Graduate Texts in Mathematics

Rajendra Bhatia

Matrix Analysis

矩阵分析

Springer

世界图书出版公司
www.wpcbj.com.cn

Rajendra Bhatia

Matrix Analysis

Springer

Rajendra Bhatia
Indian Statistical Institute
New Delhi 110 016
India

Mathematics Subject Classification (1991): 15-01, 15A16, 15A45, 47A55, 65F15

Library of Congress Cataloging-in-Publication Data
Bhatia, Rajendra, 1952–
Matrix analysis / Rajendra Bhatia.
p. cm. – (Graduate texts in mathematics ; 169)
Includes bibliographical references and index.
ISBN 0-387-94846-5 (alk. paper)
1. Matrices. I. Title. II. Series.
QA188.B485 1996
512.9'434–dc20 96-32217

ISBN 0-387-94846-5 Springer-Verlag New York Berlin Heidelberg SPIN 10524488

图书在版编目（CIP）数据

矩阵分析 = Matrix Analysis：英文/（印）巴蒂亚（Bhatia，R.）著．—影印本．—北京：世界图书出版公司北京公司，2011.3

ISBN 978-7-5100-3305-6

Ⅰ.①矩…　Ⅱ.①巴…　Ⅲ.①矩阵分析—研究生—教材—英文　Ⅳ.①O151.21

中国版本图书馆 CIP 数据核字（2011）第 029419 号

书　　名：	Matrix Analysis
作　　者：	Rajendra Bhatia
中 译 名：	矩阵分析
责任编辑：	高蓉　刘慧
出 版 者：	世界图书出版公司北京公司
印 刷 者：	三河市国英印务有限公司
发　　行：	世界图书出版公司北京公司（北京朝内大街 137 号　100010）
联系电话：	010-64021602，010-64015659
电子信箱：	kjb@ wpcbj. com. cn
开　　本：	24 开
印　　张：	15.5
版　　次：	2011 年 04 月
版权登记：	图字：01-2011-0664
书　　号：	978-7-5100-3305-6/O·882　　定　　价：45.00 元

Preface

A good part of matrix theory is functional analytic in spirit. This statement can be turned around. There are many problems in operator theory, where most of the complexities and subtleties are present in the finite-dimensional case. My purpose in writing this book is to present a systematic treatment of methods that are useful in the study of such problems.

This book is intended for use as a text for upper division and graduate courses. Courses based on parts of the material have been given by me at the Indian Statistical Institute and at the University of Toronto (in collaboration with Chandler Davis). The book should also be useful as a reference for research workers in linear algebra, operator theory, mathematical physics and numerical analysis.

A possible subtitle of this book could be *Matrix Inequalities*. A reader who works through the book should expect to become proficient in the art of deriving such inequalities. Other authors have compared this art to that of cutting diamonds. One first has to acquire hard tools and then learn how to use them delicately.

The reader is expected to be very thoroughly familiar with basic linear algebra. The standard texts *Finite-Dimensional Vector Spaces* by P.R. Halmos and *Linear Algebra* by K. Hoffman and R. Kunze provide adequate preparation for this. In addition, a basic knowledge of functional analysis, complex analysis and differential geometry is necessary. The usual first courses in these subjects cover all that is used in this book.

The book is divided, conceptually, into three parts. The first five chapters contain topics that are basic to much of the subject. (Of these, Chapter 5 is more advanced and also more special.) Chapters 6 to 8 are devoted to perturbation of spectra, a topic of much importance in numerical analysis, physics and engineering. The last two chapters contain inequalities and perturbation bounds for other matrix functions. These too have been of broad interest in several areas.

In Chapter 1, I have given a very brief and rapid review of some basic topics. The aim is not to provide a crash course but to remind the reader of some important ideas and theorems and to set up the notations that are used in the rest of the book. The emphasis, the viewpoint, and some proofs may be different from what the reader has seen earlier. Special attention is given to multilinear algebra; and inequalities for matrices and matrix functions are introduced rather early. After the first chapter, the exposition proceeds at a much more leisurely pace. The contents of each chapter have been summarised in its first paragraph.

