Orthogonalisation of Vectors

Matrix Decompositions and Approximation Problems

Rajendra Bhatia

Rajendra Bhatia is a Professor at Indian Statistical Institute, New Delhi. His interests are in analysis of matrices and operators in Hilbert space.

The Gram–Schmidt Process

The Gram–Schmidt process is one of the first things one learns in a course on vectors or matrices. Let us recall it briefly.

Let $\boldsymbol{x} = (x_1, \quad , x_n)$ be a vector with n coordinates x_j, each of which is a complex number. The collection of all such vectors is the vector space $\mathbb{C}^n$ It helps to think of $\boldsymbol{x}$ as a column vector and write $\boldsymbol{x}^*$ for the row vector with coordinates $\bar{x}_j$. The *inner product* (or the *scalar product*) between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is the number

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^* \boldsymbol{y} = \sum_{j=1}^{n} \bar{x}_j y_j.$$

The *norm* of $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\| = (\boldsymbol{x}^* \boldsymbol{x})^{\frac{1}{2}} = (\sum_{j=1}^{n} |x_j|^2)^{\frac{1}{2}}.$$

If we are given n linearly independent vectors $\boldsymbol{a}_1, \quad \boldsymbol{a}_n$, the Gram–Schmidt process constructs an orthonormal basis out of them as follows. We put $\boldsymbol{q}_1 = \boldsymbol{a}_1 / \|\boldsymbol{a}_1\|$. This vector has norm 1. We now put $\boldsymbol{v}_2 = \boldsymbol{a}_2 - \langle \boldsymbol{q}_1, \boldsymbol{a}_2 \rangle \boldsymbol{q}_1$; and $\boldsymbol{q}_2 = \boldsymbol{v}_2 / \|\boldsymbol{v}_2\|$. Then $\boldsymbol{q}_2$ is orthogonal to $\boldsymbol{q}_1$ and has norm 1. At the next stage, we put $\boldsymbol{v}_3 = \boldsymbol{a}_3 - \langle \boldsymbol{q}_1, \boldsymbol{a}_3 \rangle \boldsymbol{q}_1 - \langle \boldsymbol{q}_2, \boldsymbol{a}_3 \rangle \boldsymbol{q}_2$; and $\boldsymbol{q}_3 = \boldsymbol{v}_3 / \|\boldsymbol{v}_3\|$. Continuing this way we obtain an orthonormal basis $\boldsymbol{q}_1, \quad , \boldsymbol{q}_n$. Note that for each $1 \leq k \leq n$, the linear spans of $\boldsymbol{a}_1, \quad , \boldsymbol{a}_k$ and $\boldsymbol{q}_1, \quad , \boldsymbol{q}_k$ are equal.

How close are the vectors $\{\boldsymbol{q}_j\}$ to the original vectors $\{\boldsymbol{a}_j\}$? To make this precise let us define the distance

between two ordered sets $\{\boldsymbol{x}_1, \quad , \boldsymbol{x}_k\}$ and $\{\boldsymbol{y}_1, \quad , \boldsymbol{y}_k\}$ of vectors in $\mathbb{C}^n$ as

$$\left(\sum_{j=1}^{k} \|\boldsymbol{x}_j - \boldsymbol{y}_j\|^2\right)^{\frac{1}{2}} \tag{1}$$

Note that each $\boldsymbol{x}_j$ is an n-vector. If we write it as $\boldsymbol{x}_j = (\boldsymbol{x}_{j1}, \quad , \boldsymbol{x}_{jn})$, then the quantity in (1) is

$$\left(\sum_{j=1}^{k}\sum_{r=1}^{n} |x_{jr} - y_{jr}|^2\right)^{\frac{1}{2}} \tag{2}$$

