T h e o r e m 2. *Let* $A = D - B + C$ *be a monotone+ matrix and either* $\rho(A^{-1}C) = 1$ *or* $\sigma(A, B) = 1$. *Then, the following implications hold*

$$\text{(18)} \quad (i) \quad A^{-1} > 0, \quad \rho(A^{-1}C) = 1 \quad \Longrightarrow \quad det(D - B) = 0, diag\ D > 0,$$

$$\text{(19)} \quad (ii) \quad A^{-1} > 0, \quad \sigma(A, B) = 1 \quad \Longrightarrow \quad (D + C)^{-1} \geq 0 \wedge \not> 0.$$

P r o o f. Part (i). Under the assumptions $A^{-1} > 0$, $C \geq 0 \wedge \neq 0$ the eigenvector $\mathbf{x}$ of $A^{-1}C$ corresponding to $\rho(A^{-1}C)$ is positive (Prop. b1 in Appendix B). When $\rho(A^{-1}C) = 1$, from the equation $A^{-1}C\mathbf{x} = \mathbf{x}$, with $\mathbf{x} > 0$, it follows that $(D - B)\mathbf{x} = 0$. Thus, *diag* $D \geq 0$, and $det(D - B) = 0$. Moreover, if just one diagonal entry $a_{i_0 i_0} = 0$, then $b_{i_0 j} = 0\ \forall j$.

Let $\mathbf{y}_s > 0$ be the eigenvector of $(A + sB)^{-1}C$ corresponding to $\rho((A + sB)^{-1}C) < \rho(A^{-1}C) = 1$ for $0 < s < \sigma(A, B)$. From the eigenvalue equation

$$\Big(\frac{1}{\rho((A + sB)^{-1}C)} - 1\Big)C\mathbf{y}_s = (D - (1 - s)B)\mathbf{y}_s,$$

it follows that if just one diagonal entry $a_{i_0 i_0} = 0$, then $c_{i_0 j} = 0\ \forall j$.

In short, if just one diagonal entry of A is zero, then A should have one row equal to zero, and consequently *det* $A = 0$. It follows that *diag* $D > 0$.

Part (ii). When $\sigma(A, B) = 1$ equation (16) becomes

$$(A + B)^{-1} = (D + C)^{-1} \geq 0 \wedge \not> 0.$$

□

Remark. (r9) When $\rho(A^{-1}C) = 1$ the matrix A is composed by a singular matrix $D - B$ with $diag\ D > 0$ and $\rho(D^{-1}B) = 1$ and by a nonnegative matrix C, which is an essential component of A for its nonsingularity and inverse-positivity.

(r10) When $\sigma(A, B) = 1$, from (19) it follows that

$$(diag\ D \geq 0 \wedge \not> 0) \vee (diag\ D \leq 0).$$

This dichotomy leads to the following scenarios:

(a) $(D + C)^{-1} \geq 0$ and $diag\ D \geq 0 \wedge \not> 0$ ($diag\ D > 0 \Longrightarrow C = 0$), which imply that $D + C$ is a P-matrix,

(b) $(D+C)^{-1} \geq 0$ and $diag\ D \leq 0$, which imply that $(D+C)$ is an essentially nonnegative or an essentially positive matrix with nonnegative inverse.

In both cases the nonpositive matrix $-B$ is an essential component of A for its inverse-positivity.

The implications (11), (12), (18), (19) lead to incompatible characterizations of the diagonal matrix D:

$$A^{-1} > 0,\ \rho(A^{-1}C) \leq 1 \Longrightarrow diag\ D > 0,$$
$$A^{-1} > 0,\ \sigma(A, B) \geq 1 \Longrightarrow diag\ D \not> 0.$$

Thus, the following implications hold

$$A^{-1} > 0,\ \rho(A^{-1}C) \leq 1 \Longrightarrow \sigma(A, B) < 1,$$
$$\rho(A^{-1}C) > 1 \Longleftarrow A^{-1} > 0,\ \sigma(A, B) \geq 1.$$

Taking into account all the results obtained in this section, the scenarios put in evidence are summarized in Table 1. There are three situations involving the limit cases $\rho(A^{-1}C) = 1$ and $\sigma(A, B) = 1$; they are illustrated by the Examples (e1), (e2), (e3) in Appendix C. Situations with $\sigma(A; B) > 1$ are illustrated in the Examples (e4) in Appendix C and Figure 2.

Table 1. *Properties of the components of* $A = D - B + C$, *with* $A^{-1} > 0$, *which termine the structure of* A *when* $\rho(A^{-1}C) \leq 1$ *and* $\sigma(A, B) \geq 1$. *Conditions* (w) *are ported in* (3).

	$\rho(A^{-1}C) < 1$	$\rho(A^{-1}C) = 1$	$\rho(A^{-1}C) > 1$
$\sigma(A, B) < 1$	$(D - B)^{-1} > 0$ $diag\, D > 0$ $\rho(D^{-1}B) < 1$	$det(D - B) = 0$ $diag\, D > 0$ $\rho(D^{-1}B) = 1$	more complex structures
$\sigma(A, B) = 1$	—	—	$(D + C)^{-1} \geq 0 \wedge \not> 0$ either $diag\, D \leq 0$ or $diag\, D(\geq 0 \wedge \not> 0) \vee (\leq 0)$
$\sigma(A, B) > 1$	—	—	$(D + C)^{-1} > 0$ $diag\, D < 0$ (w)

4 - Structures of A when $\rho(A^{-1}C) > 1$ and $\sigma(A, B) < 1$

Let a monotone+ matrix $A = D - B + C$ be perturbed by $sB - tC$, with $s, t \geq 0$, and let $Z(s, t)$ its inverse:

$$(20) \quad A + sB - tC = D - (1 - s)B + (1 - t)C, \qquad Z(s, t) = (A + sB - tC)^{-1}.$$

As shown in Subsection 2.3, $Z(s, t)$ exists and is positive in a set Ω defined by

$$(21) \qquad \Omega = \{(s, t) \in R_2^+ \,|\, 0 \leq s < s^*(t), \quad 0 \leq t < t^*(s)\},$$

where $s^*(t)$ is non decreasing with t starting from $s^*(0) = \sigma(A, B)$, and $t^*(s)$ is increasing with s starting from $t^*(0) = 1/\rho(A^{-1}C)$. The special perturbations of A in (20), expressed in terms of its own off-diagonal entries, allow to restrict the set Ω. Indeed, the set Ω cannot contain points (s, t) with both $s > 1,\ t > 1$. Otherwise, for $s = t = 1$ the perturbed matrix defined in (20) should become a diagonal matrix: $A + B - C = D$, obviously with nonpositive inverse, when it exists. Thus the set Ω should be contained just in one of the following strips in the parameter space (s, t):

$$\Omega \subset \Omega_{1\infty}, \quad \Omega \subset \Omega_{\infty 1},$$

where

$$\Omega_{s_1t_1} = \{(s,t) \in R_2^+ \,|\, 0 \le s \le s_1,\ 0 \le t \le t_1\}. \tag{22}$$

Under the assumptions that

$$\rho(A^{-1}C) \le 1 \ \vee\ \sigma(A,B) \ge 1, \tag{23}$$

considered in the previous Section 3, it follows that $s^*(t) \le 1,\ t^*(s) \ge 1$ and $(\Omega \subset \Omega_{1\infty}) \wedge (\Omega \not\subset \Omega_{11})$, or $s^*(t) \ge 1,\ t^*(s) \le 1$ and $(\Omega \subset \Omega_{\infty 1}) \wedge (\Omega \not\subset \Omega_{11})$. Under the assumption

$$\rho(A^{-1}C) > 1 \ \wedge\ \sigma(A,B) < 1 \tag{24}$$

necessarily one has that neither $C = 0$ nor $B = 0$, $(B, C \ge 0 \wedge \ne 0)$. Moreover, $(s^*(t),\ t^*(s)) \in \Omega_{11}$ for s, t small enough. Then, for increasing s and t, just only one of the two curves $s^*(t)$, $t^*(s)$ may become greater than 1, or both of them remain under 1. The sign of *diag* D can be again precisely specified only when $\Omega \not\subset \Omega_{11}$. In any case, in all the examples considered with $\Omega \subset \Omega_{11}$, it was found neither *diag* $D > 0$ nor *diag* $D < 0$.