The book can be used for a variety of graduate courses. Chapters 1 to 4 should be included in any course on Matrix Analysis. After this, if perturbation theory of spectra is to be emphasized, the instructor can go on to Chapters 6,7 and 8. With a judicious choice of topics from these chapters, she can design a one-semester course. For example, Chapters 7 and 8 are independent of each other, as are the different sections in Chapter 8. Alternately, a one-semester course could include much of Chapters 1 to 5, Chapter 9, and the first part of Chapter 10. All topics could be covered comfortably in a two-semester course. The book can also be used to supplement courses on operator theory, operator algebras and numerical linear algebra. The book has several exercises scattered in the text and a section called Problems at the end of each chapter. An *exercise* is placed at a particular spot with the idea that the reader should do it at that stage of his reading and then proceed further. *Problems*, on the other hand, are designed to serve different purposes. Some of them are supplementary exercises, while others are about themes that are related to the main development in the text. Some are quite easy while others are hard enough to be contents of research papers. From Chapter 6 onwards, I have also used the problems for another purpose. There are results, or proofs, which are a bit too special to be placed in the main text. At the same time they are interesting enough to merit the attention of anyone working, or planning to work, in this area. I have stated such results as parts of the Problems section, often with hints about their solutions. This should enhance the value of the book as a reference, and provide topics for a seminar course as well. The reader should not be discouraged if he finds some of these problems difficult. At a few places I have drawn attention to some unsolved research problems. At some others, the existence of such problems can be inferred from the text. I hope the book will encourage some readers to solve these problems too.

While most of the notations used are the standard ones, some need a little explanation:

Almost all functional analysis books written by mathematicians adopt the convention that an inner product $\langle u, v \rangle$ is linear in the variable u and conjugate-linear in the variable v. Physicists and numerical analysts adopt the opposite convention, and different notations as well. There would be no special reason to prefer one over the other, except that certain calculations and manipulations become much simpler in the latter notation. If u and v are column vectors, then u^*v is the product of a row vector and a column vector, hence a number. This is the inner product of u and v. Combined with the usual rules of matrix multiplication, this facilitates computations. For this reason, I have chosen the second convention about inner products, with the belief that the initial discomfort this causes some readers will be offset by the eventual advantages. (Dirac's bra and ket notation, used by physicists, is different typographically but has the same idea behind it.)

The k-fold tensor power of an operator is represented in this book as $\otimes^k A$, the antisymmetric and the symmetric tensor powers as $\wedge^k A$ and $\vee^k A$, respectively. This helps in thinking of these objects as maps, $A \to \otimes^k A$, etc. We often study the variational behaviour of, and perturbation bounds for, functions of operators. In such contexts, this notation is natural.

Very often we have to compare two n-tuples of numbers after rearranging them. For this I have used a pictorial notation that makes it easy to remember the order that has been chosen. If $x = (x_1, \ldots, x_n)$ is a vector with real coordinates, then $x^\downarrow$ and $x^\uparrow$ are vectors whose coordinates are obtained by rearranging the numbers x_j in decreasing order and in increasing order, respectively. We write $x^\downarrow = (x_1^\downarrow, \ldots, x_n^\downarrow)$ and $x^\uparrow = (x_1^\uparrow, \ldots, x_n^\uparrow)$, where $x_1^\downarrow \geq \cdots \geq x_n^\downarrow$ and $x_1^\uparrow \leq \cdots \leq x_n^\uparrow$.