Let us consider a very simple example in the space $\mathbb{C}^2$ Let $\boldsymbol{a}_1 = (1,0), \boldsymbol{a}_2 = (\frac{4}{5}, \frac{3}{5})$. The vectors $\boldsymbol{a}_1, \boldsymbol{a}_2$ are linearly independent and each of them has norm 1. However, they are not orthogonal to each other. The Gram–Schmidt process applied to them gives the vectors $\boldsymbol{q}_1 = (1,0), \boldsymbol{q}_2 = (0,1)$. The distance between the pair $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ and the pair $\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$ is $(\frac{4}{5})^{\frac{1}{2}}$. Can we find another pair of orthonormal vectors that is closer to $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$? If we try the obvious possibilities that the forms of $\boldsymbol{a}_1, \boldsymbol{a}_2$ suggest, we soon find that the pair $\boldsymbol{y}_1 = (\frac{4}{5}, -\frac{3}{5}), \boldsymbol{y}_2 = (\frac{3}{5}, \frac{4}{5})$ is at distance $(\frac{12}{25})^{\frac{1}{2}}$ from $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$. Thus the Gram–Schmidt process while constructing an orthonormal basis can take us far away from the original set of vectors.

Another pair that is even closer to $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ is the pair $\boldsymbol{u}_1 = (\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$, $\boldsymbol{u}_2 = (\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})$. One can see that the distance of this pair from $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ is $(4 - \frac{8}{\sqrt{5}})^{\frac{1}{2}}$. Thus the three pairs $\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$, $\{\boldsymbol{y}_1, \boldsymbol{y}_2\}$ and $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ are at distances .8944, .6928 and .6498, respectively from the given pair $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$.

One can see, using Lagrange multipliers, that among all pairs of orthonormal vectors, the pair $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is the closest to $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$. We will soon see this by another argument.

The problem of finding the orthonormal basis closest to a given set of linearly independent vectors is of interest

in quantum chemistry. In many models of atomic phenomena some of the quantities of interest are represented by orthonormal vectors. Experimental observations to measure these quantities are inaccurate and thus give us vectors that are not orthonormal. We might want to stay as close to the experimental data as possible when converting these vectors to orthonormal ones demanded by the model. The process of finding the closest orthonormal basis is called the *Löwdin orthogonalisation* after the Swedish chemist P O Löwdin who introduced it. This is related to one of the basic theorems in linear algebra as we will see.

Matrix Approximation Problems

Let A be an $n \times n$ matrix with entries a_{ij}. Let A^* be the conjugate transpose of A—the matrix whose i,j entry is $\bar{a}_{ji}$. Let tr A stand for the trace of A. The *Hilbert–Schmidt norm* (or the *Frobenius norm*) of A is defined as

$$\|A\|_2 = (\sum_{i,j} |a_{ij}|^2)^{1/2} = (\text{tr } A^*A)^{1/2} \qquad (3)$$

This norm is *unitarily invariant*: if U, V are unitary matrices, then

$$\|A\|_2 = \|UAV\|_2. \qquad (4)$$

This is so because

$$\text{tr } (UAV)^*(UAV) = \text{tr } V^*A^*AV = \text{tr } A^*A. \qquad (5)$$

Note that if $\{\boldsymbol{a}_1, \quad \boldsymbol{a}_n\}$ are elements of $\mathbb{C}^n$ and if we write the $n \times n$ matrix A whose columns are $\boldsymbol{a}_1, \quad , \boldsymbol{a}_n$ as $A = [\boldsymbol{a}_1, \quad , \boldsymbol{a}_n]$, then

$$\|A\|_2^2 = \sum_j \|\boldsymbol{a}_j\|^2$$

The matrix A is invertible if and only if its columns are linearly independent as vectors, and it is unitary if and only if they are orthonormal. Thus the problem of

finding the orthonormal basis closest to a given set of n linearly independent vectors is the same as the problem of finding the unitary matrix closest to a given invertible matrix. Here the closest matrix is one whose distance in the Hilbert–Schmidt norm from the given matrix is minimal.

This is a typical example of a matrix approximation problem.

The QR and the Polar Decompositions

The Gram–Schmidt process can be represented as an interesting matrix factoring theorem:

Every invertible matrix A can be factored as $A = QR$, where Q is unitary and R is upper triangular. We can choose R so that all its diagonal entries are positive. With this restriction Q and R are unique.