4.1 - *Sign of diag D when $\Omega \not\subset \Omega_{11}$*

Theorem 3. *Let $A = D - B + C$ be a monotone+ matrix satisfying* (24). *Then the following implications hold*

$$\text{(i)}\quad (\Omega \subset \Omega_{1\infty}) \wedge (\Omega \not\subset \Omega_{11}) \iff \textit{diag } D > 0,\ B \textit{ irreducible}, \tag{25}$$

$$\text{(ii)}\quad (\Omega \subset \Omega_{\infty 1}) \wedge (\Omega \not\subset \Omega_{11}) \iff \textit{diag } D < 0,\ C \textit{ irreducible}. \tag{26}$$

Proof. Part (i). $(\Longrightarrow)$ From (24), $s^*(0) = \sigma(A,B) < 1$. Thus the assumptions $\Omega \subset \Omega_{1\infty}$ and $\Omega \not\subset \Omega_{11}$ imply that

$$\exists \hat{s} :\ 0 < \hat{s} < s^*(1) \le 1,\ t^*(s) \ge 1 \ for\ s \ge \hat{s}$$

from which

$$Z(s,1) = (A + sB - C)^{-1} = (D - (1-s)B)^{-1} > 0, \quad s \in (\hat{s}, s^*(1)), \tag{27}$$

and then $D - (1-s)B$ for $s \in (\hat{s}, s^*(1))$ is an irreducible M-matrix.

$(\Longleftarrow)$ Let $\mathbf{x} > 0$ be the eigenvector of $A^{-1}C$ associated with $\rho(A^{-1}C)$. The eigenvalue equation written as

$$\Big(1 - \frac{1}{\rho(A^{-1}C)}\Big) D^{-1}C\mathbf{x} = (D^{-1}B - I)\mathbf{x},$$

yields $\rho(D^{-1}B) > 1$.

The matrix $Z(s,1)$, given in (27), is positive for $0 < 1-s < 1/\rho(D^{-1}B)$, i.e. for

$$\hat{s} = 1 - \frac{1}{\rho(D^{-1}B)} < s < s^*(1) = 1. \tag{28}$$

Thus, from continuity arguments, $Z(s,t)$ is positive also in a sufficiently small region around $t = 1$. Moreover, for $s \in (\hat{s}, s^*(1))$: $\lim_{s\to\hat{s}} ||Z(s,1)|| = +\infty$, $Z_s(s,1) < 0$, and $Z_t(s,t) < 0$ for $t < 1$, till $t = s^*(t)$ or $t = 0$.

Part (ii). ($\Longrightarrow$) From (24), $t^*(0) = 1/\rho(A^{-1}C) < 1$. Thus, the assumptions $\Omega \subset \Omega_{\infty 1}$ and $\Omega \not\subset \Omega_{11}$ imply that

$$\exists \hat{t}:\ 0 < \hat{t} < t^*(1) \leq 1,\ s^*(t) \geq 1\ for\ t \geq \hat{t}.$$

It follows that

$$Z(1,t) = (A + B - tC)^{-1} = (D + (1-t)C)^{-1} > 0, \quad t \in (\hat{t}, t^*(1)), \tag{29}$$

and then $D + (1-t)C$ for $t \in (\hat{t}, t^*(1))$ is a W-matrix.

($\Longleftarrow$) Let $\mathbf{x} > 0$ be the eigenvector of $A^{-1}C$ associated with $\rho(A^{-1}C) > 1$. The eigenvalue equation, written as

$$\Big(1 - \frac{1}{\rho(A^{-1}C)}\Big)|D|^{-1}C\mathbf{x} = (|D|^{-1}B + I)\mathbf{x},$$

yields

$$\frac{1}{\rho(A^{-1}C)} + \frac{1}{\rho(D^{-1}C)} < 1. \tag{30}$$

Let $\Gamma = |D|^{-1}C$, $\Gamma_i,\ i = 1, 2, ..., n-1$, be the principal square submatrices of Γ, and $\rho^* = max_i\ \rho(\Gamma_i)$. From (30) it follows that $\rho(\Gamma) > 1$, and from [**36**], p. 30, that $\rho^* < \rho(\Gamma)$. The matrix $-I + (1-t)\Gamma$ satisfies conditions (w) in (3) for

$$\hat{t} = max\big(0,\ 1 - \frac{1}{\rho^*}\big) < t < t^*(1) = 1 - \frac{1}{\rho(\Gamma)}. \tag{31}$$

In fact, $(1-t)\rho(\Gamma) > 1$ and $(1-t)\rho^* < 1$. Then, the inverse of $Z(1,t) = |D|(-I + (1-t)\Gamma)$ is positive. From continuity arguments, $Z(1,t)$ is positive also in a sufficiently small region around $s = 1$. Moreover, for $t \in (\hat{t}, t^*(1))$: $\lim_{t\to\hat{t}} ||Z(1,t)|| = +\infty$, $Z_t(1,t) > 0$, and $Z_s(s,t) > 0$ for $s < 1$, till $s = t^*(t)$ or $s = 0$ □

R e m a r k . (r11) About Theorem 3, case(i): since $(D - B)^{-1} \not\geq 0$ the component C must have a lower and an upper bound to assure the inverse-positivity of A. Case(ii): since $(D + C)^{-1} \not\geq 0$ the component B must be bounded as in case(i). In all the matrices used in the numerical examples (of order $n \leq 4$), the suitable ranges of the entries of B and C to assure the inverse-positivity of A are narrow (see Examples (e5) and (e6) in Appendix C).

(r12) The matrix $A + sB - tC = D - (1 - s)B - (t - 1)C$ of case (i) is an irreducible M-matrix in the region $s \in (\hat{s}, 1)$, $t \in (1, t^*(s))$ of the (s, t) plane, wherein

$$\rho((1 - s)D^{-1}B + (t - 1)D^{-1}C) < 1. \tag{32}$$

Inequality (32) may hold in an infinite region, as in the Example (e5) in Appendix C, for which $s^*(t) = 1$ for $t \geq 1$ and $\lim_{s\to 1} t^*(s) = +\infty$. A second situation characterized by $\sigma(A, B) > 1$ and infinite Ω is given in the Example (e4) in Appendix C for which $\lim_{t\to 1} s^*(t) = +\infty$ and $\lim_{s\to +\infty} t^*(s) = 1$. The two examples are illustrated in Figure 2.

4.2 - *Special structures of A*

Only under the assumption (24) the sign of *diag* D cannot be specified. It may be either *diag* $D > 0$ or *diag* $D < 0$ when $\Omega \not\subset \Omega_{11}$, as stated in the previous Section 4.1, while it is undefined when $\Omega \subset \Omega_{11}$. In general let D be given by $D = D_+ - D_-$, with D_+ and D_- diagonal matrices having diagonal entries $0.5(|a_{ii}| + a_{ii})$ and $0.5(|a_{ii}| - a_{ii})$, respectively. Thus, A is written as

$$A = (D_+ + C) - (D_- + B).$$

Furthermore, under the assumption (24), the structure of A cannot be identified again by the properties of the splitting $A = D - B + C$, but by different decompositions of A, which put in evidence suitable properties to state the inverse-positivity of A. These decompositions may be obtained by Π-transformations defined by

$$\hat{A} = \Pi_1 A \Pi_2,$$

where Π_1 and Π_2 are permutation matrices (it may be either $\Pi_1 = I$ or $\Pi_2 = I$). These transformations preserve the inverse-positivity of A. On the other side, they do not preserve the spectrum (unless $\Pi_2 = \Pi_1^T$, similarity transformations), the irreducibility property, and change the three components D, $-B$, C. Note that different transformations of the same monotone+ matrix may put in evidence different structures (see points (III), (IV) of this subsection).

