ILLINOIS JOURNAL OF MATHEMATICS
Volume 24, Number 1, Spring 1980

UNITARY APPROXIMATION OF POSITIVE OPERATORS

BY

JOHN G. AIKEN, JOHN A. ERDOS AND JEROME A. GOLDSTEIN[1]

0. Abstract

Of concern are some operator inequalities arising in quantum chemistry. Let A be a positive operator on a Hilbert space $\mathscr{H}$. We consider the minimization of $\|U - A\|_p$ as U ranges over the unitary operators in $\mathscr{H}$ and prove that in most cases the minimum is attained when U is the identity operator. The norms considered are the Schatten p-norms. The methods used are of independent interest; application is made of noncommutative differential calculus.

1. Introduction and preliminaries

The motivation for considering the quantity $\|U - A\|_p$ as the unitary operator U varies, arises from the following considerations. In molecular orbital calculations, a finite set of vectors $\mathscr{F} = \{f_1, f_2, \ldots, f_n\}$ is chosen for physical, chemical, or computational reasons and is used to reduce the Hamiltonian of the molecule in question to a finite-dimensional matrix. This set of vectors is often replaced by an orthonormal set $\mathscr{E} = \{e_1, e_2, \ldots, e_n\}$ to reduce the complexity of the computation. The linear transformation B defined by $e_i = Bf_i$ $(1 \leq i \leq n)$ is called an *orthogonalization* of $\mathscr{F}$ (although orthonormalization seems to be a better term).

It is desirable to choose B so that the e_i are, in some sense, as close as possible to the f_i. If B is chosen to be the Löwdin orthogonalization L, then it is known that the expression

$$E(B) = \sum_{i=1}^{n} \|Bf_i - f_i\|^2$$

is as small as possible when $B = L$[1, 2, 3, 12]. In other words $E(L) \leq E(B)$ for each orthogonalization B of $\mathscr{F}$. Similar results have been proved for certain unconditional bases in the infinite-dimensional case [8, Chapter VI, Theorem 3.4].

This theorem can be interpreted as a minimization problem involving Hilbert-Schmidt norms. In the present paper we give this reformulation, shar-

[1] This author's contribution was made, in part, while he was at the University of Brasilia. He thanks Professors Geraldo S. S. Avila and Djairo G. de Figueiredo for their gracious hospitality.

Received October 11, 1977.

pen the results of [1] considerably and give extensions of the results to the operator norm and the other Schatten p-norms. We also establish that E has a unique local minimum. This is an important theoretical result in that it can be used to justify numerical methods for computing Löwdin orthogonalizations in quantum chemistry.

In this paper the term *Hilbert space* will mean *complex* Hilbert space and the inner product will be denoted by $\langle \cdot , \cdot \rangle$. The term *operator* will mean *bounded linear* operator and the spectrum of an operator A will be denoted by $\sigma(A)$. The set of all operators on a Hilbert space $\mathscr{H}$ will be denoted by $\mathscr{B}(\mathscr{H})$.

We refer to [5] or [8] for properties of the Schatten p-classes. The following is a brief summary of the results we require. For any compact operator A let $|A| = (A^* A)^{1/2}$ and $s_1(A), s_2(A), \ldots$ be the (positive) eigenvalues of $|A|$ in decreasing order and repeated according to multiplicity. If, for some $p > 0$,

$$\sum_{i=1}^{\infty} s_i(A)^p < \infty,$$

we say that A is in the *Schatten p-class* $\mathscr{C}_p$ and write

$$\|A\|_p = \left(\sum_{i=1}^{\infty} s_i(A)^p \right)^{1/p}.$$

If $p \geq 1$, $\mathscr{C}_p$ with the above norm is a Banach space; if $0 < p < 1$, $\mathscr{C}_p$ is a metric space with metric d given by $d(X, Y) = \sum_{i=1}^{\infty} s_i(X - Y)^p$. For all $p > 0$, $\mathscr{C}_p$ is an ideal of $\mathscr{B}(\mathscr{H})$.

