Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 400 (2005) 319–325
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Optimal commuting approximation of Hermitian operators

Francesco Ticozzi

Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy

Received 22 September 2004; accepted 28 November 2004
Available online 22 January 2005
Submitted by P.A. Fuhrmann

Abstract

We formulate and provide a solution to an approximation problem that occurs in various settings: Finding an optimal additive decomposition of a given Hermitian Hilbert–Schmidt operator, in a term commuting with a second Hermitian compact operator and a term as small as possible in the trace norm sense. In the finite-dimensional case, we show how to interpret our result through a Sylvester equation. An application to a quantum information problem and an interpretation in quantum probability are also sketched.

Keywords: Approximation problems; Sylvester equation; Quantum information

1. The problem

Let $\mathscr{H}$ be a separable complex Hilbert space. The set of the Hilbert–Schmidt operators $\mathscr{S}_2(\mathscr{H})$ is the set of operators A in $\mathscr{H}$ such that $\mathrm{Tr}(A^*A) < \infty$, where $*$ denotes adjoint. $\mathscr{S}_2(\mathscr{H})$ is an Hilbert space endowed with the inner product $(A, B) := \mathrm{Tr}(A^*B)$.

E-mail address: ticozzi@dei.unipd.it

0024-3795/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2004.11.028

An operator B in $\mathscr{H}$ is compact if

$$\exists \{C_k\}_{k=1}^{\infty}, \quad \operatorname{rank} C_k \leqslant k,$$

such that $\|B - C_k\| \searrow 0$ when $k \to \infty$, where $\|A\| := \sup_{x \neq 0} \frac{|Ax|}{|x|}$ indicates the operator norm.

Problem 1 (*Commuting decomposition*). Let A be a self-adjoint operators in $\mathscr{S}_2$ and B a compact Hermitian operator on $\mathscr{H}$. We are looking for an optimal decomposition for $A = A_0 + \Delta$, with $A_0 = A_0^* \in \mathscr{S}_2(\mathscr{H})$, such that $[A_0, B] := A_0 B - B A_0 = 0$, and the trace norm of Δ is minimal. That is, we want to find A_0 such that:

$$A_0 = \arg\min_X \|A - X\|,$$
$$\text{subject to:} \quad [X, B] = 0, \quad X = X^*. \tag{1}$$

2. The optimal decomposition

To solve our problem we will use the following theorems.

Theorem 1 (Hilbert–Schmidt). *Let A be a self-adjoint, compact operator on $\mathscr{H}$. Then, there is a complete orthonormal basis, $\{\phi_n\}$, for $\mathscr{H}$ so that $A\phi_n = \lambda_n \phi_n$ and* $\lim_{n \to \infty} \lambda_n = 0$.

For a proof, see e.g. [8, p. 203]. Every operator in $\mathscr{S}_2(\mathscr{H})$ is compact, hence the theorem applies to our setting, provided we consider self-adjoint operators.

Theorem 2. *Two self-adjoint, compact operators A and B on $\mathscr{H}$ commute ($AB - BA = 0$) if and only if they admit a common orthonormal basis of eigenvectors* $\left\{\hat{\phi}_n\right\}$.

Proof. It is easy to see that two operators with a common basis of eigenvectors commute. Here we will prove the converse. Suppose $AB = BA$ and take $\{\phi_n\}$ a basis of eigenvectors of A, corresponding to eigenvalues λ_n of A. The existence of such a basis is guaranteed by Theorem 1. Thus $AB\phi_n = \lambda_n B\phi_n$. Then $B\phi_n$ is an eigenvector of A relative to the same eigenvalue λ_n (or the zero vector). Suppose that $\phi_n^1, \ldots, \phi_n^n, \ldots$ are the orthonormal basis vectors that generate the eigenspace of A relative to λ_n, and define $\Lambda_n = \operatorname{span}\left\{\phi_n^1, \ldots, \phi_n^n, \ldots\right\}$. Hence $B\Lambda_n \subseteq \Lambda_n$. Define $\widehat{B}$ the restriction of B to Λ_n. $\widehat{B}$ remains a compact, self-adjoint operator on a (reduced) Hilbert space. Then, there exists $\left\{\hat{\phi}_n^i\right\}$, an orthonormal basis of eigenvectors of $\widehat{B}$ for Λ_n, that is also an orthonormal basis for Λ_n of eigenvectors of A. By substituting these $\left\{\hat{\phi}_n^i\right\}$ to the former $\left\{\phi_n^i\right\}$ for every eigenspace of A, we conclude. $\square$

The central result is the following:

Theorem 3 (Optimal decomposition). *Let us rewrite* $B = \sum_i b_i \Pi_i$, *where* b_i *are the eigenvalues of* B *and* Π_i *the orthogonal projectors on the* $\mathrm{Ker}(b_i I - B)$. *The optimal solution to the minimum problem* (1) *is given by* $\widehat{\Delta} := A - A_0$, *where*:

$$A_0 := \sum_i \Pi_i A \Pi_i, \tag{2}$$

is the sum of the reductions of A *on* B *eigenspaces.*

Proof. Reformulating the problem geometrically, we are looking for the orthogonal projection of A onto the closed linear subspace of operators commuting with B. By the theorem of the orthogonal projections [8], we know that it exists unique and corresponds to the minimum norm, commuting part.