We show a simple example on the loss of the irreducibility property. Consider the matrix

$$A = \begin{vmatrix} a & 1 & 0 \\ 1 & b & 1 \\ 1 & 0 & 0 \end{vmatrix}$$

with $a, b \neq 0$ and $det\ A = 1$. It is easy to verify that its directed graph is strongly connected, and then it is irreducible [**36**], p. 20. (Nevertheless, the linear system $A\mathbf{x} = \mathbf{k}$ is reducible). There exist two permutation matrices Π_a and Π_b (with the following entries equal to 1: $p_{a13} = p_{a22} = p_{a31} = 1$ and $p_{b13} = p_{b21} = p_{b32} = 1$) such that $\Pi_a A$ and $A\Pi_b$ are reducible matrices ([**18**], p. 50, definition 2′; [**36**], p. 18).

Special structures of complex monotone+ matrices are described in the following.

(I) Any monotone+ matrix A may be split in infinite ways as

$$A = M \pm N \quad with \quad det\ M \neq 0,\ M^{-1} \geq 0,\ N \geq 0.$$

Let N_0 be an arbitrary nonnegative matrix. Then, it is enough to take $N = \pm s N_0$ as perturbations of A, and consequently define M as

$$M = A - N,\ for\ \ 0 < s < \frac{1}{\rho(A^{-1}N_0)}; \quad M = A + N,\ for\ \ 0 < s < \sigma(A, N_0).$$

It follows

$$\rho(M^{-1}N) = \frac{\rho(A^{-1}N)}{1 + \rho(A^{-1}N)} < 1, \quad \sigma(M, N) = 1 + \sigma(A, N) > 1.$$

The first inequality is contained in ([**36**], p. 89, proof of the convergence of a regular splitting of a monotone matrix). Construction of suitable splittings can help to recognize the implicit structures, that cause the inverse-positivity.

(II) Under the assumption $D_+ + C \geq P$, with P a P-matrix, A can be expressed as

$$A = P - B_0 + C_0, \tag{33}$$

where $B_0 = D_- + B$, $C_0 = D_+ + C - P \geq 0$. Then,

$$P^{-1}B_0 \ \text{ irreducible},\ \rho(P^{-1}B_0) < 1,\ \sigma(P - B_0, C_0) > 1 \quad \Longrightarrow \quad A^{-1} > 0.$$

It can be shown that

$$\rho(A^{-1}C_0) < 1 \ \text{ while } \ \rho(A^{-1}C) > 1, \quad \sigma(P - B_0, C_0) = 1 + \sigma(A, C_0) > 1.$$

Similarly, under the assumption $D_- + B \geq P$, with P a P-matrix, A can be expressed as

$$A = -P + C_0 - B_0, \tag{34}$$

where $B_0 = D_- + B - P \geq 0$, $C_0 = D_+ + C$. Then,

$$P^{-1}C_0 \text{ irreducible, conditions } (w) \text{ in } (3) \text{ with } \Gamma = P^{-1}C_0 \implies A^{-1} > 0.$$

It can be shown that

$$\sigma(A, B_0) > 1, \quad \text{while } \sigma(A, B) < 1, \quad \rho((-P + C_0)^{-1}B_0) = \frac{\rho(A^{-1}B_0)}{1 + \rho(A^{-1}B_0)} < 1.$$

These situations are illustrated by Examples (e2), (e3), (e7), (e8) in Appendix C.

A very special situation is obtained when $A = P_1 - P_2$, wherein P_1 and P_2 are P-matrices. In such case,

$$\rho(P_1^{-1}P_2) < 1 \iff A^{-1} > 0, \quad \rho(P_2^{-1}P_1) < 1 \iff A^{-1} < 0.$$

(III) Splittings like (33), (34) can be obtained by means of transformations of type

$$\hat{A} = \Pi A = \pm P - B_0 + C_0, \quad \hat{A} = A\Pi = \pm P - B_0 + C_0,$$

where Π is a permutation matrix, P is a P-matrix, and $B_0, C_0 \geq 0$.

Let us consider the matrix A of Example (e1) in Appendix C. The matrix $D - B$ is singular: $\lambda(D^{-1}B) = 0, \pm 1$. Thus, C cannot be considered a perturbation of a monotone+ matrix. Let Π be the permutation matrix with entries $\pi_{12} = \pi_{23} = \pi_{31} = 1$, and let $\hat{A}$ be the transformed matrix

$$\hat{A} = \Pi A = P - \Pi B + \Pi C = \begin{vmatrix} -1 & 2 & -1 \\ q & -1 & 1 \\ 1 & -1 & p \end{vmatrix}$$

where $P = \Pi D$ is a P-matrix with entries $p_{12} = 2$, $p_{23} = p_{31} = 1$. Now the matrix $P - \Pi B$ is non singular: $\lambda(P^{-1}\Pi B) = 0, \pm 1/\sqrt{2}$, and $P^{-1}\Pi B$ is irreducible. ΠC may be considered as a perturbation of the monotone+ matrix $P - \Pi B$.

(IV) Let A_i be the $(n-1) \times (n-1)$ principal square submatrix of A. By means of a suitable permutation matrix Π_i we can write

$$\Pi_i A \Pi_i^T = \begin{vmatrix} A_i & U_i \\ V_i^T & a_{ii} \end{vmatrix}$$

Under the assumption that $det\ A_i \neq 0$, the inverse of $\Pi_i A \Pi_i^T$ is given by

$$\Pi_i A^{-1} \Pi_i^T = \frac{1}{b_i} \begin{vmatrix} A_i^* & -A_i^{-1} U_i \\ -V_i^T A_i^{-1} & 1 \end{vmatrix}$$

where

$$A_i^* = b_i A_i^{-1} + A_i^{-1} U_i V_i^T A_i^{-1}, \quad b_i = a_{ii} - V_i^T A_i^{-1} U_i = det\ A / det\ A_i.$$

Obviously, the transformation $\Pi_i A \Pi_i^T$ preserves the inverse-positivity. Irreducible M-matrices and W-matrices satisfy the following conditions (see Appendix A)

$$\forall i : \quad b_i > 0,\ A_i^{-1} U_i < 0,\ V_i^T A_i^{-1} < 0,$$

which imply the inverse-positivity of A. However, in order to have $A^{-1} > 0$, conditions weaker than those satisfied by either M-matrices or W-matrices can be fulfilled. From the previous arguments it follows that:

(35) $\exists i_0 :\ b_{i_0} > 0,\ A_{i_0}^{-1} > 0,\ A_{i_0}^{-1} U_{i_0} < 0,\ V_{i_0}^T A_{i_0}^{-1} < 0, \implies A^{-1} > 0,$

(36) $\exists i_0 : b_{i_0} > 0,\ A_{i_0}^{-1} < 0,\ A_{i_0}^{-1} U_{i_0} < 0,\ V_{i_0}^T A_{i_0}^{-1} < 0,\ A_{i_0}^* > 0 \implies A^{-1} > 0.$

Note that the inequalities in the left member of (35) imply $A_{i_0}^* > 0$. On the other side, when $A_{i_0}^{-1} < 0$, as in the implication (36), it is necessary to require that $A_{i_0}^* > 0$. These situations can be found in simple examples, where the conditions in the implications (35), (36) are fulfilled just for only one i. See Examples (e1), (e5), (e6), (e9), (e10) in Appendix C.

5 - Product of monotone matrices

Products of monotone+ matrices have structures showing the three components D, $-B$, C. Let us consider the case of products of M-matrices, W-matrices and P-matrices. The product of two M-matrices $D_i - X_i$ and two W-matrices $-D_i + X_i$ (with $diag\ D_i > 0$, $diag\ X_i = 0$, $X_i \geq 0$, $i = 1, 2$) look the same at a first glance:

$$(\pm D_1 \mp X_1)(\pm D_2 \mp X_2) = (D_1 D_2 + diag\ X_1 X_2) - (D_1 X_2 + X_1 D_2) + Odiag\ X_1 X_2,$$

where $Odiag\ X_1X_2 = B_1B_2$ represent the off-diagonal entries of X_1X_2. The diagonal entries are positive, and in general both nonpositive and nonnegative components are essential for the inverse positivity. For M-matrices, only when $\rho(D_1^{-1}X_1)$ and $\rho(D_2^{-1}X_2)$ are small enough, the matrix $D_1D_2 + diag\ X_1X_2 - (D_1X_2 + X_1D_2)$ is an M-matrix, and thus $Odiag\ X_1X_2$ may be considered as a perturbation.