The symbol $|||\cdot|||$ stands for a unitarily invariant norm on matrices: one that satisfies the equality $|||UAV||| = |||A|||$ for all A and for all unitary U, V. A statement like $|||A||| \leq |||B|||$ means that, for the matrices A and B, this inequality is true simultaneously for all unitarily invariant norms. The supremum norm of A, as an operator on the space $\mathbb{C}^n$, is always written as $||A||$. Other norms carry special subscripts. For example, the Frobenius norm, or the Hilbert-Schmidt norm, is written as $||A||_2$. (This should be noted by numerical analysts who often use the symbol $||A||_2$ for what we call $||A||$.)

A few symbols have different meanings in different contexts. The reader's attention is drawn to three such symbols. If x is a complex number, $|x|$ denotes the absolute value of x. If x is an n-vector with coordinates $(x_1, \ldots, x_n)$ then $|x|$ is the vector $(|x_1|, \ldots, |x_n|)$. For a matrix A, the symbol $|A|$ stands for the positive semidefinite matrix $(A^*A)^{1/2}$. If J is a finite set, $|J|$ denotes the number of elements of J. A permutation on n indices is often denoted by the symbol σ. In this case, $\sigma(j)$ is the image of the index j under the map σ. For a matrix $A, \sigma(A)$ represents the spectrum of A. The trace of a matrix A is written as tr A. In analogy, if $x = (x_1, \ldots, x_n)$ is a vector, we write tr x for the sum Σx_j.

The words matrix and operator are used interchangeably in the book. When a statement about an operator is purely finite-dimensional in content, I use the word matrix. If a statement is true also in infinite-dimensional spaces, possibly with a small modification, I use either the word matrix or the word operator. Many of the theorems in this book have extensions to infinite-dimensional spaces.

Several colleagues have contributed to this book, directly and indirectly. I am thankful to all of them. T. Ando, J.S. Aujla, R.B. Bapat, A. Ben Israel, I. Ionascu, A.K. Lal, R.-C.Li, S.K. Narayan, D. Petz and P. Rosenthal read parts of the manuscript and brought several errors to my attention. Fumio Hiai read the whole book with his characteristic meticulous attention and helped me eliminate many mistakes and obscurities. Long-time friends and coworkers M.D. Choi, L. Elsner, J.A.R. Holbrook, R. Horn, F. Kittaneh, A. McIntosh, K. Mukherjea, K.R. Parthasarathy, P. Rosenthal and K.B. Sinha, have generously shared with me their ideas and insights. These ideas, collected over the years, have influenced my writing.

I owe a special debt to T. Ando. I first learnt some of the topics presented here from his Hokkaido University lecture notes. I have also learnt much from discussions and correspondence with him. I have taken a lot from his notes while writing this book.

The idea of writing this book came from Chandler Davis in 1986. Various logistic difficulties forced us to abandon our original plans of writing it together. The book is certainly the poorer for it. Chandler, however, has contributed so much to my mathematics, to my life, and to this project, that this is as much his book as it is mine.

I am thankful to the Indian Statistical Institute, whose facilities have made it possible to write this book. I am also thankful to the Department of Mathematics of the University of Toronto and to NSERC Canada, for several visits that helped this project take shape.

It is a pleasure to thank V.P. Sharma for his LaTeXtyping, done with competence and with good cheer, and the staff at Springer-Verlag for their help and support.

My most valuable resource while writing, has been the unstinting and ungrudging support from my son Gautam and wife Irpinder. Without that, this project might have been postponed indefinitely.

Rajendra Bhatia

Contents

I

A Review of Linear Algebra

In this chapter we review, at a brisk pace, the basic concepts of linear and multilinear algebra. Most of the material will be familiar to a reader who has had a standard Linear Algebra course, so it is presented quickly with no proofs. Some topics, like tensor products, might be less familiar. These are treated here in somewhat greater detail. A few of the topics are quite advanced and their presentation is new.

I.1 Vector Spaces and Inner Product Spaces

Throughout this book we will consider finite-dimensional vector spaces over the field $\mathbb{C}$ of complex numbers. Such spaces will be denoted by symbols V, W, V_1, V_2, etc. Vectors will, most often, be represented by symbols u, v, w, x, etc., and scalars by a, b, s, t, etc. The symbol n, when not explained, will always mean the dimension of the vector space under consideration.