It is not difficult to see how this theorem follows from the Gram–Schmidt process. The columns of Q are orthonormal vectors constructed from the columns of A. The fact that $\{a_1, \quad , a_k\}$ span the same linear space as $\{q_1, \quad q_k\}$ is reflected in the upper triangular form of R. The vectors Q are unique upto a multiplication by a complex number of modulus one. So, the restriction that the diagonal entries of R be positive imposes uniqueness.

The decomposition $A = QR$ is called the QR *decomposition*. If A is singular, it still has a QR decomposition. Now some of the rows of R are zero.

There is another factoring of an invertible matrix into two factors one of which is unitary. This is the *polar decomposition*:

Every invertible matrix A can be factored uniquely as $A = UP$, where U is unitary and P is positive definite.

The factor P is the unique positive definite square root

of the positive definite matrix A^*A. If one puts $U = AP^{-1}$, then $U^*U = UU^* = I$. If A is singular, it still has a polar decomposition $A = UP$. Now the factor U is not unique, but P is.

The polar decomposition has an interesting extremal characterisation:

Theorem. Among all unitary matrices the one closest to A is the matrix U in the polar decomposition $A = UP$.

Proof. Let W be any unitary matrix. Then

$$\|A - W\|_2 = \|UP - W\|_2 = \|P - U^*W\|_2,$$

by the unitary invariance property (4). Thus to find the unitary matrix closest to A it suffices to find the one closest to P. If we show that the unitary matrix closest to P is the identity matrix I it will follow that the unitary matrix closest to UP is U

For every unitary matrix V

$$\|P-V\|_2^2 = \text{tr}\,(P-V^*)(P-V) = \text{tr}\,(P^2+I-PV-V^*P).$$

This quantity is minimum when

$$\text{tr}\,(PV + V^*P) = \text{tr}\,P(V + V^*) \qquad (6)$$

is maximum. The trace is not affected if we apply a unitary similarity (i.e., $\text{tr}\,X = \text{tr}\,WXW^*$ for all X and unitary W). The spectral theorem tells us that we can apply such a similarity to bring V to the diagonal form. Thus we may assume that V is diagonal with entries $e^{i\theta_j}, 1 \leq j \leq n$ down its diagonal. So, the quantity in (6) is

$$\text{tr}\,P(V + V^*) = 2\sum_j p_{jj}\cos\theta_j.$$

Since $p_{jj} \geq 0$, clearly this is maximised when $\cos\theta_j = 1$. This translates to the condition $V = I$. ■

Thus the polar decomposition provides the basis for the Löwdin orthogonalisation. The orthonormal basis closest to a set of linearly independent vectors $\{\boldsymbol{a}_1, \quad \boldsymbol{a}_n\}$ is obtained by writing the matrix $A = [\boldsymbol{a}_1, \quad \boldsymbol{a}_n]$, then finding its polar decomposition $A = UP$, and reading the columns of $U = [\boldsymbol{u}_1, \quad , \boldsymbol{u}_n]$ to get the desired orthonormal basis $\{\boldsymbol{u}_1, \quad , \boldsymbol{u}_n\}$.

This explains the example discussed in Section 1. We have the polar decomposition

$$\begin{bmatrix} 1 & \frac{4}{5} \\ 0 & \frac{3}{5} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$$

Since $P = WSW^*$ where W is unitary and S diagonal with positive entries, we can write $A = UP = UWSW^* = VSW^*$ where V is unitary. This is called the *singular value decomposition* of A. To find the factors here, we have to diagonalise P. This involves a more elaborate calculation than the one for the Gram–Schmidt process.

The Closest Hermitian Matrix

The problem of finding the closest Hermitian matrix to a given matrix is motivated by the same considerations as that of finding the closest unitary matrix. It is simpler to solve this.

If $A = B + iC$, where B and C are Hermitian, then

$$\|A\|_2^2 = \text{tr } A^*A = \text{tr } (B - iC)(B + iC) =$$

$$\text{tr } (B^2 + C^2) = \|B\|_2^2 + \|C\|_2^2.$$

Every matrix has a decomposition of this kind:

If we put $B = \frac{1}{2}(A + A^*)$ and $C = \frac{1}{2i}(A - A^*)$, then B, C are Hermitian and $A = B + iC$. This is analogous to the decomposition $z = x + iy$ of a complex number into its real and imaginary parts. For this reason B

and C are called the *real* and *imaginary parts* of A and the decomposition $A = B + iC$ is called the *Cartesian decomposition.*

Now, if H is any Hermitian matrix, then

$$\|A - H\|_2^2 = \|H - B\|_2^2 + \|C\|_2^2.$$

Clearly, the choice $H = B$ minimises this quantity. Thus the Hermitian matrix closest to A is the real part of A.