On the contrary, the product of an M-matrix by a W-matrix (the product of a W-matrix by an M-matrix) is a monotone matrix with negative diagonal entries and both nonnegative and nonpositive components. Finally, the product of a P-matrix by either an M-matrix or a W-matrix (the product of either an M-matrix or a W-matrix by a P-matrix) is a monotone matrix where the diagonal entries do not have the same sign.

Note that the sign of the diagonal entries of products between two of these matrices in general is not maintained for products of more than two factors, as can be shown by simple examples.

Let us consider the problem of verifying when a given monotone matrix is a product of matrices and of computing its factors. As an example, consider the following case in which the type of factors is known. Let a monotone+ matrix A be expressed as $A = I - B + C = (I - X)(I - Y)$, with $I - X$, $I - Y$ M-matrices, so that

$$B = X + Y, \quad C = XY. \tag{37}$$

Since $0 \leq X, Y \leq B$, a necessary condition for the existence of X and Y is $C \leq B^2$. From (37) it follows that X and Y must satisfy the matrix polynomial equations

$$X^2 - XB + C = 0, \quad Y^2 - BY + C = 0. \tag{38}$$

Let $g(\lambda)$ be defined as

$$g(\lambda) = det(\lambda^2 - \lambda B + C).$$

Then, every solution X, Y to equations (38) satisfies the equation $g(X) = 0$, $g(Y) = 0$ ([**18**], p. 228). This result follows from a generalization of the Hamilton-Cayley Theorem ([**18**], p. 83). Conditions for the existence of solutions to matrix polynomial equations (37) can be found in [**11**, **12**, **30**], while the computation of their solutions in [**3**, **13**, **26**].

For the computation of X and Y, introduced in the previous example, it has been used an empirical procedure. Assume X as the unknown matrix. From (37), by setting $Y = B - X$, it follows that X is a solution to the matrix equation $X(B - X) = C$. Special solutions are obtained under the assumptions that either $x_{ij} = 0$ when $b_{ij} = 0$ or $x_{ii} = 0$ $\forall i$. In these cases the number of unknowns is less than the number of equations. In all the numerical experiments carried out with $n \leq 5$, a unique solution has been obtained.

From the previous reasoning it can be deduced that the problem of stating when a given monotone matrix can be decomposed into a product of monotone matrices and of computing such factors are very difficult tasks, even when the type of factors is known.

6 - Concluding remarks

Basic monotone+ matrices have been defined as monotone+ matrices characterized by only two components in the splitting (1). From Theorem 1 it follows that any basic monotone+ matrix A with *diag* $D \neq 0$ is either an irreducible M-matrix (for which $\rho(A^{-1}C) < 1$) or a W-matrix (for which $\sigma(A, B) > 1$). Necessary and sufficient conditions for the inverse-positivity of matrices of type $A = D - B$ or $A = D + C$, with D, B, C defined as in splitting (1), can be found in Appendix A. They can also be formulated as the conditions (m) and (w) in (2) and (3).

Thus, the class of basic monotone+ matrices consists of irreducible M-matrices (for which $\rho(A^{-1}C) < 1$), W-matrices (for which $\sigma(A, B) > 1$), and matrices $A = B - C$ with B and C defined as in splitting (1) (for which (24) holds). Special monotone+ matrices $A = B - C$ are obtained by taking either $B = P$, $P^{-1}C$ irreducible or $C = P$, $P^{-1}B$ irreducible, where P is a P-matrix with zero diagonal entries:

$$A = -P + C = P(-I + P^{-1}C) \qquad A = -B + P = P(I - P^{-1}B);$$

they are P-transformations of either W-matrices or irreducible M-matrices. Moreover, perturbations of these matrices, preserving the inverse positivity, and defined by $A = -P + C - B_0$, $-B + P + C_0$, with $B_0, C_0 \geq 0$, and *diag* $B_0, C_0 = 0$, have again only two components.

Theorems 1 and 3 also hold when the diagonal matrix D is replaced by a P-matrix in the splitting (1): $A = P - B + C$.

Let a monotone+ matrix A be splitted as in (20) and let the sets Ω and Ω_{11} be defined as in (21) and (22), respectively. If $\Omega \subset \Omega_{11}$ for a monotone+ matrix A, then A may lose the inverse-positivity under the action of small perturbations. Moreover, it follows from Theorems 1 and 3 that the sign of *diag* D of A cannot be precisely specified when $\Omega \subset \Omega_{11}$. Under this condition both $\rho(A^{-1}C) > 1$ and $\sigma(A, B) < 1$ hold. The structures of monotone+ matrices may become more complex than those of the basic monotone+ matrices. In general, a monotone+ matrix A shows all the three components in the splitting (1) and they are no more suitable to describe the structure of A. More involved decompositions or transformations of A are needed to reveal the properties that cause the inverse-positivity. Some situations, which make the analysis tractable, have been described in Subsection 4.2.

An example of application that exploits the presented results is given by a topical problem which concerns modeling and computing the abundance $u(t, x)$ of a population in a vast environment, and in particular the existence of an equilibrium state. The environment Ω shows subregions either favourable or hostile to settling of individuals. Thus, the dispersion process has both random and deterministic components. The evolution balance equation can be written in the form

$$\frac{\partial u}{\partial t} - \nabla \cdot d(x)\nabla u - \int_{\Omega} K(x, y)u(t, y)dy = f(t, x),$$

$$+ \textit{ boundary and initial conditions},$$

where $d(x) > 0$ is a random dispersion coefficient, $f(t, x) > 0$ is an immigration flux from the external environment, and $K(x, y) = pK_+(x, y) - qK_-(x, y)$, with $K_\pm(x, y) \geq 0$, and p, q are positive control parameters, representing the strength of the attractive and repelling actions of the environment, respectively. Birth and death processes are included in the term $K(x, x)u(t, x)$.

Let the discrete numerical approximation of this integro-differential problem be obtained by means of the finite volume method ([**36**], p. 161, p. 250). The positivity of the solutions of both steady states and time dependent problems depend on the positivity of the inverse of certain matrices. The ranges of the parameters p and q should be determined so that such condition is satisfied.

Acknowledgments. I wish to thank the Referee for his precise criticism and suggestions. Hearty thanks to A. Aimi, M. Groppi, M. T. Mauro, M. Diligenti, University of Parma (Italy), S. Pasquali, CNR-IMATI Milano (Italy), and S. Ricci, Fivizzano (Italy). The work was performed in the frame of the activities sponsored by CNR-IMATI Milano.

References

[**1**] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical science*, SIAM, Philadelphia, 1994.

[**2**] A. Berman and R. J. Plemmons, *Eight types of matrix monotonicity*, Linear Algebra and Appl. **13** (1976), 115–123.

[**3**] D. A. Bini, G. Latouche and B. Meini, *Solving matrix polynomial equations arising in queueing problems*, Linear Algebra Appl. **340** (2002), 225–244.

[**4**] F. Bouchon, *Monotonicity of some perturbations of irreducibly diagonally dominant M-matrices*, Numer. Math. **105** (2007), 591–601.

[**5**] J. H. Bramble and B. E. Hubbard, *New monotone type approximations for elliptic problems*, Math. Comp. **18** (1964), 349–367.

[**6**] J. H. Bramble and B. E. Hubbard, *On a finite difference analogue of an elliptic boundary problem which is neither diagonally dominant nor of non-negative type*, J. Math. and Phys. **43** (1964), 117–132.

[**7**] G. Buffoni and A. Galati, Matrici essenzialmente positive con inversa positiva, Boll. Un. Mat. Ital. (4) **10** (1974), 98–103.

[**8**] G. Buffoni, *Nonnegative and skew-symmetric perturbations of a matrix with positive inverse*, Math. Comp. **54** (1990), 189-194.

[**9**] G. Buffoni, Perturbation of a matrix with positive inverse, Riv. Mat. Univ. Parma (4) **16** (1990), 251–262.

[**10**] L. Collatz, *Aufgaben monotoner Art*, Arch. Math. **3** (1952), 366–376.

