Submodular spectral functions of principal submatrices of a hermitian matrix, extensions and applications

S. Friedland,[*] and S. Gaubert[†]

June 17, 2012

Abstract

We extend the multiplicative submodularity of the principal determinants of a nonnegative definite hermitian matrix to other spectral functions. We show that if f is the primitive of a function that is operator monotone on an interval containing the spectrum of a hermitian matrix A, then the function $I \mapsto \mathrm{tr}\, f(A[I])$ is supermodular, meaning that $\mathrm{tr}\, f(A[I]) + \mathrm{tr}\, f(A[J]) \leqslant \mathrm{tr}\, f(A[I \cup J]) + \mathrm{tr}\, f(A[I \cap J])$, where $A[I]$ denotes the $I \times I$ principal submatrix of A. We discuss extensions to self-adjoint operators on infinite dimensional Hilbert space and to M-matrices. We discuss an application to CUR approximation of nonnegative hermitian matrices.

2010 Mathematics Subject Classification. 15A18, 15B57, 90C10
Key words. Operator monotone functions, Loewner theorem, submodular functions, Hadamard-Fischer inequality, M-matrices, CUR approximations, self-adjoint operators.

1 Introduction and statement of the main result

Let m be a positive integer and denote $[m] := \{1, \ldots, m\}$. A real valued function $w : 2^{[m]} \to \mathbb{R}$ defined on all subsets of $[m]$ is called *nondecreasing* if $w(I) \leqslant w(J)$ when $I \subset J \subset [m]$. It is *submodular* if

$$w(I) + w(J) \geqslant w(I \cup J) + w(I \cap J)$$

for any two subsets I, J of $[m]$. The function w is called *nonincreasing* or *supermodular* whenever $-w$ is nondecreasing or submodular, respectively. A function that is both submodular and supermodular is called *modular*.

The importance of submodular functions in combinatorial optimization is well known. Several polynomial time algorithms to minimize a submodular function under a matroid constraint are known, we refer the reader to the survey [Iwa08] for more information. The maximization of a submodular function under a matroid constraint, and specially, under a cardinality constraint, $\nu_k(w) := \max_{I \subset [m], |I| \leqslant k} w(I)$, is also of great interest. For some submodular functions w the latter problem is

---

[*]Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Chicago, Illinois 60607-7045, USA, `friedlan@uic.edu`,

[†]INRIA and CMAP, École Polytechnique, 91128 Palaiseau Cédex, `stephane.gaubert@inria.fr`

NP-hard. However, a classical result [NWF78] shows that when w is nondecreasing and submodular, the greedy algorithm allows one to compute an approximation $\nu_k^G(w)$ of $\nu_k(w)$ which is such that $\nu_k^G(w) \geqslant (1-e^{-1})\nu_k(w)$. See [CCPV07] for recent developments regarding submodular maximization.

Denote respectively by $\mathbb{C}^{m \times m} \supset \mathrm{H}_m \supset \mathrm{H}_m(\mathcal{E})$ the space of $m \times m$ complex valued matrices, the space of $m \times m$ hermitian matrices, and the subset of $A \in \mathrm{H}_m$ whose eigenvalues lie in the interval $\mathcal{E} \subset \mathbb{R}$. For $I \subset [m]$ denote by $A[I]$ the principal submatrix of A, obtained from A by deleting the rows and columns in the set $[m] \setminus I$. For $A, B \in \mathrm{H}_m$ denote by $\succeq$ the Loewner ordering, so that $A \succeq B$ if $A - B \in \mathrm{H}_m([0, \infty))$. We also write $A \gneq B$ or $A \succ B$ if $A - B \in \mathrm{H}_m([0, \infty)) \setminus \{0\}$ or $A - B \in \mathrm{H}_m((0, \infty))$ respectively.

Recall that the principal minors of a nonnegative definite matrix satisfy the multiplicative submodularity property:

$$\det A[I \cup J] \det A[I \cap J] \leqslant \det A[I] \det A[J], \text{ where } I, J \subset [m], \ A \in \mathrm{H}_m([0, \infty)). \tag{1.1}$$

In other words, the function $\log(\cdot, A) : 2^{[m]} \to \mathbb{R}$ given by

$$\log(I, A) := \log \det A[I], \quad I \subset [m], \ A \in \mathrm{H}_m((0, \infty)) \tag{1.2}$$

is submodular. This inequality has arisen in the work of several authors. It goes back to Gantmacher and Kreĭn [GK60] and Kotelyanskiĭ [Kot50], see the discussion by Ky Fan [Fan67, Fan68]. It can also be found in [KK83, JB85]. The classical Hadamard-Fischer inequality for the principal minors of nonnegative definite matrices is obtained when $I \cap J = \emptyset$, understanding that $\det A[\emptyset] = 1$. We refer the reader to [FJ00] for a survey of determinantal inequalities, and to [CT88] for their relation with information theory. It is well known that the inequality (1.1) hold also for M-matrices, e.g. [Car67].

In this paper, we derive a general submodularity result for spectral functions of hermitian matrices, for some p-trace of M-matrices and some extensions to self adjoint positive operators on infinite dimensional separable Hilbert spaces.

We first summarize our results on hermitian matrices. Recall that if $A = U D U^*$, where U is unitary and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ is the diagonal matrix with the eigenvalues of A on the diagonal, the matrix $f(A)$ is defined to be $U \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_m)) U^*$. Note in particular that $\mathrm{tr} f(A) = f(\lambda_1) + \cdots + f(\lambda_m)$.

Recall that a real function f is *operator monotone* on the interval $\mathcal{E} \subset \mathbb{R}$ if for all $m \geqslant 1$ and for all $A, B \in \mathrm{H}_m(\mathcal{E})$,

$$A \preceq B \implies f(A) \preceq f(B) \ .$$

An *operator convex* function on the interval $\mathcal{E}$ is defined by requiring that

$$t f(A) + (1 - t) f(B) \succeq f(t A + (1 - t) B)$$

for each $t \in [0, 1]$. A function f is *operator antitone* (resp. *operator concave*) on an interval if $-f$ is operator monotone (resp. operator convex) on the same interval. Operator monotone and operator convex functions are characterized by Loewner theory [Löw34]. In particular, integral representations like the one of (2.1) below are known. We refer the reader to [Bha97] for more background. Matrix inequalities involving operator monotone functions can be found in [AZ99, Zha02].

The main result of this paper is the following.

Theorem 1 *Let f be a real continuous function defined on an interval $\mathcal{E}$ of $\mathbb{R}$, and assume that f is the primitive of a function that is operator monotone on the interior of $\mathcal{E}$. Then, for every $m \times m$ hermitian matrix A with spectrum in $\mathcal{E}$, the function $2^{[m]} \to \mathbb{R}$*

$$I \mapsto \operatorname{tr} f(A[I])$$

is supermodular.

It is understood that $\operatorname{tr} f(A[\emptyset]) := 0$ for every function f.

Recall that the primitive of an operator monotone function is operator convex on the same interval, but that not all operator convex functions are obtained in this way. We shall see that the conclusion of this theorem no longer holds if f is only assumed to be operator convex.

Let us mention some immediate applications. The derivative of the map t^p, namely, pt^{p-1}, is known to be operator antitone on $(0, \infty)$ for $0 < p \leqslant 1$, and operator monotone on $[0, \infty)$ for $1 \leqslant p \leqslant 2$. The derivative of the map $t \log t$, namely, $1 + \log t$, is also known to be operator monotone on $(0, \infty)$. (See [Bha97].) Hence, the next result readily follows from the main theorem.

