AN ALGORITHM FOR MAXIMIZING A QUOTIENT OF TWO HERMITIAN FORM DETERMINANTS WITH DIFFERENT EXPONENTS

Raphael Hunger, Paul de Kerret, and Michael Joham

ABSTRACT

We investigate the maximization of a quotient of two determinants with different exponents under a Frobenius norm constraint, where each determinant is taken from a matrix-valued Hermitian form. The optimum matrix that constitutes the Hermitian forms is shown to be a scaled partial isometry. We prove monotonic convergence of the iterative algorithm, which means that the utility increases in every step. In addition, we prove structural properties of the optimum matrix that constitutes the Hermitian forms instead of the determinant operator. However, for scalar arguments, i.e., for vector-valued Hermitian forms, both operators lead to the same outcome. In contrast to our contribution, only two discrete exponents (namely one and two) in the numerator of the Hermitian form are treated in [3] and the authors come up with algorithms only for those two special cases by means of majorization.

2. PROBLEM FORMULATION

The maximization of the quotient of two determinants with different exponents has the form

\[
\maximize_{X \in \mathbb{C}^{M \times N}} \frac{|X^H B X|^\alpha}{|X^H A X|^\beta} \quad \text{s.t.: } \|X\|^2_F = N,
\]

where \( A \in \mathbb{S}_M^+ \) is positive definite and \( B \in \mathbb{S}_M \) is non-negative definite with \( \text{rank}(B) \geq N \). The optimization variable \( X \in \mathbb{C}^{M \times N} \) is constrained to have a squared Frobenius norm \( N \), where \( M \geq N \) clearly has to hold. The real-valued scalar \( \alpha \) is assumed to be larger than or equal to one. Special attention will be paid to the vector-valued variant of (1), which reads as

\[
\maximize_{x \in \mathbb{C}^M} \frac{|x^H B x|^\alpha}{|x^H A x|^\beta} \quad \text{s.t.: } \|x\|^2_2 = 1,
\]

since for this maximization, we will derive an iterative algorithm with monotonic convergence. For the matrix-valued variant in (1), we derive structural properties of the optimum matrix \( X \).

3. STRUCTURAL PROPERTIES OF MATRIX-VALUED HERMITIAN FORM DETERMINANTS

3.1. Special Case: Same Exponent

For the less relevant case, when \( \alpha = 1 \), the Frobenius norm constraint in (1) is only weakly active as a scaling of \( X \) does not influence the objective. In this case, the optimization can be regarded as unconstrained since the Frobenius norm constraint \( \|X\|^2_F = N \) can afterwards always be assured without changing the objective. The optimum matrix \( X \) follows from choosing the \( N \) principal eigenvectors of the...
matrix $A^{-1}B$ belonging to the $N$ largest arbitrarily scaled eigenvalues and afterwards rescaling them jointly such that $\|X\|_{F}^2 = N$, whereas the optimum objective is given by the product of those $N$ largest eigenvalues. This can be shown as follows: Let $X = QR'$ denote the QR-decomposition of $X$, where $R' \in \mathbb{C}^{N \times N}$ is a full-rank upper triangular matrix and $Q' \in \mathbb{C}^{M \times N}$ satisfies $Q^H Q' = I_N$. Then,

$$\frac{|X^H B X|}{X^H A X} = \frac{|Q^H B Q'|}{|Q^H A Q'|}$$

which means that the utility only depends on the basis $Q'$ and not on $R'$. Defining $Y := A^2 Q'$ and using again the QR-decomposition $Y = QR$ and the fact that only the basis $Q$ governs the determinant quotient, we obtain

$$\frac{|X^H B X|}{X^H A X} = \frac{|Q^H A^{-\frac{1}{2}} B A^{-\frac{1}{2}} Q|}{|Q^H Q|} = |Q^H C Q|$$

with $C := A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$. Subject to $Q^H Q = I_N$, above determinant is maximized when the columns of $Q$ are the $N$ unit-norm principal eigenvectors of $C$ belonging to the $N$ largest eigenvalues. We omit the proof due to lack of space. In this case, the determinant corresponds to the product of the $N$ largest eigenvalues of $C = A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$. This utility is also achieved when the columns of $X$ are chosen to be the $N$ principal eigenvectors of the matrix $A^{-1}B$, such that $B X = A X A$, where the diagonal $N \times N$ matrix $A$ contains the $N$ dominant eigenvalues of $A^{-1}B$ that are equal to those of $C$. Inserting $B X = A X A$ into the utility yields $|X^H B X|/|X^H A X| = |A|$. So, choosing the columns of $X$ as the $N$ principal eigenvectors of $A^{-1}B$ is optimal.