[**11**] J. E. Dennis, Jr., J. F. Traub and R. P. Weber, *On the matrix polynomial, lambda-matrix and block eigenvalue problems*, Computer Science Technical Reports, Cornell Univ., Ithaca, N.Y., and Carnegie-Mellon Univ., Pittsburgh, PA, 1971

[**12**] J. E. Dennis, Jr., J. F. Traub and R. P. Weber, *The algebraic theory of matrix polynomials*, SIAM J. Numer. Anal. **13** (1976), 831–845.

[**13**] J. E. Dennis, Jr., J. F. Traub and R. P. Weber, *Algorithms for solvents of matrix polynomials*, SIAM J. Numer. Anal. **15** (1978), 523–533.

[**14**] K. Fan, *Topological proofs for certain theorems on matrices with non-negative elements*, Monatsh. Math. **62** (1958), 219–237.

[**15**] L. Farina and S. Rinaldi, *Positive linear systems: Theory and applications*, Wiley-Interscience, New York, 2000.

[**16**] M. Fiedler and V. Pták, *On matrices with non-positive off-diagonal elements and positive principal minors*, Czechoslovak Math. J. **12** (1962), 382–400.

[**17**] G. Frobenius, *Über Matrizen aus nicht negativen Elementen*, S.-B. Preuss Akad. Wiss. Berlin (1912), 456–477.

[**18**] F. R. Gantmacher, *The theory of matrices, Vol. I*, Chelsea Publishing Co., New York, 1959.

[**19**] F. R. Gantmacher, *The theory of matrices, Vol. II*, Chelsea Publishing Co., New York, 1959.

[**20**] A. V. Gavrilov, *A sufficient condition for the monotonicity of a positive-definite matrix*, Comput. Math. Math. Phys. **41** (2001), 1237–1238.

[**21**] F. Goldberg, *On monotonicity-preserving perturbations of M-matrices*, arXiv:1308-0844, preprint, 2013.

[22] R. D. Haynes, M. R. Trummer and S. C. Kennedy, *Persistently positive inverses of perturbed M-matrices*, Linear Algebra Appl. **422** (2007), 742–754.

[23] J. Huang and T.-Z. Huang, *The inverse positivity of perturbed tridiagonal M-matrices*, Linear Algebra Appl. **434** (2011), 131–143.

[24] J. Huang, R. D. Haynes and T.-Z. Huang, *Monotonicity of perturbed tridiagonal M-matrices*, SIAM J. Matrix Anal. Appl. **33** (2012), 681–700.

[25] T. Ikeda, *Maximum principle in finite element models for convection-diffusion phenomena*, Mathematics Studies, North-Holland, Amsterdam, 1983.

[26] W. Kratz and E. Stickel, *Numerical solution of matrix polynomial equations by Newton's method*, IMA J. Numer. Anal. **7** (1987), 355–369.

[27] S. C. Kennedy and R. D. Haynes, *Inverse positivity of perturbed tridiagonal M-matrices*, Linear Algebra Appl. **430** (2009), 2312–2323.

[28] M. A. Krasnoselskii, *Positive solutions of operator equations*, Noordhoff, Groningen, 1964.

[29] G. I. Marčuk, *Metodi del calcolo numerico*, Editori Riuniti, Roma, 1984.

[30] T. Netzer and A. Thom, *About the solvability of matrix polynomial equations*, Bull. Lond. Math. Soc. **49** (2017), 670–675.

[31] J. M. Ortega and W. C. Rheinboldt, *Monotone iterations for nonlinear equations with application to Gauss-Seidel methods*, SIAM J. Numer. Anal. **4** (1967), 171–190.

[32] A. Ostrowski, *Über die determinanten mit überwiegender Hauptdiagonale*, Comment. Math. Helv. **10** (1937), 69–96.

[33] J. R. Rice, *Numerical methods, software and analysis*, McGraw-Hill, New York, 1983.

[34] E. E. Tyrtyshnikov, *A brief introduction to numerical analysis*, Birkhäuser, Boston, MA, 1997.

[35] R. S. Varga, *Factorization and normalized iterative methods*, Boundary problems in differential equations, R. E. Langer, ed., Univ. of Wisconsin Press, Madison, Wis., 1960, 121–142.

[36] R. S. Varga, *Matrix iterative analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962.

[37] T. Vejchodský, *Necessary and sufficient condition for the validity of the discrete maximum principle*, Proceedings of Algoritmy 2012, 1–10.

Appendix A

The main properties of M-matrices and W-matrices are here recalled.

Let $A_i,\ i = 1, 2, ..., n$, be the $(n-1)\times(n-1)$ principal square submatrix of A obtained by eliminating from A the $i-th$ row and the $i-th$ column, and let

μ_{ij}, $j = 1, 2, ..., n-1$, its eigenvalues, with $\mu_{i1} = min_j Re(\mu_{ij})$ for M-matrices and $\mu_{i1} = max_j Re(\mu_{ij})$ for W-matrices.

Fifty different but equivalent characterizations of nonsingular M-matrices are given by Berman and Plemmons ([**1**], p. 134), together with necessary and sufficient conditions for an arbitrary real matrix to be a nonsingular M-matrix ([**1**], p. 140).

Equivalent necessary and sufficient conditions for an essentially positive matrix A to have a positive inverse are given in [**7**, **14**]. The conditions given in [**14**] (Theorem 6, conditions $(b), (c'')$) are

$$(b) \quad \exists\, \gamma > 0 : A^T \gamma > 0, \qquad (c'') \quad \forall i : \mu_{i1} < 0, \tag{39}$$

while those in [**7**] are

$$\exists\, i_0 : \mu_{i_0 1} < 0, \qquad \forall\, \forall i : \frac{detA_i}{detA} > 0. \tag{40}$$

It is possible to prove that the weaker conditions (39) with $\mu_{i1} \leq 0$, and (40) with $detA_i/detA \geq 0$ are necessary and sufficient for an essentially nonnegative matrix to have a nonnegative inverse.

Table 2. *Some basic properties of irreducible M-matrices and W-matrices. Elements are the entries of the matrix A, A_i is any principal square submatrix of A of order -1, λ_j and μ_{ij} are the eigenvalues of A and A_i, J and J_i are the Jacobi matrices sociated with A and A_i.*

$A \in$ {irreducible M-matrices}	$A \in$ {W-matrices}
$a_{ii} > 0,\, a_{ij} \leq 0,\ i \neq j$	$a_{ii} < 0,\, a_{ij} \geq 0,\ i \neq j$
$det\ A > 0$	$(-1)^{n-1}\ det\ A > 0$
$\lambda_1 < Re(\lambda_j) < \lvert\lambda_j\rvert,\ j > 1$	$\lvert\lambda_j\rvert > \lambda_1 > Re(\lambda_j),\ j > 1$
$\rho(J) < 1$	$\rho(J) > 1$
$A_i^{-1} > 0$	$A_i^{-1} \leq 0$
$0 < \mu_{i1} \leq Re(\mu_{ij}),\, \forall\, i, j$	$0 > \mu_{i1} \geq Re(\mu_{ij}),\, \forall\, i, j$
$\rho(J_i) < 1$	$\rho(J_i) < 1$

Some basic properties of irreducible M-matrices and W-matrices are summarized in Table 2, where $\lambda_j,\ j = 1, 2, ..., n$, are the eigenvalues of A, with $\lambda_1 = min_j\ |\lambda_j| > 0$, and J is the point Jacobi matrix ([**36**], p. 57) associated with A. It is defined by $J = I - D^{-1}A = D^{-1}B$ for an M-matrix $A = D - B$, and by $J = I + |D|^{-1}A = |D|^{-1}C$ for a W-matrix $A = D + C$. The point Jacobi matrices J_i associated with A_i are defined similarly.

Appendix B

Proposition b1. *Let $P > 0$ and $Q \geq 0 \wedge \neq 0$, be $n \times n$ matrices. Then,*

(i) *both the spectral radius $\rho(PQ)$ and the corresponding eigenvector are positive;*

(ii) *the spectral radius $\rho(QP) = \rho(P^T Q^T)$ is again positive, while the corresponding eigenvector may have some components equal to zero.*

Proposition b2. *Under the assumptions of Proposition* (b1), *let the entries of $P = P(t)$ be strictly increasing* (*decreasing*) *of the real parameter t. Then, $\rho(P(t)Q)$ is strictly increasing* (*decreasing*).