The polar decomposition $A = UP$ can be thought of as the analogue of the polar representation $z = e^{i\theta}r$ of a complex number. Thus the statements about the closest unitary and Hermitian matrices proved above are analogues of the facts about the point on the unit circle and the point on the real line closest to a given complex number.

A matrix is said to be *normal* if $AA^* = A^*A$. This is equivalent to the condition that the factors U and P in the polar decomposition of A commute. Evidently Hermitian matrices and unitary matrices are normal.

The set of all Hermitian matrices is a real vector space; the set of all unitary matrices is a differentiable manifold. The set of all normal matrices does not have any nice geometric structure. This is one reason why the problem of finding the closest normal matrix to a given matrix turns out to be much harder than the problems we have considered. This problem is not yet solved completely. See [2] for a discussion and also for examples of other problems where the solution for normal matrices is much harder than that for Hermitian or unitary matrices.

Approximation in Other Norms

The Hilbert–Schmidt norm is the simplest norm on matrices from the point of view of approximation problems. This is because it is like the Euclidean norm on vectors. There are other norms that are of interest. For exam-

Suggested Reading

[1] A detailed discussion of the polar and the QR deompositions may be found in H Helson, *Linear Algebra*, TRIM 4, Hindustan Book Agency, 1994.

[2] A more advanced treatment of matrix approximation problems may be found in R Bhatia, *Matrix Analysis*, Springer-Verlag, 1997.

[3] The relevance of matrix approximation problems to quantum chemistry is explained in the article by J A Goldstein and M Levy, Linear algebra and quantum chemistry, *American Math. Monthly*, 78, 710–718, 1991.

[4] The Löwdin Orthogonalisation was proposed by P O Löwdin, On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals, *J. Chem. Phys.*, 18, 365–374, 1950.

[5] Algorithms for finding the QR and the Singular Value Decompositions are discussed in G Golub and C Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1983.

Box 1.

Let A^* be the matrix obtained from A by taking the transpose of A and then replacing each entry by its complex conjugate. A matrix A which satisfies $A = A^*$ is called *Hermitian.* A Hermitian matrix all whose eigenvalues are positive is called *positive definite.* An invertible matrix A for which $A^{-1} = A^*$ is called *unitary.* A is called *normal* if $AA^* = A^*A$. Hermitian matrices and unitary matrices are special kinds of normal matrices.

The *spectral theorem* says that every normal matrix A can be *diagonalised* by a *unitary conjugation;* i.e., there exists a unitary matrix U and a diagonal matrix D such that $A = UDU^*$. The diagonal entries of D are complex numbers. They are real if A is Hermitian, positive if A is positive definite, and complex numbers of modulus one if A is unitary.

ple, if we think of A as a linear operator on $\mathbb{C}^n$ then the *operator norm* of A is defined as

$$\|A\| = \max\{\|Ax\| : x \in \mathbb{C}^n \ \|x\| = 1\}.$$

Like the Hilbert–Schmidt norm, this norm is also unitarily invariant. There are several other norms on matrices that are unitarily invariant.

The answer to a minimisation problem often changes with the norm. That is natural, because the functions being minimised are different.

It is, therefore, interesting to know that for *every* unitarily invariant norm $|||\ |||$ on the space of matrices, the minimum of $|||A-W|||$ over unitary matrices is attained when W is the unitary factor in the polar decomposition of A; and the minimum of $|||A-H|||$ over Hermitian matrices is attained when $H = \frac{1}{2}(A + A^*)$.

Address for Correspondence
Rajendra Bhatia
Indian Statistical Institute
7, SJS Sansanwal Marg
New Delhi 110 016, India.

It is never possible to introduce only observable quantities in a theory. It is the theory which decides what can be observed.

Albert Einstein