The class $\mathscr{C}_1$ is called the *trace class*. If $A \in \mathscr{C}_1$ and $\{x_i\}$ is any orthonormal basis of $\mathscr{H}$ then the quantity $\tau(A)$ defined by $\tau(A) = \sum_{i=1}^{\infty} \langle Ax_i, x_i \rangle$ is independent of the choice of $\{x_i\}$ and is called the *trace* of A. Note that if $A \in \mathscr{C}_p$ then $\|A\|_p^p = \tau(|A|^p)$. The class $\mathscr{C}_2$ is called the *Hilbert-Schmidt class* and is a Hilbert space under the inner product $\langle A, B \rangle = \tau(B^* A)$. If $p \geq 1$ and $1/p + 1/q = 1$ then for $A \in \mathscr{C}_p$, $B \in \mathscr{C}_q$

$$|\tau(AB)| \leq \|AB\|_1 \leq \|A\|_p \|B\|_q.$$

We state, for reference, the following standard theorem (see, for example [9, Problem 105])

THEOREM 1.1. *If X is any bounded operator on a Hilbert space then X may be expressed uniquely as $X = UP$ where P is a positive operator and U is a partial isometry such that* ker $(U) =$ ker (P). *If X is selfadjoint then U is selfadjoint and commutes with P.*

This unique expression UP for X will be referred to as the *polar decomposition* of X. Note that $P = (X^* X)^{1/2}$. Recall that a selfadjoint unitary operator is called a *symmetry* and every symmetry V is of the form $E - F$ where E and F are orthogonal projections with $EF = 0$ and $E + F = I$.

Our main results generalize the following theorem.

THEOREM 1.2 [1], [2], [3], [8], [12]. *Let $\mathscr{F} = \{f_1, \ldots, f_n\}$ be a basis of unit vectors for a finite-dimensional Hilbert space. The expression $E(B) = \sum_{i=1}^{n} \|Bf_i - f_i\|^2$ subject to the condition that $\{Bf_i : 1 \leq i \leq n\}$ is orthonormal, has a unique minimum. This minimum is attained when B is the unique positive operator that makes $\{Bf_i\}$ orthonormal.*

We now reformulate the above. If B is any operator such that $\{Bf_i\}$ is orthonormal then clearly B is invertible and so, by the polar decomposition, $B = UL$ where U is unitary and L is invertible and nonnegative. If B' is any other operator such that $\{B'f_i\}$ is orthonormal, then $B' = WB$ where W is the unitary map defined by $WBf_i = B'f_i$. Thus $B' = U'L$ when $U' = WU$ is unitary. Therefore varying B subject to $\{Bf_i\}$ being orthonormal is equivalent to varying U over all unitary operators while keeping L fixed. Then, writing $e_i = Bf_i$ we have

$$\begin{aligned} E(B) &= \sum_{i=1}^{n} \|Bf_i - f_i\|^2 \\ &= \sum_{i=1}^{n} \|(I - B^{-1})e_i\|^2 \\ &= \sum_{i=1}^{n} \|(U - L^{-1})U^{-1}e_i\|^2 \\ &= \|U^{-1} - L^{-1}\|_2^2 \end{aligned}$$

since $\{Ue_i\}$ is an orthonormal basis of H. Thus minimizing $E(B)$ becomes a problem of minimizing $\|V - L^{-1}\|_2$ as V ranges over all unitary operators.

The generalized problem is to consider the function $F_p(U) = \|U - A\|_p^p$ for fixed positive A as U varies over all unitaries with $U - A \in \mathscr{C}_p$. We use methods of the calculus. Briefly, our main results state that for $p > 1$, the operator U is a critical point of F_p if and only if U is a symmetry commuting with A and F_p has a unique local and global minimum when $U = I$. These results can be considered as a contribution to the theory of non-commutative approximation as developed by Halmos [10], giving (in the terminology of [10]) I as a unitary approximant in $\mathscr{C}_p$ norm to any positive operator. In fact, the results are a little more general, since if X is an operator with $\ker X = \ker X^*$ then X may be written as $X = U_0 A$ where U_0 is unitary and A is positive. Since if $U - X \in \mathscr{C}_p$ for some $p > 0$ then

$$\|U - X\|_p = \|U_0^* U - A\|_p,$$

our results show that U_0 is a unitary approximant to X in $\mathscr{C}_p$ norm.

We thank the referee for doing a very thorough job and making helpful comments.