Proposition b3. *Let $||\cdot||$ be any norm of the matrix in argument. Let $A(u)$ an $n \times n$ real matrix, dependent on the real nonnegative parameter u, with $detA(u) \neq 0$ for $0 \leq u < u_0$, and $detA(u_0) = 0$. Then $||A^{-1}(u)|| \to +\infty$ as $u \to u_0$. Consequently, at least one entry of A^{-1} becomes ∞ as $u \to u_0$.*

Proof. This statement follows from $1/|\lambda(u)| \leq ||A^{-1}(u)||$, with $\lambda(u)$ any eigenvalue of $A(u)$. In particular, let $A(u) = I - uM$, where $M \geq 0,\ M \neq 0$, $\rho(M) > 0$. Then $||(I - uM)^{-1}|| \to +\infty$ as $u \to 1/\rho(M)$. In fact, the series $I + uM + u^2M^2 + ...$ is divergent. □

Proposition b4. *Consider a batch of polynomials*

$$p(s, t) = \gamma_0 s^n + \gamma_1(t)s^{n-1} + ... + \gamma_{n-1}(t)s + \gamma_n(t),$$

where $\gamma_0 = constant$, $\gamma_1(t) = \gamma_{10}t + \gamma_{11}$, $\gamma_2(t) = \gamma_{20}t^2 + \gamma_{21}t + \gamma_{22}, ...$,

$$\gamma_n(t) = \gamma_{n0}t^n + \gamma_{n1}t^{n-1} + ... + \gamma_{nn-1}t + \gamma_{nn},$$

with $\gamma_{kl} = constant$ ($k = 1, 2, ..., n,\ l = 0, 1, ..., k$).

Let $s^*(t)$ be a solution to $p(s, t) = 0$. If there exists a solution $s^*(t) = s^*_0$ independent of t for $t_0 \leq t \leq t_1$, then s^*_0 is also a solution to $p(s, t) = 0$ for $t > t_1$.

P r o o f. The polynomial $p(s,t)$ may be written as

$$p(s,t) = p_0(s) + p_1(s,t),$$

where $p_0(s) = \gamma_0 s^n + \gamma_{11} s^{n-1} + ... + \gamma_{n-1\, n-1} s + \gamma_{nn}$, and

$$p_1(s,t) = \hat{\gamma}_0 t^n + \hat{\gamma}_1(s) t^{n-1} + ... + \hat{\gamma}_{n-1}(s) t,$$

with $\hat{\gamma}_k(s)$ polynomial in s of degree k.

If $p(s_0^*, t) = 0$ for $t_0 \le t \le t_1$, then $\partial p_1(s_0^*, t)/\partial t = 0$. Thus, s_0^* is a solution to $p_0(s) = 0$, and either all the coefficients of the polynomial $\hat{\gamma}_k(s)$ are zero, or s_0^* is also a solution to $\hat{\gamma}_k(s) = 0$. The conditions $p_0(s_0^*) = 0$, $\hat{\gamma}_0 = 0$, $\hat{\gamma}_k(s_0^*) = 0$, $k = 1, , 2, ...n-1$, are independent of t, so that s_0^* is also a solution to $p(s,t) = 0$ for $t > t_1$. □

Appendix C

Examples of 3×3 matrices illustrating the limit cases when either $\rho(A^{-1}C) = 1$ or $\sigma(A,B) = 1$ considered in Section 3 are reported. Here $\delta = det\ A$ and $\lambda(A^{-1}C) =$ eigenvalues of $A^{-1}C$.

(e1) See Theorem 2, part (i): A is given by a singular matrix $D - B$, with $diag\ D > 0$ and $\rho(D^{-1}B) = 1$, plus a nonnegative matrix C:

$$A = \begin{vmatrix} 1 & -1 & p \\ -1 & 2 & -1 \\ q & -1 & 1 \end{vmatrix}$$

$$A^{-1} = \frac{1}{\delta} \begin{vmatrix} 1 & 1-p & 1-2p \\ 1-q & 1-pq & 1-p \\ 1-2q & 1-q & 1 \end{vmatrix}$$

where

$$0 < p, q < 1/2, \quad \delta = p + q - 2pq > 0, \quad \lambda(A^{-1}C) = -2pq/\delta,\ 0,\ 1,$$
$$\rho(A^{-1}C) = 1, \quad \sigma(A,B) = min(1 - \sqrt{2p},\ 1 - \sqrt{2q}) < 1.$$

(e2) See Theorem 2, part (ii), and Remark (r10), case (a): A is given by a P-matrix $D + C$, with $diag\ D \ge 0 \wedge \not> 0$, minus a nonnegative matrix B:

$$A = \begin{vmatrix} 1 & -p & -q \\ -p & 0 & 1 \\ -q & 1 & 0 \end{vmatrix}$$

$$A^{-1} = \frac{-1}{\delta} \begin{vmatrix} 1 & q & p \\ q & q^2 & 1-pq \\ p & 1-pq & p^2 \end{vmatrix}$$

where

$$0 < p, q < 1/\sqrt{2}, \quad \delta = -1 + 2pq < 0, \quad \lambda(A^{-1}C) = 0,\ 1,\ -1/\delta,$$
$$\rho(A^{-1}C) = -1/\delta > 1, \quad \sigma(A, B) = 1.$$

(e3) See Theorem 2, part (ii), and Remark (r10), case (b): A is given by an essentially positive matrix $D + C$, with *diag* $D \leq 0 \wedge \nleq 0$ and nonnegative inverse, minus a nonnegative matrix B:

$$A = \begin{vmatrix} -1 & 1 & -p \\ 1 & -2 & 1 \\ 1 & -q & 0 \end{vmatrix}$$

$$A^{-1} = \frac{1}{\delta} \begin{vmatrix} q & pq & 1-2p \\ 1 & p & 1-p \\ 2-q & 1-q & 1 \end{vmatrix}$$

where

$$0 < p < 1/2, \quad 0 < q < (1-2p)/(1-p) < 1, \quad \delta = 1 - 2p - q + pq > 0,$$
$$\rho(A^{-1}C) > (1 - p + 2pq)/\delta > 1, \quad \sigma(A, B) = 1.$$

Now examples illustrating the case $\sigma(A, B) > 1$ (*diag* $D < 0$) considered in Section 3 are presented.

(e4) See Theorem 1, part (ii): A is given by a W-matrix $D+C$ perturbed by a nonnegative matrix B. Let the matrix $A+sB-tC = D-(1-s)B+(1-t)C$ and its inverse $Z(s, t)$ be given by

$$A + sB - tC = \begin{vmatrix} -1 & q(1-t) & -p(1-s) \\ q(1-t) & -1 & q(1-t) \\ q(1-t) & 0 & -1 \end{vmatrix}$$

$$Z(s,t) =$$
$$\frac{1}{\delta} \begin{vmatrix} 1 & q(1-t) & q^2(1-t)^2 - p(1-s) \\ q(1-t)(1+q(1-t)) & 1 + pq(1-s)(1-t) & q(1-t))(1-p(1-s)) \\ q(1-t) & q^2(1-t)^2 & 1 - q^2(1-t)^2 \end{vmatrix}$$

where $p, q > 0$ and

$$\delta(s,t) = det(A + sB - tC) = -(1 + pq(1-s)(1-t)) + q^2(1-t)^2(1 + q(1-t))$$

Under the assumptions $0.76 < q < 1$ and $0 \le p < -1/q + q + q^2$ it follows that $\delta(0,0) > 0$ and $A^{-1} > 0$. From the equation $\delta(s,t) = 0$ an expression for $t^*(s)$ cannot be obtained. Nevertheless, an expression for $(t^*)^{-1}(t)$ (Figure 2), for which $s > (t^*)^{-1}(t) \Longrightarrow \delta(s,t) > 0$, can be obtained, and it is given by

$$(t^*)^{-1}(t) = 1 + \frac{1}{pq(1-t)} - \frac{1}{p} q(1-t)(1 + q(1-t)), \quad t > t^*(0) = \frac{1}{\rho(A^{-1}C)}.$$

Moreover, $s^*(t)$ (Figure 2) is obtained by imposing the positivity of the entries of $Z(s,t)$,

$$s^*(t) = 1 + \frac{1}{pq(1-t)}.$$

For $p = 0.3$, $q = 0.9$ it results: $\delta(0,0) = 0.269$, $t^*(0) = 1/\rho(A^{-1}C) = 0.0816$, $s^*(0) = \sigma(A,B) = 4.704$.