### 3.2. Special Case: Square Matrix

Given the special case with $N = M$, any unitary matrix $\tilde{X}$ with squared Frobenius norm $M$ maximizes the objective. This follows from the arithmetic-geometric inequality as the utility can be decoupled into a constant times $|X^H X|^{\alpha-1}$ which has to be maximized under the squared Frobenius norm constraint $\|X\|_{F}^2 = N$. Hence, the product of the eigenvalues of $X^H X$ has to be maximized while their sum is fixed to $N$. As a consequence, all eigenvalues have to be identical, and since they are real-valued, all of them are equal to one. The eigenvalues of $\tilde{X}$ therefore may take arbitrary complex unit norm values and thus, $\tilde{X}$ is unitary. With this choice, the optimum objective simplifies to $|B|^\alpha / |A|$.

### 3.3. General Case: Different Exponents and Tall Matrix

In the general case, where $\alpha > 1$ and $N < M$ holds, the Frobenius norm constraint is strongly active. Statements on the structure of the optimum matrix $\tilde{X}$ are deduced from setting the derivative of the Lagrangian function for (1) to zero:

$$\mathcal{L}(X, \mu) := \frac{|X^H B X|^{\alpha}}{|X^H A X|} - \mu \left[ \text{tr}(X^H X) - N \right]$$

The optimum Lagrangian multiplier $\bar{\mu}$ can be computed via

$$\text{tr} \left[ \frac{X^H \partial \mathcal{L}(X, \bar{\mu})}{\partial X^*} \right]_{X = X, \mu = \bar{\mu}} = 0$$

$$\Leftrightarrow \quad \bar{\mu} = (\alpha - 1) \frac{|X^H B X|^{\alpha - 1}}{|X^H A X|},$$

which means that $\bar{\mu}$ is $\alpha - 1$ times the optimum objective and thus vanishes when $\alpha = 1$ as mentioned in Section 3.1. Moreover, the KKT condition which the optimum matrix $\tilde{X}$ has to fulfill follows from $\partial \mathcal{L}(X, \mu)/\partial X^* |_{X = \tilde{X}, \mu = \bar{\mu}} = 0$ and reads by means of (4) as

$$\alpha B X (X^H B X)^{-1} - A X (X^H A X)^{-1} = (\alpha - 1) \tilde{X}. \quad (5)$$

Noticing that the optimum Lagrangian multiplier $\bar{\mu} > 0$ is positive, we can left-hand-side multiply the derivative of $\mathcal{L}(X, \mu)$ with respect to $X^*$ by $X^H$ and obtain

$$X^H \frac{\partial \mathcal{L}(X, \mu)}{\partial X^*} |_{X = \tilde{X}, \mu = \bar{\mu}} = 0 \Leftrightarrow \tilde{X}^H \tilde{X} = I_M, \quad (6)$$

so $\tilde{X} \in \mathbb{C}^{M \times N}$ is a partial isometry and $\tilde{X} X^H$ is a projector.

### 4. Algorithmic Solution for Vector-Valued Hermitian Forms

When $N = 1$, the matrix $X$ reduces to a column vector $x$ and the objective reduces to [cf. (2)]

$$f(x) := \frac{(x^H B x)^\alpha}{x^H A x}. \quad (7)$$

Thus, the determinant maximization problem in (1) simplifies to the one in (2), and in turn, the Lagrangian multiplier in (4) can be expressed as $\bar{\mu} = (\alpha - 1) f(\bar{x})$. Similar to the matrix-valued first order optimality condition in (5), the vector valued version reads as

$$\frac{\alpha}{x^H B x} B \bar{x} - \frac{1}{x^H A x} A \bar{x} = (\alpha - 1) \bar{x}. \quad (8)$$

From the various ways to rearrange above nonlinear fixed point equation, the following one has the nice property that an iterative fixed point algorithm based on its particular eigenproblem-structure features monotonic convergence:

$$\left[ \alpha (x^H B x)^{\alpha-1} B - (\alpha - 1) (x^H B x)^{\alpha - 1} I \right] \bar{x} = f(\bar{x}) A \bar{x}. \quad (8)$$

