The history of variation diminution

Projesh Nath Choudhury
Indian Institute of Technology Gandhinagar

(Joint with Shivangi Yadav)

Applied Matrix Positivity-II
ICMS, Edinburgh

November 05, 2024

TP and TN matrices

Definition

Suppose A is a (possibly rectangular) real matrix.

- We say A is *totally positive of order* k, denoted by TP$_k$, if all minors of size at most k are positive.

  (Similarly, *totally non-negative of order* k (TN$_k$).)

- We say A is *totally positive* (TP) (*totally non-negative* (TN)) if A is TP$_k$ (TN$_k$) for all $k \geq 1$.

These matrices have featured in diverse areas in mathematics, including:

- analysis
- matrix theory
- Gabor analysis
- integrable systems
- differential equations
- probability/statistics
- approximation theory
- cluster algebras
- combinatorics
- representation theory

Properties of TP matrices: 1. Spectra and minors

Similar to definition of positive (semi)definite matrices – but now,
(a) the matrix need not be symmetric (or square);
(b) uses all minors instead of principal minors.

Positive definite matrices have other characterizations. E.g.:

1. All eigenvalues (of all principal submatrices) are positive. Similarly:

Theorem (Gantmacher–Krein, *Compos. Math.* 1937)

Given a square matrix $A \in \mathbb{R}^{n \times n}$, the following statements are equivalent:

(a) *The matrix A is totally positive (totally non-negative).*

(b) *Every square submatrix of A has positive (non-negative) eigenvalues.*

2. All leading principal submatrices have positive determinant. Similarly:

Theorem (Fekete, *Rend. Circ. Mat.* 1912; Schoenberg, *Ann. of Math.* 1955)

Let $m, n \geq k \geq 1$ be integers. Then $A \in \mathbb{R}^{m \times n}$ is TP$_k$ if and only if all contiguous submatrices of A of size at most k have positive determinant.

Properties of TP matrices: 2. (Whitney) density

3. Recall: every positive semidefinite (PSD) matrix A can be approximated by a sequence of positive definite matrices:

$$A + \epsilon \mathrm{Id}, \ \epsilon \to 0^+.$$

A similar density result for TN_k matrices was discovered by Whitney:

Theorem (Whitney, *J. d'Analyse Math.* 1951)

Given integers $m, n \geq k \geq 1$, the set of $m \times n$ TP_k matrices is dense in the set of $m \times n$ TN_k matrices.

4. PSD matrices form a closed set (under entrywise limits).
   Similarly, so do the TN_k $m \times n$ matrices.

5. PSD matrices are not closed under products – but TN_k matrices are:

Lemma (Cauchy–Binet formula)

If A, B are TN_k matrices, and AB is defined, then AB is TN_k as well.

Examples of TP/TN matrices: 1.

- *Generalized Vandermonde matrices* are TP:
  If $0 < x_1 < x_2 < \cdots < x_m$ and $\alpha_1 < \cdots < \alpha_n$ are real numbers, then the $m \times n$ matrix $(x_j^{\alpha_k})$ is TP.

- The *Gaussian kernel* is TP on $\mathbb{R} \times \mathbb{R}$: Given $\sigma > 0$ and scalars

$$x_1 < \cdots < x_n, \qquad y_1 < \cdots < y_n,$$

  the matrix $G[\mathbf{x} : \mathbf{y}] = (e^{-\sigma(x_i - y_j)^2})_{i,j=1}^n$ is TP.

  Proof:

$$G[\mathbf{x}; \mathbf{y}] = \mathrm{diag}(e^{-\sigma x_i^2})_{i=1}^n \cdot ((e^{2\sigma x_i})^{y_j})_{i,j=1}^n \cdot \mathrm{diag}(e^{-\sigma y_j^2})_{j=1}^n,$$

  and the second matrix is a generalized Vandermonde matrix. Thus $\det G[\mathbf{x}; \mathbf{y}] > 0$. □

- *Toeplitz cosine matrices* are TN:

$$C(\theta) := \cos((j-k)\theta)_{j,k=1}^n, \quad \text{where} \quad \theta \in [0, \tfrac{\pi}{2(n-1)}].$$

Examples of TP/TN matrices: 2. PF sequences

'Historical' example: *one-sided Pólya frequency (PF) sequences*:

These are real sequences $\mathbf{a} = (a_0, a_1, a_2, \dots)$ such that the following semi-infinite Toeplitz matrix is TN:

$$T_{\mathbf{a}} := \begin{pmatrix} a_0 & 0 & 0 & \cdots \\ a_1 & a_0 & 0 & \cdots \\ a_2 & a_1 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Also associate to $\mathbf{a}$ its *generating function* $\Psi_{\mathbf{a}}(x) := a_0 + a_1 x + a_2 x^2 + \cdots$.