Remark. In this example $s^*(t)$ is strictly increasing with t, with $\lim_{t\to 1} s^*(t) = +\infty$. If in the matrix A of this example: $a_{31} = -p$, $a_{32} = q < 1$, then $s^*(t) = constant = \sigma(A,B) = 1 + 1/p^2$. When $p = 0$ it follows $\sigma(A,B) = +\infty$.

Examples illustrating the cases $\rho(A^{-1}C) > 1$ and $\sigma(A,B) < 1$ considered in Section 4 follow.

(e5) See Subsection 4.1, Theorem 3, part (i): A is characterized by $diag\ D > 0$, B irreducible, $\rho(D^{-1}B) > 1$. Let the matrix $A + sB - tC = D - (1-s)B + (1-t)C$ and its inverse $Z(s,t)$ be given by

$$A + sB - tC = \begin{vmatrix} 1 & -(1-s) & p(1-t) \\ -(1-s) & 1.5 & -(1-s) \\ 0 & -(1-s) & 1 \end{vmatrix}$$

$$Z(s,t) = \frac{1}{\delta} \begin{vmatrix} 1.5 - (1-s)^2 & (1-s)(1-p(1-t)) & (1-s)^2 - 1.5p(1-t) \\ 1-s & 1 & (1-s)(1-p(1-t)) \\ (1-s)^2 & 1-s & 1.5 - (1-s)^2 \end{vmatrix}$$

where $p > 0$ and $\delta(s,t) = det(A + sB - tC) = 1.5 - (1-s)^2(2 - p(1-t))$.

When $p = 0$: $\delta(0,0) = -0.5$, $\rho(D^{-1}B) = 2/\sqrt{3} = 1.155$, $(D-B)^{-1} < 0$, $(D - (1-s)B)^{-1} > 0$ for $s > \hat{s} = 1 - 0.5\sqrt{3} = 0.134 < s < 1$.

The function $t^*(s)$ is obtained from the equation $\delta(s,t) = 0$ ($\delta(s,t) > 0$ for $0 \le t < t^*(s)$), and $s^*(t)$ by imposing the positivity of the entries of $Z(s,t)$ (Figure 2):

$$t^*(s) = 1 - \frac{2}{p} + \frac{1.5}{p(1-s)^2}, \quad for \quad 0 \le s < 1,$$

$$s^*(t) = 1 - \sqrt{1.5p(1-t)} \quad for \quad 0 \le t \le 1; \quad s^*(t) = 1 \; for \; t \ge 1.$$

In order to have $t^*(0) > 0$ and $s^*(0) > 0$ it must be $1/2 < p < 2/3$. Then, $t^*(0) = 1/\rho(A^{-1}C) = (p-0.5)/p < 1$, $t^*(\hat{s}) = 1$, $\lim_{s\to 1} t^*(s) = +\infty$, and $s^*(0) = \sigma(A,B) = 1 - \sqrt{1.5p} < \hat{s} < 1$.

(e6) See Subsection 4.1, Theorem 3, part (ii): A is characterized by $diag\ D < 0$, C irreducible, $\rho(D^{-1}C) > 1$, $(D+C)^{-1} \not> 0$. Let the 4×4 matrix A be written as

$$A = \begin{vmatrix} M & U \\ U^T & -1 \end{vmatrix}$$

where

$$M = \begin{vmatrix} -1 & a & -p \\ a & -1 & a \\ a & 0 & -1 \end{vmatrix}$$

$U = b(1,1,1)^T$, and $a, b, p \ge 0$. Its inverse is given by

$$A^{-1} = \frac{1}{k} \begin{vmatrix} kM^{-1} + M^{-1}UU^TM^{-1} & -M^{-1}U \\ -U^TM^{-1} & 1 \end{vmatrix}$$

where $k = -1 - U^TM^{-1}U = det\ A/det\ M$ and

$$M^{-1} = \frac{1}{\delta} \begin{vmatrix} 1 & a & a^2 - p \\ a + a^2 & 1 + ap & a(1-p) \\ a & a^2 & 1 - a^2 \end{vmatrix}$$

with $\delta = \delta(a,p) = -1 + a^2 + a^3 - ap$.

When $p = 0$ and $a_0 \simeq 0.755 < a < 1$ ($\delta(a_0, 0) = 0$) it results $\delta(a,0) > 0$ and $M^{-1} > 0$. Thus, $A^{-1} = (D+C)^{-1} \not> 0$ for any $b > 0$.

For $a_0 < a < 1$ and $-1/a + a + a^2 < p < a^2$ it holds $\delta(a,p) < 0$ and $M^{-1} < 0$. Thus, $A^{-1} = (D - B + C)^{-1} > 0$ for b in a suitable interval.

With $a = 0.9$, $b = 0.2$ the limits for p are $-1/a + a + a^2 = 0.599 < p < a^2 = 0.81$. With $p = 0.7$ the range of b to have $A^{-1} > 0$ is $\simeq 0.12 \le b \le\simeq 0.22$. With $b = 0.2$: $\delta(a,p) = -0.091$ and $A^{-1} > 0$. Moreover, $\sigma(A,B) \simeq 0.1445$,

$\rho(A^{-1}C) = 18.313$, $\rho(D^{-1}C) = 1.284$. At the boundary $s = 1$ we have that $Z(1,t) = (D + (1-t)C)^{-1} > 0$ for $0.162 < t < 0.221 = 1 - 1/\rho(D^{-1}C)$.

(e7) See Subsection 4.2, item (II): A is composed by $P - B_0$ with P a P-matrix, $B_0 \geq 0$, $\rho(P^{-1}B_0) < 1$, perturbed by $C_0 \geq 0$. Let A be given by

$$A = P - B_0 + C_0 = \begin{vmatrix} 2 & q & -1 \\ -1 & -1 & p \\ q & p & -1 \end{vmatrix}$$

where $p, q > 0$, P is the symmetric P-matrix with non-zero entries $p_{*11} = 2$, $p_{*23} = p_{*32} = p$, and $B_0, C_0 \geq 0$. It follows

$$P^{-1}B_0 = \begin{vmatrix} 0 & 0 & 1/2 \\ 0 & 0 & 1/p \\ 1/p & 1/p & 0 \end{vmatrix}$$

which is singular, irreducible (the directed graph is strongly connected), and its eigenvalues are given by $\lambda(P^{-1}B_0) = 0, \pm\sqrt{1+0.5p}/p$. Thus, $\rho(P^{-1}B_0) < 1$ and $(P - B_0)^{-1} > 0$ for $p > 0.5(1+\sqrt{17}) \simeq 2.562$.

Taking $p = 2$, $q = 0.4$ one has: $A^{-1} > 0$, $\rho(A^{-1}C) \simeq 2.431$, $\sigma(A,B) \simeq 0.1056$, $\rho(A^{-1}C_0) \simeq 0.286$, $\rho(P^{-1}B_0) = 1/\sqrt{2} \simeq 0.707$, $\sigma(P - B_0, C_0) = 1.249$.

(e8) See Subsection 4.2, item (II): A is composed by $-P + C_0$ with P a P-matrix, $C_0 \geq 0$, $\rho(P^{-1}C_0) > 1$, $(-P + C_0)^{-1} > 0$, perturbed by $-B_0 \leq 0$.

Let A be given by

$$A = -P + C_0 - B_0 = \begin{vmatrix} 1 & 1 & -2 \\ -1 & 0 & 1 \\ 3 & -1 & -1 \end{vmatrix}$$

where P is the P-matrix with non-zero entries $p_{*13} = 2$, $p_{*21} = p_{*32} = 1$. We obtain $A^{-1} > 0$, $\rho(A^{-1}C) = 7.91$, $\sigma(A,B) = 0.2928$, $\sigma(A,B_0) = 3$, $(-P + C_0)^{-1} > 0$, $\rho(P^{-1}C_0) = 1.29$, $\rho((-P + C_0)^{-1}B_0) = 0.5$.