2. Differentiation of the norm in $\mathscr{C}_p$

In this section we find the derivative of the norm in the Banach space $\mathscr{C}_p$. It will be observed that the full force of these results are not used in Section 3 since the quantity $\|X + T\|_p^p - \|X\|_p^p$ is considered there only in the case when T is one-dimensional. However, we feel that the inclusion of these results is justified since they are of independent interest and are likely to have further applications. The real part of a complex number z will be denoted by $\mathscr{R}(z)$.

THEOREM 2.1. *If $p > 1$, the map $X \mapsto \|X\|_p^p$ ($\mathscr{C}_p \to \mathbf{R}$) is (Fréchet) differentiable with derivative D_X at X given by*

$$D_X(T) = \tfrac{1}{2}p\tau(|X|^{p-1}U^*T + T^*U|X|^{p-1}) = p\mathscr{R}\tau(|X|^{p-1}U^*T)$$

where $|X|$ is the positive square root of X^*X and $X = U|X|$ is the polar decomposition of X. If the underlying Hilbert space is finite-dimensional, the same result holds for $0 < p \leq 1$ at every invertible element X.

Proof. It has been shown by McCarthy [11] that $\mathscr{C}_p$ $(1 < p < \infty)$ is uniformly convex. It follows from a standard result (see e.g. [4 p. 36, Theorem 1]) that $\mathscr{C}_p$ has Fréchet differentiable norm. Hence the map $X \mapsto \|X\|_p^p$ is differentiable and for the case $1 < p < \infty$, it only remains to establish the formula for the derivative.

Let $X, T \in \mathscr{C}_p$ and let E be any projection commuting with X. We claim that

$$D_X[ET(I - E)] = 0.$$

A calculation shows that

$$(2E - I)[X + ET(I - E)](2E - I) = X - ET(I - E).$$

Observe that $2E - I$ is unitary. It is therefore clear that

$$\|X + ET(I - E)\|_p^p = \|X - ET(I - E)\|_p^p$$

or, in other words $\|X + ET(I - E)\|_p^p$ is an even function of T. Thus the derivative vanishes at zero which establishes the claim.

Now consider the case when X is positive. Then $X = \sum_{i=1}^{\infty} \lambda_i(x_i \otimes x_i)$ where $\lambda_i \geq 0$ and $\{x_i\}$ is an orthonormal basis. Let E_i be the projection onto the space spanned by x_i and let $F_n = I - \sum_{i=1}^{n} E_i$. Since, from above, $D_X(E_1TF_1) = D_X(F_1TE_1) = 0$ we have

$$D_X(T) = D_X(E_1TE_1) + D_X(F_1TF_1).$$

Repeating the above argument, a simple induction now shows that for any integer n,

$$D_X(T) = \sum_{i=1}^{n} D_X(E_iTE_i) + D_X(F_nTF_n).$$

It is elementary that

$$D_X(E_i T E_i) = p\mathscr{R}[\lambda_i^{p-1}\langle Tx_i, x_i\rangle] = p\mathscr{R}\langle X^{p-1}Tx_i, x_i\rangle.$$

(Perhaps the easiest way to see this is to evaluate

$$\frac{d}{dt}\|X + tE_i TE_i\|_p^p\bigg]_{t=0}$$

after noting that $\|X + tE_i TE_i\|_p^p = |\lambda_i + t\langle Tx_i, x_i\rangle|^p + \sum_{j\neq i}\lambda_j^p$.) Hence

$$D_X(T) = p\sum_{i=1}^{n}\mathscr{R}\langle X^{p-1}Tx_i, x_i\rangle + D_X(F_n TF_n).$$

Since (F_n) converges strongly to zero as $n\to\infty$, it follows (see [6, Lemma 2]) that $(F_n TF_n)$ converges to zero in $\mathscr{C}_p$. Since D_X is continuous, $(D_X(F_n TF_n))\to 0$ and we see that

$$D_X(T) = p\sum_{i=1}^{\infty}\mathscr{R}\langle X^{p-1}Tx_i, x_i\rangle = p\mathscr{R}\tau(X^{p-1}T)$$

which is the required formula in the case when $X\geq 0$.