Corollary 2 *Let A be a $m \times m$ nonnegative definite hermitian matrix. Then, for all $I, J \subset [m]$,*

$$\operatorname{tr} A[I]^p + \operatorname{tr} A[J]^p \quad \geqslant \quad \operatorname{tr} A[I \cup J]^p + \operatorname{tr} A[I \cap J]^p, \quad for \quad 0 \leqslant p \leqslant 1 \ , \quad (1.3)$$
$$\operatorname{tr} A[I]^p + \operatorname{tr} A[J]^p \quad \leqslant \quad \operatorname{tr} A[I \cup J]^p + \operatorname{tr} A[I \cap J]^p, \quad for \quad 1 \leqslant p \leqslant 2 \ , \quad (1.4)$$

and

$$\operatorname{tr} \big(A[I] \log A[I] \big) + \operatorname{tr} \big(A[J] \log A[J] \big) \leqslant \qquad\qquad (1.5)$$
$$\operatorname{tr} \big(A[I \cup J] \log A[I \cup J] \big) + \operatorname{tr} \big(A[I \cap J] \log A[I \cap J] \big) \ .$$

$\square$

We write $A[I]^p$ for $(A[I])^p$. It is understood that $A^0 := \lim_{p \searrow 0} A^p$, for every nonnegative definite matrix A. In particular, A^0 is the identity matrix Id if A is positive definite, and, if A is nonnegative definite, $\operatorname{tr} A^0 = \operatorname{rank} A$.

We now survey briefly the contents of our paper. In §2 we prove Theorem 1. In §3 we discuss M-matrices. We show the inequalities (1.1), (1.3) and (1.5). Moreover $\operatorname{tr} A[I]^p$ is a supermodular function for each $p < 0$. In §4 we discuss the extensions of Theorem 1 to the space $\mathrm{S}(\mathcal{H})$ of self adjoint operators on a separable Hilbert space $\mathcal{H}$. We consider the lattice of closed subspaces $\mathcal{U}$ in $\mathcal{H}$, with joint and meet operations $\operatorname{clo}(\mathbf{U} + \mathbf{V})$ and $\mathbf{U} \cap \mathbf{V}$, where clo denotes the closure of a set, together with its sublattice of finite dimensional subspaces $\mathcal{U}_f$. Let $P(\mathbf{U}) \in \mathrm{S}(\mathcal{H})$ be the orthogonal projection on $\mathbf{U} \in \mathcal{U}$. Associate with each $\mathbf{U} \in \mathcal{U}$ the operator $A(\mathbf{U})$ which is the restriction of $P(\mathbf{U})AP(\mathbf{U})$ to $\mathbf{U}$. Thus, $A(\mathbf{U})$ is an analog of $B[I]$ for a hermitian matrix B. For $\mathbf{U} \in \mathcal{U}_f$ it is straightforward to show that $w(\mathbf{U}) := \operatorname{tr} f(A(\mathbf{U}))$ is submodular under the conditions given in §2. For an infinite dimensional subspace $\mathbf{U}$, we restrict out discussion to $w_p(\mathbf{U}) := \operatorname{tr}(A(\mathbf{U}))^p$ and the von Neumann entropy $\tilde{w}(\mathbf{U}) := -\operatorname{tr}(A(\mathbf{U}) \log A(\mathbf{U}))$, under the assumption that A is positive and compact. Assuming that $\operatorname{tr} A^p < \infty$ we show that $w_p(\mathbf{U})$ is submodular for $p \in (0, 1)$ and

supermodular for $p \in (1, 2)$. We show that $\tilde{w}$ is submodular it $- \operatorname{tr}(A \log A) < \infty$. In §5 we discuss the CUR approximation [GTZ97] of nonnegative definite hermitian matrix. The main problem here is to find a good approximation to the maximum of $\operatorname{tr}(\log A[I])$ on all subsets I of $[m]$ of cardinality k. We discuss briefly the obvious greedy algorithm for this problem, give a simple condition where $\operatorname{tr}(\log A[I])$ is nondecreasing, and give an estimate for the CUR approximation obtained by the greedy algorithm in the general case. In §6 we give examples to show that in general the results of §2 are best possible.

Note added to the arXiv postprint. The authors thank D. Petz for having brought to their attention his work with K. Audenaert and F. Hiai (Strongly subaddtive functions, *Acta Math. Hungar.* 128(4):386–394, 2010), after this paper was published on line in *Linear Algebra Appl.* (December 2011, doi:10.1016/j.laa.2011.11.021). Theorem 1 of this paper is a slightly more general version of their Theorem 4.1 (the latter corresponds to the case in which the interval is $(0, \infty)$). Some results of the present Section 4 concerning the finite dimensional case (Theorem 5) can also be thought of as an extension of theirs. Their work is motivated by quantum information theory.

2 Proof of Theorem 1

Observe first that if all the eigenvalues of A belong to the interval $\mathcal{E}$, so do the eigenvalues of the principal submatrix $A[I]$. Hence, the function $I \mapsto \operatorname{tr} f(A[I])$ is well defined for all $I \subset [m]$.

It suffices to consider the case in which the interval $\mathcal{E}$ is bounded. Moreover, if the result is established for every matrix A the spectrum of which is included in the interior of $\mathcal{E}$, arguing by density, the result must also hold whenever the spectrum of A is included in $\mathcal{E}$. Hence, we may assume that $\mathcal{E}$ is open. Since the property to be established is invariant by a translation and a scaling of the interval $\mathcal{E}$, we finally assume that $\mathcal{E} = (-1, 1)$. Then, a theorem of Loewner (Corollary V.4.5 in [Bha97]) shows that every operator monotone function g on $(-1, 1)$ can be written as

$$g(t) = g(0) + g'(0) \int_{-1}^{1} \frac{t}{1 - \lambda t} d\mu(\lambda) \tag{2.1}$$

where μ is a probability measure on $[-1, 1]$. Moreover, $g'(0) \geqslant 0$.

Assume now that f is a primitive of g, so that

$$f(t) = a + tg(0) + g'(0) \int_{0}^{t} ds \int_{-1}^{1} \frac{s}{1 - \lambda s} d\mu(\lambda) \ ,$$

for some constant a. Observe that for $|t| < 1$, the denominator $1 - \lambda s$ is bounded below by $1 - |t|$, and so, the above double integral is absolutely convergent. Applying Fubini's theorem, we get

$$
\begin{aligned}
f(t) &= a + tg(0) + g'(0) \int_{-1}^{1} d\mu(\lambda) \int_{0}^{t} \frac{s}{1 - \lambda s} ds \\
&= a + tg(0) + g'(0) \int_{-1}^{1} \varphi(\lambda, t) d\mu(\lambda)
\end{aligned}
$$

where

$$\varphi(\lambda, t) := -\frac{t}{\lambda} - \frac{1}{\lambda^2} \log(1 - \lambda t) \ .$$

4

Although λ appears at the denominator, this expression defines a function of (λ, t) that extends continuously to $[-1, 1] \times (-1, 1)$. (In particular, $\varphi(0, t) = \frac{1}{2}t^2$.)

For all $\lambda \in [-1, 1]$, consider now the functions $\bar{w}$ and w_λ from $2^{[m]} \to \mathbb{R}$,

$$w_\lambda(I) := \operatorname{tr} \varphi(\lambda, A[I]) \ , \qquad \bar{w}(I) := \operatorname{tr}(a\operatorname{Id} + g(0)A)[I] \ ,$$

so that

$$\operatorname{tr} f(A[I]) = \bar{w}(I) + g'(0) \int_{-1}^{1} w_\lambda(I) d\mu(\lambda) \ . \tag{2.2}$$

Observe first that the function $I \mapsto -\frac{1}{\lambda} \operatorname{tr} A[I]$ is modular, for all $\lambda \neq 0$. The eigenvalues of the matrix $\operatorname{Id} - \lambda A$ belong to the interval $[1 - \lambda, 1 + \lambda]$, which implies that this matrix is positive definite. Since the sum of supermodular functions is supermodular, it follows from the multiplicative submodularity property of the determinant, Eqn (1.1), that

$$w_\lambda(I) = -\frac{1}{\lambda} \operatorname{tr} A[I] - \frac{1}{\lambda^2} \log(I, \operatorname{Id} - \lambda A)$$

is supermodular, as soon as $\lambda \neq 0$. Since $w_\lambda(I)$ depends continuously of λ, the same is true when $\lambda = 0$. Note finally that the map $\bar{w}$ is modular. Since the supermodularity property is preserved by taking positive linear combinations and integrals with respect to positive measures, the result follows from the representation (2.2). □