Most often, our vector space will be an inner product space. The inner product between the vectors u, v will be denoted by $\langle u, v\rangle$. We will adopt the convention that this is conjugate-linear in the first variable u and linear in the second variable v. We will always assume that the inner product is definite; i.e., $\langle u, u\rangle = 0$ if and only if $u = 0$. A vector space with such an inner product is then a finite-dimensional Hilbert space. Spaces of this type will be denoted by symbols $\mathcal{H}, \mathcal{K}$, etc. The norm arising from the inner product will be denoted by $\|u\|$; i.e., $\|u\| = \langle u, u\rangle^{1/2}$.

As usual, it will sometimes be convenient to deal with the standard Hilbert space $\mathbb{C}^n$. Elements of this vector space are column vectors with

n coordinates. In this case, the inner product $\langle u, v\rangle$ is the matrix product u^*v obtained by multiplying the column vector v on the left by the row vector u^*. The symbol $*$ denotes the conjugate transpose for matrices of any size. The notation u^*v for the inner product is sometimes convenient even when the Hilbert space is not $\mathbb{C}^n$.

The distinction between column vectors and row vectors is important in manipulations involving products. For example, if we write elements of $\mathbb{C}^n$ as column vectors, then u^*v is a number, but uv^* is an $n \times n$ matrix (sometimes called the "outer product" of u and v). However, it is typographically inconvenient to write column vectors. So, when the context does not demand this distinction, we may write a vector x with scalar coordinates $x_1, \ldots, x_n$, simply as $(x_1, \ldots, x_n)$. This will often be done in later chapters. For the present, however, we will maintain the distinction between row and column vectors.

Occasionally our Hilbert spaces will be real, but we will use the same notation for them as for the complex ones. Many of our results will be true for infinite-dimensional Hilbert spaces, with appropriate modifications at times. We will mention this only in passing.

Let $X = (x_1, \ldots, x_k)$ be a k-tuple of vectors. If these are column vectors, then X is an $n \times k$ matrix. This notation suggests matrix manipulations with X that are helpful even in the general case.

For example, let $X = (x_1, \ldots, x_k)$ be a linearly independent k-tuple. We say that a k-tuple $Y = (y_1, \ldots, y_k)$ is **biorthogonal** to X if $\langle y_i, x_j\rangle = \delta_{ij}$. This condition is expressed in matrix terms as $Y^*X = I_k$, the $k \times k$ identity matrix.

Exercise I.1.1 *Given any k-tuple of linearly independent vectors X as above, there exists a k-tuple Y biorthogonal to it. If $k = n$, this Y is unique.*

The Gram-Schmidt procedure, in this notation, can be interpreted as a matrix factoring theorem. Given an n-tuple $X = (x_1, \ldots, x_n)$ of linearly independent vectors the procedure gives another n-tuple $Q = (q_1, \ldots, q_n)$ whose entries are orthonormal vectors. For each $k = 1, 2, \ldots, n$, the vectors $\{x_1, \ldots, x_k\}$ and $\{q_1, \ldots, q_k\}$ have the same linear span. In matrix notation this can be expressed as an equation, $X = QR$, where R is an upper triangular matrix. The matrix R may be chosen so that all its diagonal entries are positive. With this restriction the factors Q and R are both unique. If the vectors x_j are not linearly independent, this procedure can be modified. If the vector x_k is linearly dependent on $x_1, \ldots, x_{k-1}$, set $q_k = 0$; otherwise proceed as in the Gram-Schmidt process. If the kth column of the matrix Q so constructed is zero, put the kth row of R to be zero. Now we have a factorisation $X = QR$, where R is upper triangular and Q has orthogonal columns, some of which are zero. Take the nonzero columns of Q and extend this set to an orthonormal basis. Then, replace the zero columns of Q by these additional basis vectors. The new matrix Q now has orthonormal columns, and we still have $X = QR$, because the new columns of Q are matched with zero rows of R. This is called the **QR decomposition.**