Examples:

- $\mathbf{a} = (1, \alpha, 0, 0, \dots)$ for $\alpha > 0$.
- If $\mathbf{a}, \mathbf{b}$ are one-sided PF sequences, then $T_{\mathbf{a}} T_{\mathbf{b}} = T_{\mathbf{c}}$ for some $\mathbf{c}$. Moreover:
  - $\mathbf{c}$ is also a one-sided PF sequence (Cauchy–Binet);
  - $\Psi_{\mathbf{a}}(x)\Psi_{\mathbf{b}}(x) = \Psi_{\mathbf{c}}(x)$.

Examples of TP/TN matrices: 2. PF sequences (cont.)

- Combining these facts, every real sequence $\mathbf{a}_n$ with generating function

$$\Psi_{\mathbf{a}_n}(x) = a_0(1+\alpha_1 x)\cdots(1+\alpha_n x), \qquad a_0 \geq 0,\ \alpha_1,\dots,\alpha_n \geq 0$$

  is a one-sided PF sequence, hence yields TN matrices.

- The converse is also true:

Theorem (Schoenberg, *Ann. of Math.* 1955)

These constitute the generating functions of all 'finite' PF sequences.

Examples: Let $a_0 = 1, \gamma > 0$, and set $\alpha_1 = \cdots = \alpha_n = \gamma/n$. Then:

- $\Psi_{\mathbf{a}_n}(x) = (1+\gamma x/n)^n$ generates a (finite) PF sequence.
- Taking limits, $e^{\gamma x}$ generates a (one-sided) PF sequence, i.e., the matrix T_γ is TN: (Fekete, *Rend. Circ. Math.* 1912)

$$T_\gamma := \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ \gamma & 1 & 0 & 0 & \cdots \\ \frac{\gamma^2}{2!} & \gamma & 1 & 0 & \cdots \\ \frac{\gamma^3}{3!} & \frac{\gamma^2}{2!} & \gamma & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Variation diminishing property

Totally non-negative matrices have the variation diminishing (VD) property. These two notions share a common history (briefly discussed here).

Definition

For a vector $0 \neq x \in \mathbb{R}^n$, denote by $S^-(x)$ the number of changes in sign (*variations*) after deleting all zero entries in x. Set $S^-(0) := 0$.

- Variation diminution (VD) simply means: $S^-(Ax) \leq S^-(x)$.
  This term was coined by Pólya (in German: 'variationsvermindernd'), with 'variation' denoting the number of sign changes in a vector/function.

- The study of variations began with Descartes (1600s).
  Laguerre then presented numerous refinements – a sample result:

Theorem (Laguerre, *J. Math. Pures Appl.* 1883)

If $f(x)$ is a (real) polynomial and $\gamma \geq 0$, then the variations $var(e^{\gamma x} f(x))$ in the Maclaurin coefficients of $e^{\gamma x} f(x)$ are non-increasing in $0 \leq \gamma < \infty$, hence are bounded above by $var(f(x)) < \infty$.

History of the variation diminishing property (cont.)

This was reformulated in the language of total positivity (for 'semi-infinite matrices') by Fekete (1912), and proved as follows:

Theorem (Fekete, *Rend. Circ. Mat.* 1912)

Suppose the following real Toeplitz (semi-infinite) matrix is TN:

$$T_{\mathbf{a}} := \begin{pmatrix} a_0 & 0 & 0 & \cdots \\ a_1 & a_0 & 0 & \cdots \\ a_2 & a_1 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Then $S^-(T_{\mathbf{a}}\mathbf{c}) \leq S^-(\mathbf{c})$ *for all* $\mathbf{c} = (c_0, c_1, \dots, c_k, 0, 0, \dots)^T$ *(and all* $k \geq 0$*).*

Now specialize this to Fekete's 'historical' example of $T_{\mathbf{a}} = T_\gamma$ for $e^{\gamma x}$. Letting $f(x) = c_0 + \cdots + c_k x^k$ and $\mathbf{c} = (c_0, c_1, \dots, c_k, 0, 0, \dots)^T$,

$$var(e^{\gamma x} f(x)) = S^-(T_\gamma \mathbf{c}) \leq S^-(\mathbf{c}) = var(f(x)). \quad \square$$

Thus we go from Descartes (1637) to Laguerre (1883) to Fekete–Pólya (1912).