3 Submodularity and super-modularity inequalities for M-matrices

Denote by $\mathbb{R}_+^{m \times m}$ the set of nonnegative matrices. For $A, B \in \mathbb{R}^{m \times m}$ we denote $A \geqslant B$ if $A - B \in \mathbb{R}_+^{m \times m}$. We also write $A \gneqq B$ and $A > B$ if $A - B \in \mathbb{R}_+^{n \times n} \setminus \{0\}$ and $A - B$ has positive entries respectively. For $B \in \mathbb{R}^{m \times m}$ denote by $\rho(B)$ the spectral radius of B. The Perron-Frobenius theorem yields that $\rho(B)$ is an eigenvalue of a matrix $B \in \mathbb{R}_+^{m \times m}$. Recall that $A \in \mathbb{R}^{m \times m}$ is called an M-matrix if all off-diagonal entries of A are nonpositive and all the principal minors of A are nonnegative. Equivalently, A is an M-matrix if $A = s\operatorname{Id} - B$ for some $B \in \mathbb{R}_+^{m \times m}$ and $s \geqslant \rho(B)$. In this section we assume that A is an M-matrix and $B \in \mathbb{R}_+^{n \times n}$ unless stated otherwise. Note that $A = s\operatorname{Id} - B$ is invertible if and only if $s > \rho(B)$. Assume that A is an invertible M-matrix. Then for any $p \in \mathbb{R}$ we define

$$A^p = s^p \sum_{i=0}^{\infty} \binom{p}{i}(-s)^{-i}B^i, \qquad \log A = (\log s)\operatorname{Id} - \sum_{i=1}^{\infty} \frac{1}{is^i}B^i.$$

Since $s > \rho(B)$, a standard argument shows that the above series absolutely converges. When A is singular, we define, for every $p \geqslant 0$, $\operatorname{tr} A^p := \lim_{t \to s^+} \operatorname{tr}(t\operatorname{Id} - B)^p$. Indeed, one readily checks that the spectrum of the matrix $(t\operatorname{Id} - B)^p$ as a pointwise limit as $t \to s^+$, if $p \geqslant 0$, so that limit of $\operatorname{tr}(t\operatorname{Id} - B)^p$ as $t \to s^+$ does exist. (However, we warn the reader of the abusive character of the notation $\operatorname{tr} A^p$: the matrix $A^p := \lim_{t \to s^+}(t\operatorname{Id} - B)^p$ may not exist if 0 is not a semi-simple eigenvalue of A.) Since $\rho(B[I]) \leqslant \rho(B)$ for any $I \subset [n]$, it follows that $A[I]$ is also an M-matrix.

5

Furthermore, if A is invertible then $A[I]$ is invertible. We finally agree that if A is singular then $\operatorname{tr} \log A = -\infty$ and $\operatorname{tr} A^p = \infty$ if $p < 0$.

The main result of this section is the following.

Theorem 3 *Let A be an M-matrix. Then, for all $I, J \subset [m]$,*

$$\operatorname{tr} A[I]^p + \operatorname{tr} A[J]^p \;\geqslant\; \operatorname{tr} A[I \cup J]^p + \operatorname{tr} A[I \cap J]^p, \quad for \quad 0 \leqslant p \leqslant 1 \;,$$
$$\operatorname{tr} A[I]^p + \operatorname{tr} A[J]^p \;\leqslant\; \operatorname{tr} A[I \cup J]^p + \operatorname{tr} A[I \cap J]^p, \quad for \quad p < 0 \quad or \quad 1 \leqslant p \leqslant 2 \;,$$

and

$$\operatorname{tr}\big(A[I] \log A[I]\big) + \operatorname{tr}\big(A[J] \log A[J]\big) \leqslant$$
$$\operatorname{tr}\big(A[I \cup J] \log A[I \cup J]\big) + \operatorname{tr}\big(A[I \cap J] \log A[I \cap J]\big) \;.$$

The proof of this theorem relies on the following lemma.

Lemma 4 *For any $B \in \mathbb{R}_+^{m \times m}$*

$$\operatorname{tr} B[I]^n + \operatorname{tr} B[J]^n \leqslant \operatorname{tr} B[I \cup J]^n + \operatorname{tr} B[I \cap J]^n$$

for every integer $n \geqslant 1$, and the equality holds for $n = 1$.

Proof. Consider the complete directed graph with m nodes. Define a *closed walk* of *length n* to be a sequence $\alpha = (i_1, \ldots, i_{n+1})$ of elements of $[m]$ such that $i_{n+1} = i_1$. We say that α is *included* in I, and we write $\alpha \subset I$, if $i_1, \ldots, i_{n+1} \in I$. The *weight* of this walk is $|\alpha| := B_{i_1 i_2} \ldots B_{i_n i_{n+1}}$. By a classical result [Sta97, Th. 4.7.1], $\operatorname{tr} B[I]^n$ is the sum of the weights of all closed walks of length n included in I. By a disjunction of cases, we deduce that

$$\operatorname{tr} B[I \cup J]^n \;=\; \sum_{\substack{\alpha \subseteq I \\ \alpha \not\subseteq J}} |\alpha| + \sum_{\substack{\alpha \subseteq I \\ \alpha \subseteq J}} |\alpha| + \sum_{\substack{\alpha \subseteq J \\ \alpha \not\subseteq I}} |\alpha| + \sum_{\substack{\alpha \subseteq I \cup J \\ \alpha \not\subseteq I \\ \alpha \not\subseteq J}} |\alpha| \;, \tag{3.1}$$

with the convention that all sums are restricted to the walks α of length n. Adding the two first sums in (3.1) yields $\operatorname{tr} B[I]^n$. Moreover, adding $\operatorname{tr} B[I \cap J]^n$ to the third sum in (3.1) yields $\operatorname{tr} B[J]^n$. We conclude that

$$\operatorname{tr} B[I \cup J]^n + \operatorname{tr} B[I \cap J]^n - \operatorname{tr} B[I]^n - \operatorname{tr} B[J]^n = \sum_{\substack{\alpha \subseteq I \cup J \\ \alpha \not\subseteq I \\ \alpha \not\subseteq J}} |\alpha| \geqslant 0 \;,$$

since the entries of the matrix B are nonnegative. Moreover, when $n = 1$, the latter sum trivially vanishes. □

Proof of Theorem 4. Suppose first that $s > \rho(B)$. We first consider the function $\operatorname{tr} A[I]^p$. We have

$$\operatorname{tr} A[I]^p = s^p \operatorname{tr} \operatorname{Id}[I] - s^{p+1} p \operatorname{tr} B[I] + w_p'(I)$$

where

$$w_p'(I) := \sum_{i \geqslant 2} (-1)^i s^{p+i} \frac{p(p-1) \cdots (p-i+1)}{i!} \operatorname{tr} B[I]^i \;.$$

Observe that for $p \in (-\infty, 2]$, all the coefficients of the latter sum have the same sign (or vanish), and this sign is positive if $p \in (-\infty, 0) \cup (1, 2)$ and negative if $p \in (0, 1)$. Hence, it follows from Lemma 4 that the function w_p' is supermodular in the former case and submodular in the latter case. Since $I \mapsto \operatorname{tr} A[I]^p$ is the sum of the modular function $I \mapsto s^p \operatorname{tr} \operatorname{Id}[I] - s^{p+1} p \operatorname{tr} B[I]$ and of w_p', the announced submodularity and supermodularities of the map $I \mapsto \operatorname{tr} A[I]^p$ hold for this matrix A.

Similarly, since $-(1-x)\log(1-x) - x$ has negative Taylor coefficients, it follows as above that $-\operatorname{tr}(A[I] \log A[I])$ is submodular.