Similarly, a change of orthogonal bases can be conveniently expressed in these notations as follows. Let $X = (x_1, \ldots, x_k)$ be any k-tuple of vectors and $E = (e_1, \ldots, e_n)$ any orthonormal basis. Then, the columns of the matrix E^*X are the representations of the vectors comprising X, relative to the basis E. When $k = n$ and X is an orthonormal basis, then E^*X is a unitary matrix. Furthermore, this is the matrix by which we pass between coordinates of any vector relative to the basis E and those relative to the basis X. Indeed, if

$$u = a_1e_1 + \cdots + a_ne_n = b_1x_1 + \cdots + b_nx_n,$$

then we have

$$\begin{array}{lll} u = Ea, & a_j = e_j^*u, & a = E^*u, \\ u = Xb, & b_j = x_j^*u, & b = X^*u. \end{array}$$

Hence,

$$a = E^*Xb \quad \text{and} \quad b = X^*Ea.$$

Exercise I.1.2 *Let X be any basis of $\mathcal{H}$ and let Y be the basis biorthogonal to it. Using matrix multiplication, X gives a linear transformation from $\mathbb{C}^n$ to $\mathcal{H}$. The inverse of this is given by Y^*. In the special case when X is orthonormal (so that $Y = X$), this transformation is inner-product-preserving if the standard inner product is used on $\mathbb{C}^n$.*

Exercise I.1.3 *Use the QR decomposition to prove* **Hadamard's inequality**: *if $X = (x_1, \ldots, x_n)$, then*

$$|\det X| \le \prod_{j=1}^{n} \|x_j\|.$$

Equality holds here if and only if either the x_j are mutually orthogonal or some x_j is zero.

I.2 Linear Operators and Matrices

Let $\mathcal{L}(V, W)$ be the space of all linear operators from a vector space V to a vector space W. If bases for V, W are fixed, each such operator has a unique matrix associated with it. As usual, we will talk of operators and matrices interchangeably.

For operators between Hilbert spaces, the matrix representations are especially nice if the bases chosen are orthonormal. Let $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$, and let $E = (e_1, \ldots, e_n)$ be an orthonormal basis of $\mathcal{H}$ and $F = (f_1, \ldots, f_m)$ an orthonormal basis of $\mathcal{K}$. Then, the (i, j)-entry of the matrix of A relative to these bases is $a_{ij} = f_i^* Ae_j = \langle f_i, Ae_j \rangle$. This suggests that we may say that the matrix of A relative to these bases is F^*AE.

In this notation, composition of linear operators can be identified with matrix multiplication as follows. Let $\mathcal{M}$ be a third Hilbert space with orthonormal basis $G = (g_1, \ldots, g_p)$. Let $B \in \mathcal{L}(\mathcal{K}, \mathcal{M})$. Then

$$\begin{aligned}
(\text{matrix of } B \cdot A) &= G^*(B \cdot A)E \\
&= G^*BF\,F^*AE \\
&= (G^*BF)(F^*AE) \\
&= (\text{matrix of } B)\ (\text{matrix of } A).
\end{aligned}$$

The second step in the above chain is justified by Exercise I.1.2.

The **adjoint** of an operator $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is the unique operator A^* in $\mathcal{L}(\mathcal{K}, \mathcal{H})$ that satisfies the relation

$$\langle z, Ax \rangle_{\mathcal{K}} = \langle A^* z, x \rangle_{\mathcal{H}}$$

for all $x \in \mathcal{H}$ and $z \in \mathcal{K}$.