A_0 is the desired operator if and only if:

[i] $[A_0, B] = 0$,
[ii] $(A - A_0, C) = 0, \quad \forall C \text{ s.t. } [C, B] = 0.$

In fact, for the candidate A_0 it holds that:

$$\begin{aligned} A_0 B - B A_0 &= \sum_i \Pi_i A \Pi_i \sum_j b_j \Pi_j - \sum_j b_j \Pi_j \sum_i \Pi_i A \Pi_i \\ &= \sum_i b_i \Pi_i A \Pi_i - \sum_i b_i \Pi_i A \Pi_i = 0, \end{aligned}$$

and

$$\begin{aligned} (A, C) - \left(\sum_i \Pi_i A \Pi_i, C \right) &= \mathrm{Tr}\left(A \sum_j c_j P_j \right) - \mathrm{Tr}\left(\sum_i \Pi_i A \Pi_i \sum_j c_j P_j \right) \\ &= \sum_n \left\{ \left\langle \phi_n, A \sum_j c_j P_j \phi_n \right\rangle - \left\langle \phi_n, \sum_j c_j P_j A P_j \phi_n \right\rangle \right\} \\ &= \sum_n \{ \langle \phi_n, c_n A \phi_n \rangle - \langle \phi_n, c_n P_n A P_n \phi_n \rangle \} \\ &= \sum_n \langle \phi_n, (c_n A - c_n A) \phi_n \rangle = 0, \end{aligned}$$

where we have used Theorem 2: If $[C, B] = 0$, then exists an orthonormal basis $\{\phi_n\}$ of common eigenvectors and we can write $C = \sum_n c_n P_n$, where P_n is the projector on the eigenspace generated by the nth common eigenvector. Thus, for every n there exists a unique $i(n)$ such that $\Pi_{i(n)} P_n = P_n \neq 0$.

The optimal solution A_0 defined in (2) is Hermitian:

$$A_0^* = \sum_i \Pi_i^* A^* \Pi_i^* = \sum_i \Pi_i A \Pi_i = A_0, \tag{3}$$

providing the desired solution for Problem 1. □

For many applications it is crucial to compute or estimate the norm of the "error" $\widehat{\Delta}$. Since $\widehat{\Delta}$ and A_0 are orthogonal in the Hilbert space of the Hilbert–Schmidt operators, it holds that:

$$\left\|\widehat{\Delta}\right\|^2 = \|A\|^2 - \|A_0\|^2. \tag{4}$$

3. Applications

3.1. Optimal solution for a class of Sylvester equations

The Hilbert–Schmidt operators space is a quite general, infinite dimensional setting in which is natural to formulate the problem geometrically. Nevertheless, some natural applications of Theorem 3 arise considering Hermitian or symmetric matrices. Here we present an an interpretation in terms of a class of Sylvester equations.

The condition $[A_0, B] = 0$ can be rewritten obtaining:

$$0 = [A_0, B] = [A, B] - [\Delta, B].$$

Defining $C := AB - BA$, we have:

$$[\Delta, B] = \Delta B - B\Delta = C.$$

This equation is a particular Sylvester equation [1, p. 203] that admits solutions (e.g. $\Delta = A$). Since $\sigma(B) \cap \sigma(B) = \sigma(B) \neq 0$ the equation has infinite solutions, and (1) is equivalent to find the minimum-norm (symmetric) solution of the above Sylvester equation. Thus, Problem 1 is equivalent to the following:

Problem 2. Let A and B be two Hermitian matices. We are looking for a self-adjoint Hilbert–Schmidt operator Δ that is the minimum trace-norm solution of the Sylvester equation:

$$\Delta B - B\Delta = C, \tag{5}$$

with $C := AB - BA$.

The optimal solution in the trace norm sense, is A_0, as defined in (2). Sylvester equations are crucial in different areas of filtering and control theory [4]. In particular, they play a fundamental role in spectral factorization problems and realization theory [3,5].

3.2. Application to quantum information theory

The choice of the trace norm is natural for the application fields we have in mind, as quantum mechanics and quantum information, where the trace is related to probabilities in statistical mixtures and measures [9,7].

In quantum mechanics the best knowledge we can have about a system is represented by a unit vector in a suitable separable Hilbert space, a pure state. Even if in the pure state description there is an amount of intrinsic uncertainty, we may need to add further uncertainty. It can be due to our ignorance about the state or about the transformations that have occurred or to a statistical description of a set of identical systems. The natural way to describe the system is then the density operator formalism. The state is described by a positive definite, unit trace Hilbert–Schmidt operator ρ on the system Hilbert space, called *density operator*. Here we consider finite dimensional Hilbert spaces for simplicity.

To apply our result to this setting, a relevant observation is the following:

Proposition 1. *Let ρ be a density matrix and $B = \sum_i b_i \Pi_i$ an Hermitian compact operator on $\mathscr{H}$. Then*

$$\hat{\rho} := \sum_i \Pi_i \rho \Pi_i$$

is a density operator too.