Assume now that $s = \rho(B)$, i.e., that A is a singular M-matrix. Let $A_t = t\operatorname{Id} - B$ for $t > \rho(B)$. Letting $t \searrow \rho(B)$ we deduce all the above inequalities for A except for $p = 0$. Letting $p \searrow 0$ we deduce that $\operatorname{tr} A[I]^0$ is a submodular function. (Note that we cannot use $p < 0$ since $\operatorname{tr} A^p$ in this case in ∞.) □

Note that since $\log(1 - x)$ has negative Taylor coefficients, it follows as above that the function $I \mapsto \operatorname{tr}(\log A[I])$ is submodular, i.e., we have obtained yet another proof of the inequality (1.1) for M-matrices [Car67].

4   Submodular functions on the lattice of closed subspaces of a Hilbert space

Let $\mathcal{H}$ be a separable Hilbert space over $\mathbb{C}$ with the inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ and the norm $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Denote by $\mathrm{S}(\mathcal{H}) \supset \mathrm{S}_+(\mathcal{H})$ the space of bounded self adjoint linear operators $A : \mathcal{H} \to \mathcal{H}$ and the cone of bounded positive self adjoint operators. So $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle$ for $A \in \mathrm{S}(\mathcal{H})$ and $\langle A\mathbf{x}, \mathbf{x} \rangle \geqslant 0$ for $A \in \mathrm{S}_+(\mathcal{H})$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{H}$. For $A \in \mathrm{S}(\mathcal{H})$ the operator norm $\|A\|$ is given by $\sup_{\|\mathbf{x}\| \leqslant 1} |\langle A\mathbf{x}, \mathbf{x} \rangle|$. More precisely, the spectrum of $A \in \mathrm{S}(\mathcal{H})$ lies in the interval $[a(A), b(A)]$ where

$$a(A) = \inf_{\|\mathbf{x}\| \leqslant 1} \langle A\mathbf{x}, \mathbf{x} \rangle, \ b(A) = \sup_{\|\mathbf{x}\| \leqslant 1} \langle A\mathbf{x}, \mathbf{x} \rangle, \ \|A\| = \max(|a(A)|, |b(A)|). \qquad (4.1)$$

Assume that $A \in \mathrm{S}(\mathcal{H})$. We first discuss the analog of $A[I]$ when I is a finite subset of $\mathbb{N}$ of cardinality l. Let $\mathbf{U}$ be an l-dimensional subspace of $\mathcal{H}$. Denote by $P(\mathbf{U}) \in \mathrm{S}_+(\mathcal{H})$ the orthogonal projection on $\mathbf{U}$. So $\mathcal{H} = \mathbf{U} \oplus \mathbf{U}^\perp = \mathbf{U} \oplus (\operatorname{Id} - P(\mathbf{U}))\mathcal{H}$. Consider the operator $P(\mathbf{U})AP(\mathbf{U})$. Clearly $P(\mathbf{U})AP(\mathbf{U})\mathbf{U}^\perp = \{\mathbf{0}\}, P(\mathbf{U})AP(\mathbf{U})\mathbf{U} \subseteq \mathbf{U}$. Denote by $A(\mathbf{U})$ the restriction of $P(\mathbf{U})AP(\mathbf{U})$ to $\mathbf{U}$. As $\|P\mathbf{x}\| \leqslant \|\mathbf{x}\|$ for $\mathbf{x} \in \mathcal{H}$ the characterization (4.1) yields

$$a(A) \leqslant a(A(\mathbf{U})) \leqslant b(A(\mathbf{U})) \leqslant b(A). \qquad (4.2)$$

In particular, if the spectrum of A lies in a given interval $\mathcal{E}$, so does the spectrum of $A(\mathbf{U})$.

Let $\mathbf{V}$ be another finite dimensional vector space of $\mathcal{H}$. Then $A(\mathbf{U}), A(\mathbf{V}), A(\mathbf{U}+\mathbf{V}), A(\mathbf{U} \cap \mathbf{V})$ are the analogs of $A[I], A[J], A[I \cup J], A[I \cap J]$ respectively.

Indeed, let $\mathbf{W} = \mathbf{U} + \mathbf{V}$ and choose an orthonormal basis $W := \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ in $\mathbf{W}$, such that $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$ is a basis in $\mathbf{U}$, $Z := \{\mathbf{x}_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_q\}$ is a basis in $\mathbf{U} \cap \mathbf{V}$ and $Y := \{\mathbf{x}_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_m\}$ is a basis in $\mathbf{V}$. It is easy to see that the

operators $A(\mathbf{W}), A(\mathbf{U}), A(\mathbf{V}), A(\mathbf{U} \cap \mathbf{V})$ are represented in the above bases by the following matrices

$$A(W) = [\langle A\mathbf{x}_i, \mathbf{x}_j \rangle]_{i=j=1}^{m}, \qquad A(X) = [\langle A\mathbf{x}_i, \mathbf{x}_j \rangle]_{i=j=1}^{q},$$
$$A(Y) = [\langle A\mathbf{x}_i, \mathbf{x}_j \rangle]_{i=j=p}^{m}, \qquad A(Z) = [\langle A\mathbf{x}_i, \mathbf{x}_j \rangle]_{i=j=p}^{q}.$$

Denote by $\mathcal{U} \supset \mathcal{U}_f$ the set of all closed subspaces of $\mathcal{H}$ and the subset of all finite dimensional subspaces, respectively. A function $w : \mathcal{U}_f \to \mathbb{R}$ is said to be *submodular* if $w(\mathbf{U}) + w(\mathbf{V}) \geqslant w(\mathbf{U}+\mathbf{V}) + w(\mathbf{U} \cap \mathbf{V})$ for all $\mathbf{U}, \mathbf{V} \in \mathcal{U}_f$. It is *supermodular* if $-w$ is submodular, and *modular* if it is both submodular and supermodular. The following result is an immediate consequence of Theorem 1 and of the previous observations.

Theorem 5 *Let f be a real continuous function defined on an interval $\mathcal{E}$ of $\mathbb{R}$, and assume that f is the primitive of a function that is operator monotone on the interior of $\mathcal{E}$. Let $\mathcal{H}$ be an infinite dimensional separable Hilbert space space. Then, for every $A \in \mathrm{S}(\mathcal{H})$ with spectrum in $\mathcal{E}$, the function $w : \mathcal{U}_f \to \mathbb{R}$ given by*

$$\mathbf{U} \mapsto \operatorname{tr} f(A(\mathbf{U}))$$

is supermodular. □

A function $w : \mathcal{U}_f \to \mathbb{R}$ is *extendable* to $\mathcal{U}$ if the following condition holds for each closed infinite dimensional subspace $\mathbf{U} \subset \mathcal{H}$. Let $\mathbf{U}_i, i \geqslant 1$ be an increasing sequence of finite dimensional subspaces of $\mathbf{U}$ such that $\cup_{i=1}^{\infty} \mathbf{U}_i$ is dense in $\mathbf{U}$. Then the sequence $w(\mathbf{U}_i), i \geqslant 1$ converges to a unique value independent of the sequence $\mathbf{U}_i, i \geqslant 1$. We denote this limit by $w(\mathbf{U})$.

In the rest of this section, we discuss the extension of certain submodular and supermodular functions of a subspace $\mathcal{U}$, arising from Theorem 5.