Now let X be any element of $\mathscr{C}_p$ and let $X = U|X|$ be its polar decomposition. It is well known that there exists V such that either V or V^* is an isometry and such that V and U coincide on $(\ker |X|)^\perp$. Thus $X = V|X|$. If V^* is an isometry then for any $T\in\mathscr{C}_p$ we have (using $VV^* = I$)

$$|X + T|^2 = |X|^2 + |X|V^*T + T^*V|X| + T^*VV^*T = ||X| + V^*T|^2$$

and so $\|X + T\|_p^p = \||X| + V^*T\|_p^p$. Therefore $D_X(T) = D_{|X|}(V^*T)$. Thus

$$D_X(T) = D_{|X|}(V^*T) = p\mathscr{R}\tau(|X|^{p-1}V^*T) = p\mathscr{R}\tau(|X|^{p-1}U^*T)$$

as required. In the case when V is an isometry, taking the adjoint and performing a similar computation yields

$$D_X(T) = D_{|X^*|}(VT^*) = p\mathscr{R}\tau(|X^*|^{p-1}VT^*).$$

Use $|X^*|^{p-1} = V|X|^{p-1}V^*$ and the desired result follows easily.

In finite dimensions, if X is invertible then the eigenvalues of $|X|$ are bounded away from zero and the above proof of the formula holds also in the case $0 < p \leq 1$. However, in this case a separate argument is needed to show that the map $X\mapsto\|X\|_p^p$ is differentiable. To see this we decompose the map thus: choose n such that $2^n p > 1$ and consider the maps

$$X\mapsto |X|^2\mapsto |X|^{1/2^n}\mapsto \||X|^{1/2^n}\|_{2^n p}^{2^n p} = \|X\|_p^p.$$

The derivative of the first map at A is the map $X\mapsto A^*X + X^*A$. That the final map is differentiable follows from the main result. To show the differentiability of the remaining map it is sufficient to show that the map $X\mapsto X^{1/2}$ is differentiable as a map between strictly positive operators. By the inverse mapping theorem it is enough to show that the inverse map, that is $X \mapsto X^2$, has an invertible derivative. The derivative of $X \mapsto X^2$ at A is $X \to AX + XA$ and it follows from a result of Rosenblum (see e.g. [13 p. 8, Corollary 0. 13]) that this map is invertible whenever $\sigma(A) \cap \sigma(-A) = \emptyset$. Since this last condition is satisfied if A is a positive invertible operator it follows that the map $X \mapsto \|X\|_p^p$ is differentiable and the proof is complete. ▮

Note that in infinite dimensions the conditions $X \in \mathscr{C}_p$ and X invertible are incompatible and the conclusion of the theorem is meaningless for $p < 1$. Also, if $p = 1$, the result is clearly false at every singular element.

3. Approximation theorems

The main part of this section consists of applying the results of Section 2 to prove unitary approximation theorems in the Schatten p-classes. For completeness we first prove a simple result in the operator norm.

THEOREM 3.1. *Let A be a positive operator on a Hilbert space $\mathscr{H}$. Then for every unitary operator U on $\mathscr{H}$, $\|I - A\| \leq \|U - A\| \leq \|I + A\|$.*

Proof. If $\|x\| = 1$, then

$$\begin{aligned}\|(U - A)x\|^2 &= \langle (I + A^2 - U^*A - AU)x, x\rangle \\ &\geq 1 + \|Ax\|^2 - 2\|Ax\| \\ &= (1 - \|Ax\|)^2.\end{aligned}$$

Since, for positive A,

$$\inf_{\|x\|=1} \|Ax\| = \inf_{\|x\|=1} \langle Ax, x\rangle, \quad \sup_{\|x\|=1} \|Ax\| = \sup_{\|x\|=1} \langle Ax, x\rangle,$$

we have

$$\begin{aligned}\|U - A\| &\geq \sup_{\|x\|=1} |1 - \|Ax\| \,| \\ &= \sup_{\|x\|=1} |1 - \langle Ax, x\rangle| \\ &= \sup_{\|x\|=1} |\langle (I - A)x, x\rangle| \\ &= \|I - A\|.\end{aligned}$$

This proves the left inequality.