SR and SSR matrices

Theorem (Schoenberg, *Math. Z.* 1930)

If $A \in \mathbb{R}^{m \times n}$ is TN, then $S^-(Ax) \leq S^-(x)$ for all $x \in \mathbb{R}^n$.

- Clearly, so does $-A$. Characterize the matrices having this property?
- Achieved by Motzkin in PhD thesis (1936) $\rightsquigarrow$ sign-regular matrices.

Definition

- A is said to be *strictly sign-regular* (SSR) if there exists a sequence of signs $\epsilon_k \in \{\pm 1\}$ such that every $k \times k$ minor has sign ϵ_k, for all $k \geq 1$.
  Example: If $\epsilon_k = 1 \ \forall \ k \geq 1$, A is called totally positive (TP).
- If one replaces 'minor' by 'nonzero minor' in the preceding definition, this yields *sign-regular* (SR) matrices (resp. totally non-negative matrices (TN)).

Theorem (Motzkin, 1936)

Suppose $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(A) = n$. Then A is SR if and only if $S^-(Ax) \leq S^-(x)$ for all $x \in \mathbb{R}^n$.

Characterization of SR/TN via VD

Remarkably, Motzkin's thesis also contained:

- Motzkin transposition theorem.
- Fourier–Motzkin Elimination (FME) algorithm.
- A convex polyhedral set is the Minkowski sum of a compact (convex) polytope and a polyhedral cone.

Gantmacher–Krein (1950s) characterized totally positive matrices in terms of the variation diminishing property.

This was later refined – for both TP and TN:

Theorem (Brown–Johnstone–MacGibbon, *J. Amer. Statist. Assoc.* 1981)

Given a real $m \times n$ matrix A, the following statements are equivalent.

1. *A is totally non-negative.*
2. *For all $x \in \mathbb{R}^n$, $S^-(Ax) \leq S^-(x)$. If moreover equality occurs and $Ax \neq 0$, the first (last) nonzero component of Ax has the same sign as the first (last) nonzero component of x.*

Characterization of TP/SSR via VD

Definition: For $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, $S^+(x)$ denotes the maximum number of sign changes in the sequence $x_1, \ldots, x_n$ where zero terms are arbitrarily assigned values $+1$ or -1.

- We set $S^+(0) := n$.
- $S^-(x) \leq S^+(x)$ for all $x \in \mathbb{R}^n$.

Examples:

- $S^-(1, 0, 2, -3, 0, 1) = 2$.
- $S^+(1, 0, 2, -3, 0, 1) = 4$.

Theorem (Brown–Johnstone–MacGibbon, *J. Amer. Statist. Assoc.* 1981)

Given a real $m \times n$ matrix A, the following statements are equivalent.

1. *A is totally positive.*
2. *For all $0 \neq x \in \mathbb{R}^n$, $S^+(Ax) \leq S^-(x)$. If moreover equality occurs and $Ax \neq 0$, the first (last) component of Ax (if zero, then the unique sign required to determine $S^+(Ax)$) has the same sign as the first (last) component of x.*

Characterization of TP/SSR via VD

Theorem (Gantmacher–Krein, 1950)

Let $A \in \mathbb{R}^{m\times n}$ *with* $m > n$. *Then* A *is SSR if and only if* $S^+(Ax) \leq S^-(x)$ *for all* $0 \neq x \in \mathbb{R}^n$.

Questions:

- Can we remove the restrictions imposed by Gantmacher–Krein and Motzkin?
- Can we characterize SR/SSR matrices for other sequences of sign patterns using VD?

Answer: Yes!

Characterization of SSR via VD

Theorem (C.–Yadav, *Proc. AMS*, in press)

Given $A \in \mathbb{R}^{m \times n}$, A satisfies $S^+(Ax) \leq S^-(x)$ for all $0 \neq x \in \mathbb{R}^n$ if and only if A is SSR.

Remarks:

- The above theorem ensures that if a matrix A satisfies the VD property then it is SSR.
- It does not give any information about the sign pattern of A.