Denote by $\mathrm{CS}_+(\mathcal{H}) \subset \mathrm{S}_+(\mathcal{H})$ the closed ideal of positive compact operators. For $A \in \mathrm{CS}_+(\mathcal{H})$, the Hilbert space $\mathcal{H}$ has a countable orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \dots$ such that

$$A\mathbf{u}_i = \lambda_i(A)\mathbf{u}_i, i \geqslant 1, \quad \lambda_1(A) \geqslant \lambda_2(A) \geqslant \dots \geqslant 0, \lim_{i \to \infty} \lambda_i(A) = 0.$$

We now state some known results about characterizations of eigenvalues of $A \in \mathrm{S}_+(\mathcal{H})$ that we will use here. Consult with [Fri73]. Let $\mathcal{X}_m$ be the set of all orthonormal system of m vectors $X := \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ in $\mathcal{H}$. The *convoy principle* (see Lemma 1 in [Fri73]) can be stated in the form of Corollary 4.4.3 in [Fbk].

$$\sup_{X \in \mathcal{X}_m} \lambda_i(A(X)) = \lambda_i(A) \text{ for } i = 1, \dots, m. \qquad (4.3)$$

The supremum is achieved for $X = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$. This in particular implies the Ky-Fan inequalities

$$\operatorname{tr}(A(X)) = \sum_{i=1}^{m} \langle A\mathbf{x}_i, \mathbf{x}_i \rangle \leqslant \sum_{i=1}^{m} \lambda_i(A).$$

Let $\mathbf{U} = \operatorname{span}(\mathbf{x}_1, \dots, \mathbf{x}_m)$. Then the eigenvalues of $A(X)$ are the eigenvalues of $A(\mathbf{U})$. The following approximation result is well known and can be straightforwardly deduced from the arguments in [Fri73].

8

Lemma 6 *Let $\mathcal{H}$ be an infinite dimensional separable Hilbert space. Let $A \in \mathrm{CS}_+(\mathcal{H})$ and $\mathbf{U}$ be an infinite dimensional closed subspace of $\mathcal{H}$. Assume that $\mathbf{U}_i, i \geqslant 1$ is an increasing sequence of finite dimensional subspaces in $\mathbf{U}$, such that $\cup_{i \geqslant 1}\mathbf{U}_i$ is dense in $\mathbf{U}$. Then for each $j \geqslant 1$, the sequence $\lambda_j(A(\mathbf{U}_i)), i \geqslant 1, \dim \mathbf{U}_i \geqslant j$ is nondecreasing and converges to $\lambda_j(A(\mathbf{U}))$.*

Let $p > 0$. Denote by $w_p : \mathcal{U}_f \to \mathbb{R}$ the function $w_p(\mathbf{U}) := \mathrm{tr}\, A(\mathbf{U})^p$ for $A \in \mathrm{S}_+(\mathcal{H})$. The interlacing properties of hermitian matrices yield that w_p is a nondecreasing function, i.e. $w_p(\mathbf{U}) \leqslant w_p(\mathbf{V})$ for $\mathbf{U} \subset \mathbf{V}$. For $p > 0$ denote by $\mathrm{CS}_{+,p}(\mathcal{H})$ the subset of all positive compact operators A on $\mathcal{H}$ such that $\mathrm{tr}\, A^p := \sum_{j=1}^{\infty} \lambda_j(A)^p < \infty$. Clearly $\mathrm{CS}_{+,p}(\mathcal{H}) \subsetneqq \mathrm{CS}_{+,q}(\mathcal{H})$ for $0 < p < q$.

Lemma 7 *Let $\mathcal{H}$ be an infinite dimensional separable Hilbert space. Assume that $p > 0$ and $A \in \mathrm{CS}_{+,p}(\mathcal{H})$. Then the function $\mathbf{U} \mapsto \mathrm{tr}\, A(\mathbf{U})^p$ is extendable from $\mathcal{U}_f$ to $\mathcal{U}$. Moreover w_p is a nondecreasing function.*

Proof. Let $\mathbf{U}_i, i \geqslant 1$ be a sequence of finite dimensional subspaces of $\mathbf{U}$ such that $\cup_{i=1}^{\infty}\mathbf{U}_i$ is dense in $\mathbf{U}$. Assume that $\dim \mathbf{U}_i = n_i$. Then the positivity of $A(\mathbf{U}_i)$ and (4.3) imply that $0 \leqslant \lambda_j(A(\mathbf{U}_i)) \leqslant \lambda_j(A(\mathbf{U})) \leqslant \lambda_j(A)$, for $j = 1, \dots, n_i$. Consider now, for each $i \geqslant 1$, the function $\varphi_i : \{1, 2, \dots\} \to [0, \infty)$ such that $\varphi_i(j) = \lambda_j(A(\mathbf{U}_i))^p$ for $1 \leqslant j \leqslant n_i$ and $\varphi_i(j) = 0$ for $j > n_i$. Consider also the function $\varphi : \{1, 2, \dots\} \to [0, \infty)$ such that $\varphi(j) := \lambda_j(A(\mathbf{U}))^p$. Then, by Lemma 6, the sequence of nonnegative functions $\varphi_i, i \geqslant 1$ converges monotonically to the function φ as $i \to \infty$, and then, the monotone convergence theorem implies that $\mathrm{tr}\, A(\mathbf{U}_i)^p = \sum_{j \geqslant 1} \varphi_i(j)$ converges, as $i \to \infty$, to $\sum_{j \geqslant 1} \varphi(j) = \mathrm{tr}\, A(\mathbf{U})^p \leqslant \mathrm{tr}\, A^p < \infty$. Finally, the convoy principle implies that $w : \mathcal{U} \to \mathbb{R}$ is a nondecreasing function. $\qquad\square$

Lemma 8 *Let $\mathcal{H}$ be an infinite dimensional separable Hilbert space. Assume that $A \in \mathrm{CS}_+(\mathcal{H})$, and that $-\mathrm{tr}(A \log A) < \infty$. Then the function $\tilde{w} : \mathbf{U} \mapsto -\mathrm{tr}(A(\mathbf{U}) \log A(\mathbf{U}))$ is extendable from $\mathcal{U}_f$ to $\mathcal{U}$.*

Proof. Let $\mathbf{U}_i, i \geqslant 1$ be as in the proof of Lemma 7, still with $n_i = \dim \mathbf{U}_i$. Observe first that the function $-x \log x$ is increasing and nonnegative on $[0, e^{-1}]$. As $\lim_{j \to \infty} \lambda_j(A) = 0$, we can find an index $\bar{j}$ such that $\lambda_j(A) \leqslant e^{-1}$ for all $j \geqslant \bar{j}$. Let us fix such a j. Then, by the convoy principle, we have $0 \leqslant \lambda_j(A(\mathbf{U}_i)) \leqslant \lambda_j(A) \leqslant e^{-1}$, for all i, and since the sequence $\lambda_j(A(\mathbf{U}_i)), i \geqslant 1$ is nondecreasing, it follows that the sequence $-\lambda_j(A(\mathbf{U}_i)) \log \lambda_j(A(\mathbf{U}_i)), i \geqslant 1$ is nondecreasing and nonnegative. Applying the monotone convergence theorem as in the proof of Lemma 7, we deduce that

$$
\begin{aligned}
0 \leqslant \lim_{i \to \infty} \sum_{j \geqslant \bar{j}} -\lambda_j(A(\mathbf{U}_i)) \log \lambda_j(A(\mathbf{U}_i)) \;&=\; \sum_{j \geqslant \bar{j}} -\lambda_j(A(\mathbf{U})) \log \lambda_j(A(\mathbf{U})) \\
&\leqslant\; \sum_{j \geqslant \bar{j}} -\lambda_j(A) \log \lambda_j(A) \;, \qquad (4.4)
\end{aligned}
$$

again with the convention that $\lambda_j(\mathbf{U}_i) = 0$ for $j > n_i$. Since the sum in (4.4) differs from the sum $\sum_{j \geqslant 1} -\lambda_j(A) \log \lambda_j(A) = -\mathrm{tr}(A \log(A)) < \infty$ only by a finite

number of terms, we conclude that the sum in (4.4) cannot be equal to ∞. Since for all $0 \leqslant j < \bar{\jmath}$, we also have $\lambda_j(A(\mathbf{U}_i)) \to \lambda_j(A(\mathbf{U}))$ as $i \to \infty$, we finally get

$$-\infty < \lim_{i \to \infty} - \operatorname{tr}(A(\mathbf{U}_i) \log A(\mathbf{U}_i)) = - \operatorname{tr}(A(\mathbf{U}) \log A(\mathbf{U})) < \infty \ .$$