Next,

$$\begin{aligned}\|U - A\| &= \sup_{\|x\|=1} \|Ux - Ax\| \\ &\leq \sup_{\|x\|=1} (1 + \|Ax\|) \\ &= \sup_{\|x\|=1} \langle (I + A)x, x\rangle \\ &= \|I + A\|. \quad \blacksquare\end{aligned}$$

Our results will show that for $p > 1$, the best unitary approximation to a positive operator A in the norm of $\mathscr{C}_p$ is I (providing any such approximation exists). The next result shows that for any $p > 0$, if $U - A \in \mathscr{C}_p$ for some unitary operator, then at least I is a candidate for an approximant of A; that is, $I - A \in \mathscr{C}_p$.

THEOREM 3.2. *If $\mathscr{I}$ is any two-sided ideal of $\mathscr{B}(\mathscr{H})$ and $U - A \in \mathscr{I}$ for some unitary U and positive A, then $I - A \in \mathscr{I}$. In particular $U - A \in \mathscr{C}_p$ implies that $I - A \in \mathscr{C}_p$ $(0 < p \leq \infty)$.*

Proof. Since every two-sided ideal of $\mathscr{B}(\mathscr{H})$ is selfadjoint, we have $U^* - A \in \mathscr{I}$ and so $-AU^* + UA = (U - A)U^* - U(U^* - A) \in \mathscr{I}$. Hence

$$I - A^2 = (U - A)(U^* + A) + AU^* - UA \in \mathscr{I}.$$

As A is positive, $I + A$ is invertible and so

$$I - A = (I - A^2)(I + A)^{-1} \in \mathscr{I}. \quad \blacksquare$$

Note that in the case when $\mathscr{I} = \mathscr{C}_\infty$ (the set of all compact operators) there is a one line proof of the above result depending on the fact that $\mathscr{B}(\mathscr{H})/\mathscr{C}_\infty$ is a C^*-algebra: for then the image of U under the canonical map $\mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})/\mathscr{C}_\infty$ is positive and unitary and so it is the identity in $\mathscr{B}(\mathscr{H})/\mathscr{C}_\infty$.

The following lemma is the key result for the approximation theorem.

LEMMA 3.3. *Let A be a strictly positive operator and let*

$$\mathscr{U}_A = \{U : U \text{ unitary}, U - A \in \mathscr{C}_p\}.$$

If $\mathscr{U}_A$ is non-empty, let $F_p(U) = \|U - A\|_p^p$, $U \in \mathscr{U}_A$. Then for V to be a local maximum or a local minimum of F_p $(p > 1)$, it is necessary that V be a symmetry commuting with A.

If $0 < p \leq 1$ the same result holds in a finite-dimensional Hilbert space provided that $V - A$ is invertible.

Proof. For any unit vector z and any real θ let $W_z(\theta)$ be the unitary operator defined by $W_z(\theta)x = e^{i\theta}\langle x, z\rangle z + x - \langle x, z\rangle z$, (that is, $W_z(\theta)$ multiplies the z-component of any vector by $e^{i\theta}$ and acts like the identity on the orthogonal complement of z). Since for $p > 1$, the derivative of F_p exists everywhere, if F_p has a local extremum at V, it is clearly necessary that for each z, $dF_p(VW_z(\theta))/d\theta$ vanishes at $\theta = 0$.

Let $V - A = U|V - A|$ be the polar decomposition of $V - A$. Then an application of the chain rule to the composition of the maps $\theta \mapsto VW_z(\theta) \to F_p(VW_z(\theta))$ and the result of Theorem 2.1 show that

$$\frac{d}{d\theta} F_p[VW_z(\theta)]\bigg|_{\theta=0} = p\mathscr{R}\tau[|V - A|^{p-1}U^*Vi(z \otimes z)] = 0.$$

Evaluating the trace using an orthonormal basis containing z, we find that $\langle |U - A|^{p-1}U^*Vz, z\rangle$ is real. Since this holds for any z, it follows that $|V - A|^{p-1}U^*V$ is selfadjoint.