Question: Can we impose additional conditions along with the VD property on A to obtain the sign pattern of A? Yes!

Single vector test of SSR via variation diminution

Theorem (C.–Yadav, *Proc. AMS*, in press)

Given $A \in \mathbb{R}^{m \times n}$ and a sign pattern $\epsilon = (\epsilon_1, \ldots, \epsilon_{\min\{m,n\}})$, the following statements are equivalent.

1. *A is SSR(ϵ).*
2. *For all $0 \neq x \in \mathbb{R}^n$, we have $S^+(Ax) \leq S^-(x)$. Further, for all $x \in \mathbb{R}^n$ with $Ax \neq 0$ and $S^+(Ax) = S^-(x) = r$, if $0 \leq r \leq \min\{m,n\} - 1$, then the sign of the first (last) component of Ax (if zero, the unique sign given in determining $S^+(Ax)$) agrees with $\epsilon_r \epsilon_{r+1}$ times the sign of the first (last) nonzero component of x.*

Natural to ask: Can the test vectors $\mathbb{R}^n$ (of Gantmacher–Krein and C.-Yadav) be reduced to a finite set?

Answer: Yes! In fact, we reduce the test set to a *single* test vector.

Definition: A *contiguous submatrix* is one whose rows and columns are indexed by consecutive integers.

Single vector test of SSR via variation diminution

3. For every contiguous square submatrix A_k of A of size $k \times k$, where $1 \leq k \leq \min\{m, n\}$, and given any fixed vector $0 \neq v := (\alpha_1, -\alpha_2, \ldots, (-1)^{k-1}\alpha_k)^T \in \mathbb{R}^k$ with all $\alpha_j \geq 0$, we define the vector
$$x^{A_k} := \operatorname{adj}(A_k)v.$$
Then $S^+(A_k x^{A_k}) \leq S^-(x^{A_k})$. Moreover, if $S^+(A_k x^{A_k}) = S^-(x^{A_k}) = r$, and if $0 \leq r \leq k-1$, then the sign of the first (last) component of $A_k x^{A_k}$ (if zero, the unique sign given in determining $S^+(A_k x^{A_k})$) agrees with $\epsilon_r \epsilon_{r+1}$ times the sign of the first (last) nonzero component of x^{A_k}.

Question: Can these results hold with vectors from any other open bi-orthant?

Answer: No!

Do test vectors from other orthant work?

Theorem (C.-Yadav, *Proc. AMS*, in press)

Suppose $x_0 \in \mathbb{R}^n$ has nonzero coordinates and at least two successive coordinates have the same sign. Let $A \in \mathbb{R}^{n\times n}$ be SSR$_{n-1}$. Then

1. *A satisfies $S^+(Ax_0) \leq S^-(x_0)$.*
2. *Further, if $Ax_0 \neq 0$ and $S^+(Ax_0) = S^-(x_0) = r$ for some $0 \leq r \leq n-2$, then the sign of the first (last) component of Ax_0 (if zero, the unique sign given in determining $S^+(Ax_0)$) equals $\epsilon_r\epsilon_{r+1}$ times the sign of the first (last) component of x_0.*

It remains to produce matrices which are SSR$_{n-1}$ but not SSR$_n$, for each n. This follows from the total positivity of the Gaussian kernel.

- Take any $B \in \mathbb{R}^{(n-1)\times(n-1)}$ such that B is SSR(ϵ) for a fixed ϵ.
- Define $B' := B \oplus \{0\} \in \mathbb{R}^{n\times n}$. Then B' is SR with $\operatorname{rank}(B') = n-1$.
- Define $A := F_\sigma^{(n)} B' F_\sigma^{(n)}$ where $F_\sigma^{(n)} = (\exp^{-\sigma(i-j)^2})_{i,j=1}^n$ for $\sigma > 0$. By the Cauchy–Binet formula, A is SSR$_{n-1}(\epsilon)$ and $\det A = 0$.

In this sense, our results are 'best possible'.

Single vector test of TP via variation diminution

Theorem (C., *Bull. London Math. Soc.* 2022)

- TP matrices are characterized by their contiguous square submatrices A_r ($r \in [k]$) satisfying the variation diminishing property, on the entire open bi-orthant $\mathbb{R}^r_{\mathrm{alt}}$;
- or on single test vectors which turn out to lie in this bi-orthant.
- Test-vectors from any open orthant apart from the open bi-orthant $\mathbb{R}^r_{\mathrm{alt}}$, can not be used to characterize total positivity via variation diminution.