Given this fixed point equation, we can readily define the fixed point iteration, which serves as the core of the algorithm:

$$x_{n+1} = \lambda_{n+1} A x_{n+1}. \quad (9)$$

Note that $x_{n+1}$ must be chosen to have norm one due to the constraint in (2) and that $x_{n+1}$ is the principal eigenvector of
the matrix $\mathbf{A}^{-1}\mathbf{M}_n$ belonging to the largest eigenvalue $\lambda_{n+1}$, where $\mathbf{M}_n$ is defined via

$$M_n := \alpha(x_n^H \mathbf{B} x_n)^{\alpha-1} \mathbf{B} - (\alpha - 1)(x_n^H \mathbf{B} x_n)^\alpha \mathbf{I}. \quad (10)$$

Let $x_0 \notin \text{null}(\mathbf{B})$ define an arbitrary unit-norm initialization vector. With $x_0$, the matrix $\mathbf{M}_0$ is composed and the principal eigenvector of the matrix $\mathbf{A}^{-1}\mathbf{M}_0$ constitutes the vector $\mathbf{x}_1$ of the next iteration. After that, the matrix $\mathbf{M}_1$ is computed, left-hand side multiplied with the inverse of $\mathbf{A}$, and its unit-norm principal eigenvector is chosen for $x_2$ and so on. In general, the eigenvalue $\lambda_{n+1}$ in iteration $n + 1$ reads as

$$\lambda_{n+1} = \frac{x_{n+1}^H \mathbf{M}_n x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}} = \frac{x_{n+1}^H \mathbf{M}_n x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}}. \quad (11)$$

A key observation is that the utility $f(x_n)$ in iteration $n$ can not only be expressed as a quotient of Hermitian forms where the matrices $\mathbf{A}$ and $\mathbf{B}$ are involved as in (7), but also as a function of the matrix $\mathbf{M}_n$ [cf. (8)]:

$$f(x_n) = \frac{x_n^H \mathbf{M}_n x_n}{x_n^H \mathbf{A} x_n}. \quad (12)$$

With this definition, we can now prove that the sequence $f(x_n)$ is increasing in $n$ as long as $n < \infty$, i.e., as long as the algorithm has not converged yet. First of all, we observe that the sequence $f(x_n)$ is bounded via

$$f(x_n) \leq \frac{\text{maxeig}(\mathbf{B})}{\text{mineig}(\mathbf{A})} < \infty, \quad (13)$$

since $\mathbf{A}$ is positive definite (its minimum eigenvalue is larger than zero) and $\mathbf{B}$ has a finite Frobenius norm. Due to the fact that $x_{n+1}$ is chosen as the generalized principal eigenvector of the matrix pair $\mathbf{M}_n$ and $\mathbf{A}$, see (9), the Lagrangian multiplier $\lambda_{n+1}$ from (11) is always larger than the objective $f(x_n)$ defined in (12) for finite $n$:

$$f(x_n) = \frac{x_n^H \mathbf{M}_n x_n}{x_n^H \mathbf{A} x_n} \leq \frac{x_{n+1}^H \mathbf{M}_n x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}} = \lambda_{n+1}. \quad (14)$$

Equality in (14) only holds for $x_n = x_{n+1}$, i.e., for $n \to \infty$ when the algorithm has converged, or when the initial vector $x_0$ is chosen to lie in the nullspace of $\mathbf{B}$, i.e., when $x_0 \in \text{null}(\mathbf{B})$. In this case, the matrix $\mathbf{M}_0$ in (10) is zero and both the objective $f(x_0)$ and the Lagrangian multiplier $\lambda_1$ are zero according to (9). In the next step with $n = 1$, the vector $\mathbf{x}_1$ can be chosen arbitrarily as long as it has unit-norm, since $0 = 0 \cdot \mathbf{A} \mathbf{x}_1$ from (9) holds for any $\mathbf{x}_1$. This time, $\mathbf{x}_1$ should be chosen not to lie in the null-space of $\mathbf{B}$ because otherwise, the objective will remain zero whenever $x_n$ is chosen to satisfy $x_n \in \text{null}(\mathbf{B})$. However, this can easily be detected and avoided and we therefore exclude the case $x_0 \in \text{null}(\mathbf{B})$. The difference between two consecutive objectives $f(x_{n+1})$ and $f(x_n)$ can be written as [cf. (12)]