Observe that V^*U is a partial isometry and

$$\ker (V^*U) = \ker U = \ker |V - A| = \ker |V - A|^{p-1}.$$

Hence $V^*U|V - A|^{p-1}$ exhibits the unique polar decomposition of a selfadjoint operator. It follows from Theorem 1.1 that V^*U is selfadjoint and commutes with $|V - A|^{p-1}$. Therefore V^*U commutes with every power of $|V - A|^{p-1}$, in particular with $|V - A|$. Thus

$$V^*(V - A) = V^*U|V - A| = |V - A|V^*U = |V - A|U^*V = (V^* - A)V$$

and so $V^*A = AV$ showing that V^*A is selfadjoint. Since A is strictly positive, $0 = \ker A = \ker V^*$ and it is easy to see that V is a symmetry commuting with A.

Note that the condition $p > 1$ was used in the above proof only to guarantee the existence of the derivative. Hence, if $V - A$ is an invertible operator on a finite-dimensional Hilbert space, this fact follows from Theorem 2.1 and the same proof establishes the statement covering the cases $0 < p \leq 1$. ∎

COROLLARY 3.4. *Let A be a non-negative operator on a Hilbert space and let $\mathscr{U}_A$ ($\neq \emptyset$) and F_p be defined as in the lemma. Then for V to be a local maximum or a local minimum of F_p ($p > 1$) it is necessary that* $\ker (A)$ *be a reducing subspace of V and that V restricted to* $\ker (A)^\perp$ *be a symmetry commuting with A.*

If $0 < p \leq 1$ the same result holds in a finite-dimensional Hilbert space provided that $V - A$ is invertible.

Proof. It follows from the proof of Lemma 3.3 that if V is a local maximum or a local minimum of F_p then $V^*A = AV$, since strict positivity was not used up to that point. Let E be the orthogonal projection onto $\ker (A)^\perp$. Then V^*EA is the unique polar decomposition of a selfadjoint operator. Thus V^*E is selfadjoint, that is, $V^*E = EV$. Multiplying on the left by V and on the right by V^* yields $EV^* = VE$. Thus

$$EVE = (V^*E)E = V^*E = EV \quad \text{and} \quad EVE = E(EV^*) = EV^* = VE.$$

Hence E commutes with V, so ker A reduces V and the equality $(VE)^2 = V(EVE) = VV^*E = E$ shows that V restricted to ker $(A)^\perp$ is a symmetry.

Some parts of the result below have been proved, using different methods, by Fan and Hoffman [7] and van Riemsdijk [14, Theorem 8].

THEOREM 3.5. *Let A be a strictly positive operator on a Hilbert space $\mathscr{H}$ and let*

$$\mathscr{U}_A = \{U : U \text{ unitary}, U - A \in \mathscr{C}_p\}.$$

If $\mathscr{U}_A$ is non-empty then for $p > 1$ the function $F_p(U) = \|U - A\|_p^p$ has a unique local minimum which occurs at $U = I$ and which is also a global minimum. In particular, if $U \in \mathscr{U}_A$, $U \neq I$, then

$$\|I - A\|_p < \|U - A\|_p.$$

If $\mathscr{H}$ is finite-dimensional, F_p has a unique local maximum which occurs at $U = -I$ and which is a global maximum. In particular if U is unitary, $U \neq \pm I$, then

$$\|I - A\|_p < \|U - A\|_p < \|I + A\|_p.$$

Proof. We first show that if the global minimum of F_p is attained then it is attained at $U = I$ and at no other point. From Theorem 3.2, if $\mathscr{U}_A \neq \emptyset$ then $I \in \mathscr{U}_A$. Also, since a global minimum is a local minimum, from Lemma 3.3, it can only be attained at some symmetry V, commuting with A. But then $I - V$ and $I - A$ are commuting compact normal operators and so have a common orthonormal basis $\{x_i\}$ of eigenvectors. Let $\lambda_i = \langle Ax_i, x_i\rangle$, $\mu_i = \langle Vx_i, x_i\rangle$. Then $|\mu_i| = 1$ and

$$\|V - A\|_p^p = \sum_{i=1}^{\infty} |\mu_i - \lambda_i|^p \geq \sum_{i=1}^{\infty} |1 - \lambda_i|^p = \|I - A\|_p^p$$

with equality holding only in case $\mu_i = 1$ for all i; that is, only if $V = I$. Similarly if the global maximum of F_p is attained then it is attained at $U = -I$ and at no other point. (Clearly this can only happen if $\mathscr{H}$ is finite-dimensional since the conditions $I + A \in \mathscr{C}_p$ and $A \geq 0$ are incompatible in infinite dimensions).