$\square$

Theorem 9 *Let $\mathcal{H}$ be an infinite dimensional separable Hilbert space. Assume that $p > 0$ and $A \in \mathrm{CS}_{+,p}(\mathcal{H})$. Then the function $w_p : \mathcal{U} \to \mathbb{R}$ given by $w_p(\mathbf{U}) := \operatorname{tr}(A(\mathbf{U}))^p$ is nondecreasing, it is submodular for $p \in (0, 1)$, supermodular for $p \in (1, 2]$ and modular for $p = 1$. Furthermore, if $A \in \mathrm{CS}_{+}(\mathcal{H})$ is such that $- \operatorname{tr}(A \log A) < \infty$, and in particular, if $A \in \mathrm{CS}_{+,p}(\mathcal{H})$ for some $p \in (0, 1)$, then the function $\tilde{w}(\mathbf{U}) := - \operatorname{tr}(A(\mathbf{U}) \log A(\mathbf{U})), \mathbf{U} \in \mathcal{U}$ is submodular.*

Proof. The results for the function $w_p(\cdot)$ follows from Corollary 2 and Lemma 7. The submodularity property for the function $\tilde{w}(\cdot)$ follows from the same corollary and Lemma 8. Finally, for $p \in (0, 1)$, $(-x \log x)/x^p \to 0$ as $x \searrow 0$, and since $\lambda_j(A) \searrow 0$ as $j \nearrow \infty$, the convergence of the series $\operatorname{tr} A^p = \sum_j \lambda_j(A)^p$ implies the convergence of the series $- \operatorname{tr}(A \log A) = \sum_j -\lambda_j(A) \log \lambda_j(A)$. Hence, $- \operatorname{tr}(A \log(A)) < \infty$ holds as soon as $A \in \mathrm{CS}_{+,p}(\mathcal{H})$ for some $p \in (0, 1)$. $\square$

5 CUR approximation of nonnegative definite matrices

Let $A \in \mathbb{C}^{m \times n}$ be of rank $r = \operatorname{rank} A$. Assume that $k \in [r - 1]$. The best rank k approximation of A, denoted as A_k, is given by the *singular value decomposition*, abbreviated here as SVD, [GVL96]. When m and n are very big, e.g. $n, m \geqslant 10^6$, finding A_k is not computationally possible, since the complexity of computing A_k is $O(kmn)$. A good alternative for A_k is the *CUR* approximation, where $C \in \mathbb{R}^{m \times k}, R \in \mathbb{C}^{k \times n}$ are submatrices of A of the form $A[[m], J], A[I, [n]]$, where $J \subset [n], I \subset [m]$ are subsets of columns and rows of A of cardinality k. Here $U \in \mathbb{C}^{k \times k}$ is a suitably chosen matrix. The best choice $U^\star$ for the Frobenius norm, i.e. $\arg \min_U \operatorname{tr}((A - CUR)(A - CUR)^*)$, is given by $C^\dagger A R^\dagger$, where $B^\dagger$ denotes the Moore-Penrose inverse [FMMN]. Again, for $m, n \gg 1$ it is unfeasible to compute $U^\star$.

Assume that $A[I, J] \in \mathbb{C}^{k \times k}$, the submatrix of A based on the rows and columns I, J respectively, is invertible. Then $CA[I, J]^{-1}R$ has the same I rows and J columns as A. Hence the best choice of $I^\star, J^\star$ seems to be given by [GTZ97]

$$\mu_k(A) := \max\{|\det A[I, J]|, I \subset [m], J \subset [n], |I| = |J| = k\} = |\det A[I^\star, J^\star]|. \quad (5.1)$$

More precisely, for $A = [a_{ij}]_{i=j=1}^{m,n} \in \mathbb{C}^{m \times n}$ let $\|A\|_{\infty,e} := \max_{i \in [m], j \in [n]} |a_{ij}|$ be the entrywise max-norm of A. Denote by $\sigma_1(A) \geqslant \ldots \geqslant \sigma_r(A) > 0$ the r-positive singular values of A. If $\det A[I, J] \neq 0$ one has the inequality [GTZ97]

$$\|A - CA[I, J]^{-1}R\|_{\infty,e} \leqslant \frac{\mu_k(A)}{|\det A[I, J]|}(k + 1)\sigma_{k+1}(A). \quad (5.2)$$

(See [Fbk, §4.13] for a simple proof of the above inequality.)

In this section we discuss the maximal problem (5.1) and the related problem of *CUR* approximation where $A \in \mathrm{H}_{m,+}$.

Lemma 10 *For $A \in \mathrm{H}_{m,+}$ and $k \in [m]$, the maximum in (5.1) is achieved for $I^\star = J^\star$, i.e.*

$$\mu_k(A) = \max_{I \subset [m], |I| = k} \det A[I]. \tag{5.3}$$

Proof. Let $A = [a_{ij}]_{i=j=1}^m$. We first consider the easy case $k = 1$. Recall that $a_{ii} \geqslant 0$ and each 2 principal minor of A is nonnegative. Hence $a_{ii}a_{jj} - |a_{ij}|^2 \geqslant 0$. Thus $\max(a_{ii}, a_{jj}) \geqslant |a_{ij}| = |a_{ji}|$. Hence (5.3) holds for $k = 1$. Assume that $k > 1$. Let $C_k(A) = \wedge^k A \in \mathbb{C}^{\binom{m}{k} \times \binom{m}{k}}$ be the k-th compound of A. (The entries of $C_k(A)$ are all $\det A[I, J]$, $|I| = |J| = k$ arranged in a lexicographical order.) Recall that $C_k(A) \in \mathrm{H}_{\binom{m}{k},+}$. Hence the previous argument applied now to any 2×2 principal minor of the matrix $C_k(A)$ yields (5.3). □

We now introduce a simple greedy algorithm for finding an approximation of $\mu_k(A)$ for $A \in \mathrm{H}_{m,+}$ and the corresponding row index I.

Algorithm 11
Input*: $A = [a_{ij}] \in \mathrm{H}_{m,+}, k \in [m], I = \emptyset, K = [m], t = 1$*
Output*: A subset I of $[m]$ of cardinality k and $\det A[I](= t)$*
  Find $i \in K$ $a_{ii} = \max_{j \in K} a_{jj}$.
    If $a_{ii} = 0$ *set* $I = [k], t = 0$ exit
    $I = I \cup \{i\}, K = K \setminus \{i\}, t = ta_{ii}$
    If $|I| = k$ exit
    For $p, q \in K$ $a_{pq} = a_{pq} - \frac{a_{pi}a_{iq}}{a_{ii}}$

To analyse this algorithm, assume for the simplicity of argument that $\det A > 0$. Then, we may introduce the submodular function $w(I) := \mathrm{tr}(\log A[I])$, and consider, as in the introduction, the maximum $\nu_k(w)$ of the function w over all subsets $J \subset [m]$ of cardinality k, so that $\nu_k(w) = \log \mu_k(A)$.

Algorithm 11 should be compared with the standard greedy algorithm [NWF78] to approximate the maximum $\nu_k(w)$, which consists in constructing a sequence of sets $\emptyset = I^0 \subset I^1 \subset \cdots \subset I^k \subset [m]$ such that at each step r, $I^r = I^{r-1} \cup \{i_r\}$ where the index i_r is chosen to so that $w(I^{r-1} \cup \{i_r\}) = \max_{i \in [m] \setminus I^{r-1}} w(I^{r-1} \cup \{i\})$. We denote by $\nu^G(w) := w(I^k)$ the value of the solution I^k returned by this greedy algorithm. Algorithm 11 is nothing but an implementation of this greedy algorithm, in which at each step, the augmenting index i_r is obtained as a maximal pivot (third line of the algorithm) and the value $w(I^r)$ is obtained incrementally from $w(I^{r-1})$ by one Gaussian elimination step (fifth and seventh lines of the algorithm). In particular, denoting by $\mu_k^G(A)$ the value returned by Algorithm 11, we get $\nu_k^G(w) = \log \mu_k^G(A)$.