$$f(x_{n+1}) - f(x_n) = \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}} - \frac{x_n^H \mathbf{M}_n x_n}{x_n^H \mathbf{A} x_n} \geq \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1} - x_n^H \mathbf{M}_n x_n}{x_{n+1}^H \mathbf{A} x_{n+1}}$$

$$\geq \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1} - x_n^H \mathbf{M}_n x_n}{x_{n+1}^H \mathbf{A} x_{n+1}} \geq \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}} - \frac{x_n^H \mathbf{M}_n x_n}{x_n^H \mathbf{A} x_n} = \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1} - x_n^H \mathbf{M}_n x_n}{x_{n+1}^H \mathbf{A} x_{n+1}}$$

$$= \frac{x_{n+1}^H \mathbf{B} x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}} = f(x_{n+1}) \cdot d(\beta) \geq 0.$$

$$f(x_{n+1}) - f(x_n) = \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}} - \frac{x_n^H \mathbf{M}_n x_n}{x_n^H \mathbf{A} x_n} \geq \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1} - x_n^H \mathbf{M}_n x_n}{x_{n+1}^H \mathbf{A} x_{n+1}}$$

$$\geq \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1} - x_n^H \mathbf{M}_n x_n}{x_{n+1}^H \mathbf{A} x_{n+1}} \geq \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}} - \frac{x_n^H \mathbf{M}_n x_n}{x_n^H \mathbf{A} x_n} = \frac{x_{n+1}^H \mathbf{M}_{n+1} x_{n+1} - x_n^H \mathbf{M}_n x_n}{x_{n+1}^H \mathbf{A} x_{n+1}}$$

$$= \frac{x_{n+1}^H \mathbf{B} x_{n+1}}{x_{n+1}^H \mathbf{A} x_{n+1}} = f(x_{n+1}) \cdot d(\beta) \geq 0.$$
Again, \( d(\beta) \) is larger than or equal to zero and therefore, the difference \( f(x_{n+1}) - \lambda_{n+1} \) is nonnegative and equal to zero only when \( d(\beta) = 0 \), i.e., when \( \beta = 1 \).

A pseudo-code algorithmic implementation is shown in Algorithm 1. In Lines 1-3, the iteration counter \( n \) is initialized and the relative accuracy \( \epsilon \) is defined. Inside the loop, Lines 6-8 first set up the matrix \( M_n \) and afterwards compute the principal eigenvector of the matrix product \( A^{-1}M_n \) to have unit norm. As a termination criterion, the relative change in the utility is checked against the threshold \( \epsilon \) in Line 10.

5. SIMULATION RESULTS

For the simulation results, we used the problem dimension \( M = 5 \) and randomly picked the two positive definite complex-valued matrices \( A \) and \( B \). The initialization vector \( x_0 \) was chosen randomly as well and the threshold for the relative change in the utility that serves as a termination criterion was set to \( \epsilon = 10^{-4} \) as used in Algorithm 1. A large scale view on the utility \( f(x_n) \) and the eigenvalue \( \lambda_n \) versus the iteration counter \( n \) is shown in Fig. 1. The dashed curve corresponds to the utilities starting at \( n = 0 \) whereas the solid one shows the eigenvalues starting at \( n = 1 \). We observe that the initialization vector \( x_0 \) achieves a very low utility but only a single iteration increases the utility almost to its maximum. In addition, we verify the interlacing property \( f(x_n) \leq \lambda_{n+1} \leq f(x_{n+1}) \) from (14) and (17). The vertical plot with the circle marker denotes the iteration index at which the relative change in the utility is smaller than the threshold \( \epsilon \) which corresponds to the iteration index \( n \) at which the algorithm terminates. Fig. 2 shows the absolute error \( f(\hat{x}) - f(x_n) \) versus the iteration index \( n \). Its linear characteristic in the semilogarithmic figure indicates a linear speed of convergence. Due to finite word-length precision, the curve saturates, and the circle marker denotes the smallest \( n \) for which the relative error is at most \( \epsilon \).

6. CONCLUSION

In this paper, we derived an algorithm for the maximization of a quotient of two Hermitian forms with different exponents under a Frobenius norm constraint. It is based on an iterative computation of the principal eigenvector of a matrix that depends on the eigenvector of the previous iteration and we have shown that it monotonically converges to a local optimum. For the matrix valued case where the determinant operator is applied to the matrix-valued Hermitian forms, we found out that the optimum matrix is a partial isometry.

7. REFERENCES