Proof. The key observation is that the optimal A_0, as defined in Theorem 3, has the same trace of A:

$$\mathrm{Tr}\left(A - \sum_i \Pi_i A \Pi_i\right) = \mathrm{Tr}\left(A - \left(\sum_i \Pi_i\right) A\right) = \mathrm{Tr}(A - A) = 0.$$

Hence, thanks to linearity of the trace, $\mathrm{Tr}(A) = \mathrm{Tr}(A_0)$. This means that if A is a density matrix, A_0 has unit trace and it is a density matrix too (positive definiteness is guaranteed by (2) and by the fact that A is positive definite). □

Suppose, on the one hand, we extract information from a finite-dimensional quantum system in a "noisy" environment, obtaining an estimated density operator ρ (e.g. by quantum state tomography [7]). On the other hand, we expected, by a theoretical analysis, ρ to be a statistical mixture of the states $\{\psi_i\}$, that are an orthonormal basis for the state space (e.g. the energy eigenstates of the system). We want to find the best approximation of ρ compatible with our "a priori" analysis. Theorem 3 tells us that the answer is:

$$\hat{\rho} = \sum_i \Pi_i \rho \Pi_i,$$

where every Π_i is the orthogonal projector onto the subspace generated by ψ_i. Hence, $\widehat{\Delta} = \rho - \hat{\rho}$ may be considered as the effect of the noise.

Such an experimental situation is very similar to the effect of a Von Neumann measurement in the case we do not know the outcome of the experiment. Consider an Hermitian operator B with spectral decomposition:

$$B = \sum_k b_k \Pi_k.$$

We consider the non-degenerate case, $b_k \neq b_l$ if $k \neq l$. B represents an observable for the quantum system and, according to the basic measurement postulates [2], measuring B on a system described by a density operator ρ induces the following state change:

$$\rho \to \hat{\rho} = \sum_i \mathrm{Tr}(\rho \Pi_i)\Pi_i. \tag{6}$$

Proposition 2. *Let ρ be a density operator and $B = \sum_i b_i \Pi_i$ an Hermitian operator on $\mathscr{H}$. Then we have*:

$$\sum_i \mathrm{Tr}(\rho \Pi_i)\Pi_i = \sum_i \Pi_i \rho \Pi_i.$$

Proof. Under the present hypothesis, we can rewrite

$$\Pi_i[\cdot] = \langle \psi_i, \cdot \rangle \psi_i.$$

Since $\langle \psi_i, \cdot \rangle$ is a complex number, using the linearity of the scalar product we obtain:

$$\begin{aligned}\sum_i \mathrm{Tr}(\rho \Pi_i)\Pi_i[\cdot] &= \sum_i \sum_j \langle \psi_j, \rho \langle \psi_i, \psi_j \rangle \psi_i \rangle \langle \psi_i, \cdot \rangle \psi_i \\ &= \sum_i \langle \psi_i, \rho \psi_i \rangle \langle \psi_i, \cdot \rangle \psi_i \\ &= \sum \langle \psi_i, \rho \langle \psi_i, \cdot \rangle \psi_i \rangle \psi_i \\ &= \sum_i \Pi_i \rho \Pi_i[\cdot]. \qquad \square\end{aligned}$$

Thus, a variational interpretation of the statistical description of these "quantum jumps" is provided: the density operator collapses to the best approximation in the trace norm sense of the before-measurement density operators that commutes with the observable involved.

This interpretation is mathematically equivalent to the fact that the best estimate for the density operator after the measurement is the prior estimate conditioned by the measurement (see e.g. [6], Appendix 2). In fact, in the non-commutative quantum probability setting the conditional distribution is the restriction (projection) of the former density operator to the measurement-involved subspaces, that represent the conditioning events.

References

[1] R. Bhatia, Matrix Analysis, Springer-Verlag, New York, 1997.

[2] C. Cohen-Tannoudji, B. Diu, F. Laloe, Quantum Mechanics, Wiley & Sons, 1977.

[3] A. Ferrante, M. Pavon, S. Pinzoni, On the relation between addittive and multiplicative decompositions of rational matrix functions, Int. J. Control 76 (2003) 366–385.

[4] H. Flanders, H.K. Wimmer, On the matrix equation $AX - XB = C$ and $AX - YB = C$, SIAM J. Appl. Math. 32 (1977) 707–710.

[5] P.A. Fuhrmann, On canonical Wiener–Hopf factorizations, in: G. Picci, D.S. Gilliam (Eds.), Dynamical Systems, Control, Coding, Computer Vision, Birkhauser, 1999, pp. 69–91.

[6] P.A. Meyer, Quantum Probability for Probabilists, Lecture Notes in Mathematics 1538, Springer Verlag, 1995.

[7] M.A. Nielsen, I.L. Chuang, Quantum Computation and Information, Cambridge University Press, 2002.

[8] M. Reed, B. Simon, Methods of Modern Mathematical Physics, vol. 1: Functional Analysis, Academic Press, 1980.

[9] J.J. Sakurai, Modern Quantum Mechanics, revised edition, Addison-Wesley, 1994.