Assume now that w is nondecreasing. Then, it is straightforward to show that if $m \geqslant 2$, then $\det A[J] \geqslant 1$ for each $J \subset [m]$, so that w is nonnegative. Then, the result of [NWF78] can be applied to w, showing that the greedy algorithm has an approximation factor $1 - e^{-1}$, i.e., $\nu_k^G(w) \geqslant (1 - e^{-1})\nu_k(w)$. It follows that the set I returned by Algorithm 11 satisfies

$$\det A[I] = \mu_k^G(A) \geqslant \mu_k(A)^{1-e^{-1}}. \tag{5.4}$$

Therefore, we arrive at the following estimate of the CUR approximation error of the greedy algorithm.

Proposition 12 *Let $m \geqslant 2$, and let $A \in \mathrm{H}_{m,+}$ be such that $w(I) := \operatorname{tr} \log A[I]$ is nondecreasing. Then, the subset I returned by Algorithm 11 satisfies:*

$$\|A - A[[m], I]A[I]^{-1}A[[m], I]^*\|_{\infty, e} \leqslant (k+1)\mu_k(A)^{e^{-1}}\sigma_{k+1}(A).$$

Proof. Combine Inequality (5.4) with Inequality (5.2). □

We now give a simple condition for w to be a nondecreasing function.

Lemma 13 *Let $A \in \mathrm{H}_{m,+}$. Assume that $\lambda_m(A) \geqslant 1$. Then w is a nondecreasing function.*

Proof. Let I be a nontrivial subset of m and let $i \in [m] \setminus I$. As the eigenvalues of $B := A[I]$ interlace the eigenvalues of $C := A[I \cup \{i\}]$ we deduce that

$$\det C = \prod_{j=1}^{|I|+1} \lambda_j(C) \geqslant (\det B)\lambda_{|I|+1}(C) \geqslant (\det B)\lambda_m(A) \geqslant \det B.$$

□

We conclude this section with an upper estimate of the CUR approximation error of $A \in \mathrm{H}_{m,+}$ using Algorithm 11. This estimate relies on the Hadamard determinant inequality and remains valid even when w is not nondecreasing.

Proposition 14 *Let $A \in \mathrm{H}_{m,+}$ and $k \in [m-1]$. Assume that $k \leqslant \operatorname{rank} A$. Let I be any subset of $[m]$ of cardinality k returned by Algorithm 11. Let $a_1, \ldots, a_k$ be the k maximal entries out of the m diagonal entries of A. Then*

$$\|A - A[[m], I]A[I]^{-1}A[[m], I]^*\|_{\infty, e} \leqslant \frac{\prod_{i=1}^k a_i}{\det A[I]}(k+1)\sigma_{k+1}(A).$$

Proof. Let $I^\star$ be such that $\mu_k(A) = \det A[I^\star]$. The Hadamard determinant inequality yields that $\det A[I^\star] \leqslant \prod_{j \in I^\star} a_{jj}$. Clearly $\prod_{j \in I^\star} a_{jj} \leqslant a_1 \ldots a_k$. Use (5.2) to deduce the lemma. □

6   Examples and further comments

We now give counter-examples showing that several natural generalizations of the present results do not hold. In particular, we shall see that the assumptions in Corollary 2 under which the function $I \mapsto T_p(I, A) := \operatorname{tr} A[I]^p$ is sub or supermodular, for a $m \times m$ nonnegative definite matrix A, are tight.

To this end, we shall consider a hermitian $m \times m$ matrix B with a block partition of the form

$$B = \begin{array}{cc} & \begin{array}{ccc} L_1 & L_2 & L_3 \end{array} \\ \begin{array}{c} L_1 \\ L_2 \\ L_3 \end{array} & \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{12}^* & B_{22} & B_{23} \\ B_{13}^* & B_{23}^* & B_{33} \end{pmatrix} \end{array}$$

so that $[m]$ is the disjoint union of the sets L_1, L_2, L_3. Given $I, J \subset [m]$, it is convenient to set

$$\delta(f, I, J, B) := \operatorname{tr} f(B[I]) + \operatorname{tr} f(B[J]) - \operatorname{tr} f(B[I \cup J]) - \operatorname{tr} f(B[I \cap J]) \ .$$

Hence, $\delta(f, I, J, B) \geqslant 0$ for all I, J whenever $I \mapsto \operatorname{tr} f(B[I])$ is submodular.

We now choose

$$I := L_1 \cup L_2, \qquad J := L_2 \cup L_3 \ .$$

Using the interpretation of the matrix product in terms of concatenation of paths, we readily get

$$\delta(t^2, I, J, B) = -2 \operatorname{tr}(B_{13} B_{13}^*) \leqslant 0 \ . \tag{6.1}$$

Moreover, the inequality is strict as soon as $B_{13} \neq 0$. Since $\delta(t^p, I, J, B)$ depends continuously of p, we get the following proposition, in which one can take $A = B$.

Proposition 15 *For all $m \geqslant 3$, and for p close enough to 2, the strict super-modularity inequality for $T_p(\cdot, A)$ may hold.* $\qquad\square$

In order to give examples in which the strict opposite inequality may hold, we observe that

$$B_{13} = 0 \implies \delta(t^3, I, J, B) = 0 \quad \text{and} \quad \delta(t^4, I, J, B) = -4 \operatorname{tr} B_{12}^* B_{12} B_{23} B_{23}^* \ . \tag{6.2}$$

Note that $\delta(t^4, I, I, B) < 0$, unless the nonnegative definite matrix $B_{12} B_{23} B_{23}^* B_{12}^*$ is zero. We assume in the sequel that this is not the case.

Let us consider

$$A = \operatorname{Id} + sB \ .$$

Using (6.1) and (6.2), together with the trivial fact that $\delta(t, I, J, B) = 0$, and $\operatorname{tr}(I + sB)^p = \operatorname{tr} I + ps \operatorname{tr} B + \cdots + \frac{p(p-1)(p-2)(p-3)}{4!} s^4 \operatorname{tr} B^4 + O(s^5)$, we identify the first terms of the Taylor expansion of $\delta(t^p, I, J, A)$ in s,

$$\delta(t^p, I, J, A) = -\frac{p(p-1)(p-2)(p-3)}{3!} s^4 \operatorname{tr} B_{12} B_{12}^* B_{23} B_{23}^* + O(s^5) \tag{6.3}$$

For s sufficiently small, the sign of the latter expression is positive for $p \in (2, 3)$, and negative for $p > 3$ or $p < 0$.

Proposition 16 *For all $m \geqslant 3$, and for all $p \in (2, 3)$ (resp $p < 0$ or $p > 3$) the strict submodularity (resp. supermodularity) inequality for $T_p(\cdot, A)$ may hold.* $\qquad\square$

Remark 17 The absence of systematic inequalities, except for $p \in [0, 1]$ or $p \in [1, 2]$ is related to the absence of majorization properties. Indeed, consider again the matrix B above, assume that it is of size 3, so that the blocks B_{12} and B_{23} are scalars, and assume that $B_{13} = 0$, so that B is a Jacobi matrix. We choose again $I = \{1, 2\}, J = \{2, 3\}$. The spectra of the matrices $B, B[I], B[J], B[I \cap J]$ are

$$\left(\sqrt{B_{12}^2 + B_{23}^2}, 0, -\sqrt{B_{12}^2 + B_{23}^2}\right), \quad \left(|B_{12}|, -|B_{12}|\right), \quad \left(|B_{23}|, -|B_{23}|\right), \quad (0) \ .$$

Consider $A := \mathrm{Id} + B$, so that $A > 0$ for $1 > \sqrt{B_{12}^2 + B_{23}^2}$. The joint sequences of the eigenvalues of A with $A[I \cap J]$ and of $A[I]$ with $A[J]$ are

$$\left(\sqrt{B_{12}^2 + B_{23}^2}+1, 1, 1, -\sqrt{B_{12}^2 + B_{23}^2}+1\right), \quad \left(|B_{12}|+1, |B_{23}|+1, -|B_{23}|+1, -|B_{12}|+1\right).$$

Note that for $B_{12}B_{23} \neq 0$ none of the joint eigenvalue sequences majorizes the other one. This example should be compared with a theorem of Bapat and Sunder [BS85], concerning the special case in which $I \cap J = \emptyset$: then, the sequence obtained by concatenating the eigenvalues of $A[I]$ and $A[J]$ is majorized by the sequence of eigenvalues of $A[I \cup J]$, and so, the inequality (1.4) for the map $I \mapsto \mathrm{tr}\, A[I]^p$ holds for for all $p \in (-\infty, 0) \cup [1, \infty)$ if $I \cap J = \emptyset$.