In the case when $\mathscr{H}$ is finite-dimensional the unitary operators form a compact set and clearly F_p attains its extreme values. Thus from above it follows that for U unitary, $U \neq \pm I$,

$$\|I + A\|_p > \|U - A\|_p > \|I - A\|_p.$$

Suppose now that $\mathscr{H}$ is infinite-dimensional and let $U \in \mathscr{U}_A$, $U \neq I$. Then $U - I$ is compact and normal and so has an orthonormal basis $\{x_i\}$ of eigenvectors. Let $H_n = \text{span}\,\{x_1, x_2, \ldots, x_n\}$ and let E_n be the orthogonal projection onto H_n. Then, since $E_n U E_n$ is a unitary operator on H_n, the finite-dimensional result shows that

$$\|E_n U E_n - E_n A E_n\|_p > \|E_n - E_n A E_n\|_p.$$

Since (E_n) converges strongly to I, it follows from [6, Lemma 2] that the sequence $(E_n X E_n)$ converges to X in $\mathscr{C}_p$. Hence $\|U - A\|_p \geq \|I - A\|_p$. This shows that the global minimum is attained at I and the first part of the proof shows that $\|U - A\|_p > \|I - A\|_p$.

It remains to show that F_p has no other local extremum and for this, in view of Lemma 3.3, it is sufficient to check every symmetry $V \neq \pm I$ which commutes with A. Then $V = E - F$ where E and F are projections commuting with A such that $E + F = I$ and $EF = 0$. Now if $U \neq V$ is any unitary operator which acts like the identity on the range of E and commutes with E then it is easy to see that

$$F_p(U) - F_p(V) = \|FUF - FAF\|_p^p - \| - F - FAF\|_p^p.$$

Thus the global result, applied to the range of F, shows that $F_p(U) - F_p(V) < 0$ and so V is not a local maximum. Similarly, V is not a local minimum. This completes the proof of the theorem. ▮

The following corollary applies also to the case $p = 1$. Note however that the strict inequalities do not hold in this case even when A is strictly positive (see Example 3.8).

Corollary 3.6. *If A is a non-negative operator and $U - A \in \mathscr{C}_p$ for some $p \geq 1$ then*

$$\|U - A\|_p \geq \|I - A\|_p.$$

If $\mathscr{H}$ is finite-dimensional then, in addition,

$$\|I + A\|_p \geq \|U - A\|_p.$$

Proof. For $p > 1$ this is proved in the same way as the theorem using Corollary 3.4 in place of Lemma 3.3. The weaker hypothesis merely results in the loss of uniqueness.

If $p = 1$ and $\mathscr{H}$ is finite-dimensional then the results follow from the fact that for fixed X, $\|X\|_p$ is a continuous function of p. The extension to the infinite-dimensional case is then proved using a sequence of finite dimensional projections as in the theorem. ▮

We now give some results that apply to the case $0 < p < 1$.

Theorem 3.7. *Let A be a strictly positive operator,*

$$\mathscr{U}_A = \{U : U \text{ unitary}, U - A \in \mathscr{C}_p\}$$

and, if $\mathscr{U}_A \neq \emptyset$ *let* $F_p(U) = \|U - A\|_p^p$ $(p > 0)$. *Then:*

(i) $F_p(U) \geq F_p(I)$ *for all unitary operators* U *commuting with* A.

(ii) *If* $A > I$ *or if* $A < I$, (*that is, if the unit circle does not separate the spectrum of* A) *then, for all* $U \in \mathscr{U}_A$, $F_p(U) \geq F_p(I)$, *and if* $\mathscr{H}$ *is finite-dimensional then* $F_p(U) \leq F_p(-I)$.