Exercise I.2.1 *For fixed bases in $\mathcal{H}$ and $\mathcal{K}$, the matrix of A^* is the conjugate transpose of the matrix of A.*

For the space $\mathcal{L}(\mathcal{H}, \mathcal{H})$ we use the more compact notation $\mathcal{L}(\mathcal{H})$. In the rest of this section, and elsewhere in the book, if no qualification is made, an operator would mean an element of $\mathcal{L}(\mathcal{H})$.

An operator A is called **self-adjoint** or **Hermitian** if $A = A^*$, **skew-Hermitian** if $A = -A^*$, **unitary** if $AA^* = I = A^*A$, and **normal** if $AA^* = A^*A$.

A Hermitian operator A is said to be **positive** or **positive semidefinite** if $\langle x, Ax \rangle \geq 0$ for all $x \in \mathcal{H}$. The notation $A \geq 0$ will be used to express the fact that A is a positive operator. If $\langle x, Ax \rangle > 0$ for all nonzero x, we will say A is **positive definite**, or **strictly positive** . We will then write $A > 0$. A positive operator is strictly positive if and only if it is invertible. If A and B are Hermitian, then we say $A \geq B$ if $A - B \geq 0$.

Given any operator A we can find an orthonormal basis $y_1, \ldots, y_n$ such that for each $k = 1, 2, \ldots, n$, the vector Ay_k is a linear combination of $y_1, \ldots, y_k$. This can be proved by induction on the dimension n of $\mathcal{H}$. Let λ_1 be any eigenvalue of A and y_1 an eigenvector corresponding to λ_1, and $\mathcal{M}$ the 1-dimensional subspace spanned by it. Let $\mathcal{N}$ be the orthogonal complement of $\mathcal{M}$. Let $P_{\mathcal{N}}$ denote the orthogonal projection on $\mathcal{N}$. For $y \in \mathcal{N}$, let $A_{\mathcal{N}}y = P_{\mathcal{N}}Ay$. Then, $A_{\mathcal{N}}$ is a linear operator on the $(n-1)$-dimensional space $\mathcal{N}$. So, by the induction hypothesis, there exists an orthogonal basis $y_2, \ldots, y_n$ of $\mathcal{N}$ such that for $k = 2, \ldots, n$ the vector $A_{\mathcal{N}}y_k$ is a linear combination of $y_2, \ldots, y_k$. Now $y_1, \ldots, y_n$ is an orthogonal basis for $\mathcal{H}$, and each Ay_k is a linear combination of $y_1, \ldots, y_k$ for $k = 1, 2, \ldots, n$. Thus, the matrix of A with respect to this basis is upper triangular. In other words, every matrix A is **unitarily equivalent** (or **unitarily similar**) to an upper triangular matrix T, i.e., $A = QTQ^*$, where Q is unitary and T is upper triangular. This triangular matrix is called a **Schur Triangular Form** for A. An orthonormal basis with respect to which A is upper triangular is called a **Schur basis** for A. If A is normal, then T is diagonal and we have $Q^*AQ = D$, where D is a diagonal matrix whose diagonal entries are the eigenvalues of A. This is the **Spectral Theorem** for normal matrices.

The Spectral Theorem makes it easy to define functions of normal matrices. If f is any complex function, and if D is a diagonal matrix with $\lambda_1, \ldots, \lambda_n$ on its diagonal, then $f(D)$ is the diagonal matrix with $f(\lambda_1), \ldots, f(\lambda_n)$ on its diagonal. If $A = QDQ^*$, then $f(A) = Qf(D)Q^*$. A special consequence, used very often, is the fact that every positive operator A has a unique positive square root. This square root will be written as $A^{1/2}$.

Exercise I.2.2 *Show that the following statements are equivalent:*

(i) A is positive.

(ii) $A = B^*B$ for some B.

(iii) $A = T^*T$ for some upper triangular T.

(iv) $A = T^*T$ for some upper triangular T with nonnegative diagonal entries.