(e9) See Subsection 4.2, item (IV): the inverse of the principal square submatrix A_3 of A is positive. Let the matrix $A + sB - tC = D - (1-s)B + (1-t)C$ and its inverse $Z(s,t)$ be given by

$$A + sB - tC = \begin{vmatrix} 2 & -(1-s) & p(1-t) \\ -(1-s) & 1 & -(1-s) \\ -(1-s) & q(1-t) & -0.1 \end{vmatrix}$$

$$Z(s,t) = \frac{1}{\delta}\begin{vmatrix} -0.1+q(1-s)(1-t) & 0.1(1-s)+pq(1-t)^2 & (1-s)^2-p(1-t) \\ (1-s)(1.1-s)] & -0.2+p(1-s)^2(1-t) & (1-t)[2-p(1-t)] \\ (1-s)[1-q(1-t)] & (1-s)^2-2q(1-t) & 2-(1-s)^2 \end{vmatrix}$$

where

$$\delta(s,t) = det(A+sB-tC) = (1-s)\left[\phi(s)+(p+2q)(1-t)-pq(1-t)^2\right],$$

with

$$\phi(s) = \frac{-0.2}{1-s} + 0.1(1-s) - (1-s)^2 < 0, \quad 0 \le s < 1.$$

Under the assumptions $0.2 < p < 1$, $(1.1-p)/(2-p) < q < 0.5$ it results $\delta(0,0) > 0$ and $A^{-1} > 0$. The expression of $t^*(s)$ is obtained from the equation $\delta(s,t) = 0$:

$$t^*(s) = 1 - \frac{p+2q}{2pq}\left[1 - \sqrt{1 + \frac{4pq}{(p+2q)^2}\phi(s)}\,\right].$$

Moreover $s^*(t)$ is given by

$$s^*(t) = 1 - \sqrt{p(1-t)}.$$

For $p = 0.9$ and $q = 0.4$ it follows: $\delta(0,0) = 0.24$, $t^*(0) = 1/\rho(A^{-1}C) = 0.226$, $s^*(0) = \sigma(A,B) = 0.0513$. $Z(s,t) > 0$ for $(t^*)^{-1}(t) < s < s^*(t)$: the two curves $(t^*)^{-1}(t)$ and $s^*(t)$ intersect at a point $\hat{t} \simeq 0.6$, $(t^*)^{-1}(\hat{t}) = s^*(\hat{t}) \simeq 0.4$ (see Figure 2).

(e10) See Subsection 4.2, item (IV): the inverse of the principal square submatrix A_3 of A is negative. Let the matrix $A+sB-tC = D-(1-s)B+(1-t)C$ and its inverse $Z(s,t)$ be given by

$$A+sB-tC = \begin{vmatrix} -1 & 1-t & -p(1-s) \\ 1-t & -2 & 1-t \\ 1-t & -p(1-s) & 0.2 \end{vmatrix}$$

$$Z(s,t) = \frac{1}{\delta}\begin{vmatrix} -0.4+p(1-s)(1-t) & -0.2(1-t)+p^2(1-s)^2 & (1-t)^2-2p(1-s) \\ (1-t)[-0.2+(1-t)] & -0.2+p(1-s)(1-t) & (1-t)[1-p(1-s)] \\ (1-t)[-p(1-s)+2] & -p(1-s)+(1-t)^2 & 2-(1-t)^2 \end{vmatrix}$$

where

$$\delta(s,t) = det(A + sB - tC)$$
$$= p^2(1-s)^2(1-t) - 3p(1-s)(1-t) - 0.2(1-t)^2 + (1-t)^3 + 0.4.$$

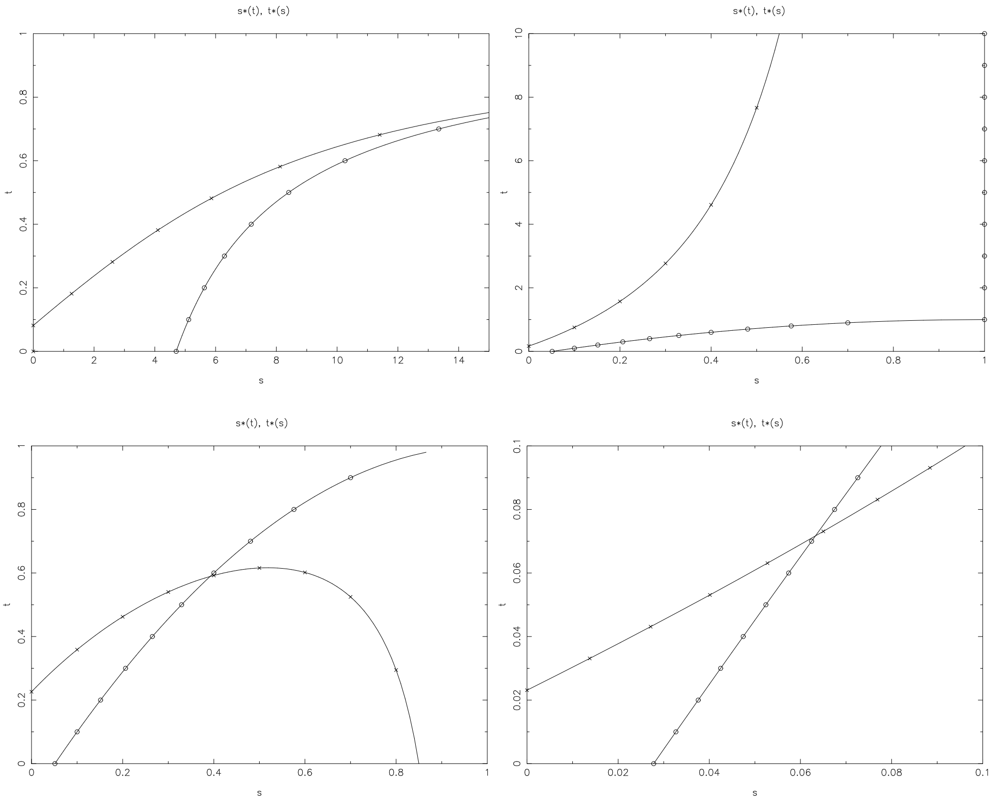

Fig. 2. $s^*(t)$ (circles) and $t^*(s)$ (crosses) for Examples (4) (top left), (5) (top right), (9) (bottom left), (10) (bottom right)

Under the assumption $\sqrt{0.2} = 0.448 < p < 1.5 - \sqrt{1.05} = 0.475$ it results $\delta(0,0) > 0$. As in the Example (4), $(t^*)^{-1}(t)$ is obtained from the equation $\delta(s,t) = 0$:

$$(t^*)^{-1}(t) = 1 - \frac{\phi(t)}{p}, \quad t > t^*(0) = \frac{1}{\rho(A^{-1}C)}$$

where

$$\phi(t) = 0.5\left(3 - \sqrt{9 - 1.6/(1-t) + 0.8(1-t) - 4(1-t)^2}\right).$$

On the other side $s^*(t)$ is given by

$$s^*(t) = 1 - \frac{\sqrt{0.2(1-t)}}{p}.$$

$t^*(0) = 1/\rho(A^{-1}C)$ is obtained from the equation $1 - \phi(t)/p = 0$, $s^*(0) = \sigma(A, B) = 1 - \sqrt{0.2}/p < 1$.

For $p = 0.46$: $t^*(0) = 0.0231$, $s^*(0) = 0.028$. $Z(s,t) > 0$ for $(t^*)^{-1}(t) < s < s^*(t)$: the two curves $(t^*)^{-1}(t)$ and $s^*(t)$ intersect at a point $\hat{t} \simeq 0.07$, $(t^*)^{-1}(\hat{t}) = s^*(\hat{t}) \simeq 0.062$ (see Figure 2).

Giuseppe Buffoni
CNR - IMATI
Via Bassini 15
20133 Milano, Italy
e-mail: giuseppe.buffoni9av2@alice.it