Proof. We merely sketch the proof since it is along the lines of previous results. If U commutes with A then U and A have a common basis of eigenvectors and so (i) follows as in Theorem 3.5. To prove (ii), note that if $A > I$ or if $A < I$ then $V - A$ is injective for every unitary operator V. Hence, if $\mathscr{H}$ is finite dimensional $V - A$ is always invertible. Thus, for finite dimensions, (ii) is proved just as Theorem 3.5 using the last part of Lemma 3.3. The result is extended to infinite dimensions using the same technique as before; note that although Lemma 2 of [6] is stated for normed ideals, the proof obviously holds for the metric ideals $\mathscr{C}_p$ $(0 < p < 1)$. ∎

We conclude with an example which shows that in the case $0 < p < 1$, the identity operator is not always a minimum of $F_p(U)$. In these cases the minimum is presumably attained at points V such that $V - A$ is not injective.

Example 3.8. Let

$$A = \begin{pmatrix} 1+\alpha & 0 \\ 0 & 1-\alpha \end{pmatrix},$$

where $0 < \alpha < 1$ and let

$$V_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Then an elementary calculation shows that the eigenvalues $s_1(\theta) \geq s_2(\theta) \geq 0$ of $|V_\theta - A|$ satisfy

$$s_1^2 + s_2^2 = 2(2 + \alpha^2 - 2\cos\theta), \qquad s_1^2 s_2^2 = (\alpha^2 + 2\cos\theta - 2)^2.$$

Thus, for sufficiently small θ we have $s_1 s_2 = \alpha^2 + 2\cos\theta - 2$ and so $s_1 + s_2 = 2\alpha$. This shows that I is not a unique global minimum of $F_1(U)$; indeed $F_1(V_\theta) = F_1(I)$ for all θ in some neighborhood of 0, (cf. Corollary 3.6). Also, elementary calculus shows that, for $0 < p < 1$, $F_p(V_\theta)$ has, as a function of θ, a strict maximum at $\theta = 0$.

References

1. J. G. Aiken, H. B. Jonassen and H. S. Aldrich, *Löwdin orthogonalization as a minimum energy perturbation*, J. Chem. Phys., vol. 62 (1975), pp. 2745–2746.
2. B. C. Carlson and J. M. Keller, *Orthogonalization procedures and localization of Wannier functions*, Phys. Rev., vol. 105 (1957), pp. 102–103.
3. A. J. Coleman, *Structure of fermion density matrices*, Rev. Mod. Phys., vol. 35 (1963), pp. 668–689.

4. J. DIESTEL, *Geometry of Banach spaces*, Lecture Notes in Mathematics No. 485, Springer-Verlag, Berlin, 1975.
5. N. DUNFORD and J. T. SCHWARTZ, *Linear operators, Part II*, Interscience, New York, 1963.
6. J. A. ERDOS, *On the trace of a trace class operator*, Bull. London Math. Soc., vol. 6 (1974), pp. 47–50.
7. K. FAN and A. J. HOFMANN, *Some metric inequalities in the space of matrices*, Proc. Amer. Math. Soc., vol. 6 (1955), pp. 111–116.
8. I. C. GOHBERG and M. G. KREIN, *Introduction to the theory of linear nonselfadjoint operators*, Transl. Math Monographs, vol. 18, Amer. Math. Soc., Providence, R. I., 1969.
9. P. R. HALMOS, *A Hilbert space problem book*, Van Nostrand, Princeton, 1967.
10. ———, *Positive approximants of operators*, Indiana Univ. Math. J., vol. 21 (1972), pp. 951–960.
11. C. A. MCCARTHY, $\mathscr{C}_p$, Israel J. Math., vol. 5 (1967), pp. 249–271.
12. G. W. PRATT JR. and S. F. NEUSTADTER, *Maximal orthogonal orbitals*, Phys. Rev., vol. 101 (1956), pp. 1248–1250.
13. H. RADJAVI and P. ROSENTHAL, *Invariant subspaces*, Springer-Verlag, Berlin, 1973.
14. D. J. VAN RIEMSDIJK, *Some metric inequalities in the space of bounded linear operators on a separable Hilbert space*, Nieuw Archief voor Wiskunde, vol. 20 (1972), pp. 216–230.

TULANE UNIVERSITY
NEW ORLEANS, LOUISIANA
TANO CORPORATION
METAIRE, LOUISIANA
KING'S COLLEGE
LONDON
TULANE UNIVERSITY
NEW ORLEANS, LOUISIANA