To construct matrices which are TP$_{n-1}$ but not TP$_n$, for each n follows from studying a Pólya frequency function of Karlin (1964).

Example of TP$_{n-1}$ but not TP$_n$ matrix

Example

We now explain how to construct a multi-parameter family of real $n \times n$ matrices for every integer $n \geq 3$, each of which is TP$_{n-1}$ but has negative determinant. This construction involves non-integer powers of a certain Pólya frequency function, studied by Karlin in [*Trans. Amer. Math. Soc.* 1964]:

$$\Omega(x) := \begin{cases} xe^{-x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

- Karlin showed that if $\alpha \in \mathbb{Z}^{\geq 0} \cup [k-2, \infty)$ then $\Omega(x)^\alpha$ is TN$_k$, i.e., given $r \in [k]$ and real $x_1 < \cdots < x_r$ and $y_1 < \cdots < y_r$, the matrix $B := (\Omega(x_i - y_j)^\alpha)_{i,j=1}^k$ is TN.
- Recently, Khare (2020) showed that if $y_1 < \cdots < y_k < x_1 < \cdots < x_k$, then $(\Omega(x_i - y_j)^\alpha)_{i,j=1}^k$ is TP$_k$ if $\alpha > k-2$ and not TN$_k$ if $\alpha \in (0, k-2) \setminus \mathbb{Z}$.
- Thus, consider $n \geq 3$ and $\alpha \in (n-3, n-2)$, and choose real scalars $y_1 < \cdots < y_n < x_1 < \cdots < x_n$. Then the matrix $A := (\Omega(x_i - y_j)^\alpha)_{i,j=1}^n$ is TP$_{n-1}$ but not TN$_n$, because $\det A < 0$.

Refinement of Motzkin's characterization of SR matrices

Recall:

- Motzkin (Ph.D. Thesis, 1936): Characterized $m \times n$ SR matrix with rank n using the VD property for all $x \in \mathbb{R}^n$.
- Gantmacher–Krein (1950): Characterized $m \times n$ totally non-negative matrices using the VD property and sign non-reversal property.

Questions: Akin to strict sign regularity:

- Can we characterize SR matrices for other sequences of sign patterns using VD?
- Can this be strengthened to a single-vector test?

Theorem (C.-Yadav, *Proc. AMS*, in press)

Yes! (for both of these questions).

SSR construction

Question: Can we construct a strictly sign regular matrix of any size and sign pattern?

- In 1950 Gantmacher–Krein showed the existence of SSR matrices for any dimension and sign pattern.
- We developed an algorithm to explicitly construct an SSR matrix of any given size and sign pattern.

Theorem (C.–Yadav, forthcoming)

Given integers $m, n \geq 1$ and an $m \times n$ SSR matrix, it is possible to add a line to any of its borders so that the resulting matrix remains SSR. If minors of a new size occur, they can be made either positive or negative.

Theorem (C.–Yadav, forthcoming)

Let $m, n \geq 1$ be integers and fix an $m \times n$ SSR matrix A. Then, a line can be inserted between any two consecutive columns (or rows) of A such that the resulting matrix remains SSR. If minors of a new size occur, then $\epsilon'_{min\{m,n\}+1}$ can be made either positive or negative.

References

[1] P.N. Choudhury.
Characterizing total positivity: single vector tests via Linear Complementarity, sign non-reversal, and variation diminution.
Bull. London Math. Soc., 54:891–911, 2022.

[2] P.N. Choudhury and S. Yadav.
Sign regular matrices and variation diminution: single-vector tests and characterizations, following Schoenberg, Gantmacher–Krein, and Motzkin.
Proc. Amer. Math. Soc., in press.

[3] P.N. Choudhury and S. Yadav.
On line insertion and construction of strictly sign regular matrices.
Forthcoming.

[4] M. Fekete and G. Pólya.
Über ein Problem von Laguerre.
Rend. Circ. Mat. Palermo, 34:89–120, 1912.

[5] F.R. Gantmacher and M.G. Krein.
Oscillyacionye matricy i yadra i malye kolebaniya mehaničeskih sistem.
Gosudarstv. Isdat. Tehn.-Teor. Lit., Moscow-Leningrad, 1950.