Example 18 There are negative values of p for which $T_p(\cdot, A)$ is not supermodular. Consider for instance the positive definite matrix

$$A = \begin{pmatrix} 5 & -12 & 9 \\ -12 & 33 & -24 \\ 9 & -24 & 19 \end{pmatrix}, \quad I = \{1, 2\}, J = \{1, 3\}, \quad \delta(t^{-1}, I, J, A) = 16/35 > 0 \ .$$

The following counter example shows that Theorem 1 does not carry over to operator convex maps.

Example 19 The map $f(t) = t^2(\lambda+t)^{-1}$ is known to be operator convex [Bha97, Problem V.5.5]. Consider the positive definite matrix

$$A = \begin{pmatrix} 1 & -2 & -2 \\ -2 & 6 & 4 \\ -2 & 4 & 8 \end{pmatrix}, \quad \lambda = 1, \quad I = \{1, 2\}, J = \{1, 3\} \ .$$

Then, $\delta(f, I, J, A) = 44/1085 > 0$, showing that the map $K \mapsto \mathrm{tr}\, f(A[K])$ is not supermodular.

We now summarize in Table 1 the submodularity and supermodularity properties we know, omitting the proofs of the two easy cases.

$p = -1, m \geqslant 3$	may not be supermodular	Ex. 18
$p < 0, m \geqslant 3$	may not be submodular	Prop. 16
$p < 0, m = 2$	supermodular	Omitted
$0 \leqslant p \leqslant 1$	submodular	Coro. 2
$1 \leqslant p \leqslant 2$	supermodular	Coro. 2
$p > 2, m = 2$	supermodular	Omitted
$p = 2, m \geqslant 3$	may not be submodular	Prop. 15
$p \in (2, 3), m \geqslant 3$	may not be supermodular	Prop. 16
$p > 3, m \geqslant 3$	may not be submodular	Prop. 16

Table 1: Summary of submodularity and supermodularity properties of the spectral function $T_p(I, A) = \mathrm{tr}\, A[I]^p$ for a $m \times m$ nonnegative definite Hermitean matrix A.

Recall that for $A, B \in \mathrm{H}_m$ and $A \leqslant B$ we have that $\mathrm{tr}\, A \leqslant \mathrm{tr}\, B$. Observe also that $\mathrm{H}_m(\mathcal{E})$ is a convex subset of H_m. Since the primitive of an operator monotone function is operator convex we deduce from Theorem 1:

14

Corollary 20 *Let f satisfies the assumptions of Theorem 1. Then for each $I \subseteq [m]$ the function $A \mapsto \operatorname{tr} f(A[I])$ is a convex function on $\mathrm{H}_m(\mathcal{E})$.* $\qquad\qquad\square$

Remark 21 A different submodularity inequality involving spectral functions appeared in [BGS08], with an application to an experiment design problem. It is shown there that for all nonnegative definite hermitian matrices A, B, C, and for all $0 \leqslant p \leqslant 1$, $\operatorname{tr}(A + B + C)^p + \operatorname{tr} C^p \leqslant \operatorname{tr}(A + C)^p + \operatorname{tr}(B + C)^p$.

References

[AZ99] T. Ando and X. Zhan. Norm inequalities related to operator monotone functions. *Math. Ann.*, 315(4):771–780, 1999.

[BGS08] M. Bouhtou, S. Gaubert, and G. Sagnol. Optimization of network traffic measurement : A semidefinite programming approach. In *International Conference on Engineering Optimization (EngOpt 2008)*, Rio de Janeiro, June 1-5 2008. Paper # 0615, Printed and CDROM proceedings, ISBN 978-85-7650-152-7.

[Bha97] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[BS85] R. B. Bapat and V. S. Sunder. On majorization and Schur products. *Linear Algebra Appl.*, 72:107–117, 1985.

[CCPV07] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák. Maximizing a submodular set function subject to a matroid constraint (extended abstract). In *Integer programming and combinatorial optimization*, volume 4513 of *Lecture Notes in Comput. Sci.*, pages 182–196. Springer, Berlin, 2007.

[Car67] D. Carlson, Weakly sign-symmetric matrices and some determinantal inequalities, *Colloq. Math.* 17 (1967), 123129.

[CT88] T. M. Cover and J. A. Thomas. Determinant inequalities via information theory. *SIAM J. Matrix Anal. Appl.*, 9(3):384–392, 1988.

[Fan67] Ky Fan. Subadditive functions on a distributive lattice and an extension of Szász's inequality. *J. Math. Anal. Appl.*, 18:262–268, 1967.

[Fan68] Ky Fan. An inequality for subadditive functions on a distributive lattice, with application to determinantal inequalities. *Linear Algebra and Appl.*, 1(1):33–38, 1968.

[FJ00] S. M. Fallat and C. R. Johnson. Determinantal inequalities: ancient history and recent advances. In *Algebra and its applications (Athens, OH, 1999)*, volume 259 of *Contemp. Math.*, pages 199–212. Amer. Math. Soc., Providence, RI, 2000.

[Fri73] S. Friedland, Extremal eigenvalue problems for convex sets of symmetric matrices and operators, *Israel J. Math.* 15 (1973), 311-331.

[Fbk] S. Friedland, *Matrices*, a book in preparation, http://www2.math.uic.edu/∼friedlan/bookm.pdf

[FMMN] S. Friedland, V. Mehrmann, A. Miedlar and M. Nkengla, Fast low rank approximations of matrices and tensors, *Electronic Journal of Linear Algebra*, *to appear*, http://www2.math.uic.edu/∼friedlan/FriMMN08_7.2.pdf

[Gan59] F. R. Gantmacher, "*The theory of matrices*", Vols. I and II, Chelsea, New York, 1959.

[GK60] F. R. Gantmacher and M. G. Kreĭn. *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme.* Wissenschaftliche Bearbeitung der deutschen Ausgabe: Alfred Stöhr. Mathematische Lehrbücher und Monographien, I. Abteilung, Bd. V. Akademie-Verlag, Berlin, 1960.

[GVL96] G.H. Golub and C.F. Van Loan, *Matrix Computations*, John Hopkins Univ. Press, 3rd Ed., Baltimore, 1996.

[GTZ97] S.A. Goreinov, E.E. Tyrtyshnikov, and N.L. Zamarashkin, A theory of pseudo-skeleton approximations of matrices, *Linear Algebra Appl.* 261 (1997), 1 – 21.

[Iwa08] S. Iwata. Submodular function minimization. *Math. Program.*, 112(1, Ser. B):45–64, 2008.

[JB85] C. R. Johnson and W. W. Barrett. Spanning-tree extensions of the Hadamard-Fischer inequalities. *Linear Algebra Appl.*, 66:177–193, 1985.

[KK83] A. K. Kelmans and B. N. Kimel′fel′d. Multiplicative submodularity of a matrix's principal minor as a function of the set of its rows, and some combinatorial applications. *Discrete Math.*, 44(1):113–116, 1983.

[Kot50] D. M. Kotelyanskiĭ. On the theory of nonnegative and oscillating matrices. *Ukrain. Mat. Žurnal*, 2(2):94–101, 1950.

[Löw34] K. Löwner. über monotone matrixfunktionen. *Math. Z*, 38:177–216, 1934.

[NWF78] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. I. *Math. Programming*, 14(3):265–294, 1978.

[Sta97] R. Stanley. *Enumerative combinatorics*, volume I, Cambridge Universiy Press, 1997.

[Zha02] X. Zhan. *Matrix inequalities*, volume 1790 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002.