If A is positive definite, then the factorisation in (iv) is unique. This is called the **Cholesky Decomposition** *of A.*

Exercise I.2.3 *(i) Let $\{A_\alpha\}$ be a family of mutually commuting operators. Then, there is a common Schur basis for $\{A_\alpha\}$. In other words, there exists a unitary Q such that $Q^*A_\alpha Q$ is upper triangular for all α.*

(ii) Let $\{A_\alpha\}$ be a family of mutually commuting normal operators. Then, there exists a unitary Q such that $Q^*A_\alpha Q$ is diagonal for all α.

For any operator A the operator A^*A is always positive, and its unique positive square root is denoted by $|A|$. The eigenvalues of $|A|$ counted with multiplicities are called the **singular values** of A. We will always enumerate these in decreasing order, and use for them the notation $s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A)$.

If rank $A = k$, then $s_k(A) > 0$, but $s_{k+1}(A) = \cdots = s_n(A) = 0$. Let S be the diagonal matrix with diagonal entries $s_1(A), \ldots, s_n(A)$ and S_+ the $k \times k$ diagonal matrix with diagonal entries $s_1(A), \ldots, s_k(A)$. Let $Q = (Q_1, Q_2)$ be the unitary matrix in which Q_1 is the $n \times k$ matrix whose columns are the eigenvectors of A^*A corresponding to the eigenvalues $s_1^2(A), \ldots, s_k^2(A)$ and Q_2 the $n \times (n-k)$ matrix whose columns are the eigenvectors of A^*A corresponding to the remaining eigenvalues. Then, by the Spectral Theorem

$$Q^*(A^*A)Q = \begin{pmatrix} S_+^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that

$$Q_1^*(A^*A)Q_1 = S_+^2, \quad Q_2^*(A^*A)Q_2 = 0.$$

The second of these relations implies that $AQ_2 = 0$. From the first one we can conclude that if $W_1 = AQ_1S_+^{-1}$, then $W_1^*W_1 = I_k$. Choose W_2 so that $W = (W_1, W_2)$ is unitary. Then, we have

$$W^*AQ = \begin{pmatrix} W_1^*AQ_1 & W_1^*AQ_2 \\ W_2^*AQ_1 & W_2^*AQ_2 \end{pmatrix} = \begin{pmatrix} S_+ & 0 \\ 0 & 0 \end{pmatrix}.$$

This is the **Singular Value Decomposition:** for every matrix A there exist unitaries W and Q such that

$$W^*AQ = S,$$

where S is the diagonal matrix whose diagonal entries are the singular values of A.

Note that in the above representation the columns of Q are eigenvectors of A^*A and the columns of W are eigenvectors of AA^* corresponding to the eigenvalues $s_j^2(A)$, $1 \le j \le n$. These eigenvectors are called the **right** and **left singular vectors** of A, respectively.

Exercise I.2.4 *(i) The Singular Value Decomposition leads to the* **Polar Decomposition:** *Every operator A can be written as $A = UP$, where U is unitary and P is positive. In this decomposition the positive part P is unique, $P = |A|$. The unitary part U is unique if A is invertible.*

(ii) An operator A is normal if and only if the factors U and P in the polar decomposition of A commute.

(iii) We have derived the Polar Decomposition from the Singular Value Decomposition. Show that it is possible to derive the latter from the former.

Every operator A can be decomposed as a sum

$$A = \operatorname{Re} A + i \operatorname{Im} A,$$

where $\operatorname{Re} A = \frac{A+A^*}{2}$ and $\operatorname{Im} A = \frac{A-A^*}{2i}$. This is called the **Cartesian Decomposition** of A into its "real" and "imaginary" parts. The operators $\operatorname{Re} A$ and $\operatorname{Im} A$ are both Hermitian.

The **norm** of an operator A is defined as

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

We also have

$$\|A\| = \sup_{\|x\|=\|y\|=1} |\langle y, Ax\rangle|.$$

When A is Hermitian we have

$$\|A\| = \sup_{\|x\|=1} |\langle x, Ax\rangle|.$$