

Methods in Computational Science - Eigenvalue and singular values decompositions (ch.6)

Johan Hoffman

Complex vector spaces

Let $z \in C$ denote a complex number

$$z = a + ib,$$

with the real part $\text{Re}(z) = a$ and the imaginary part $\text{Im}(z) = b$, both real numbers, and the imaginary unit $i^2 = -1$. The complex conjugate of z takes the form

$$\bar{z} = a - ib,$$

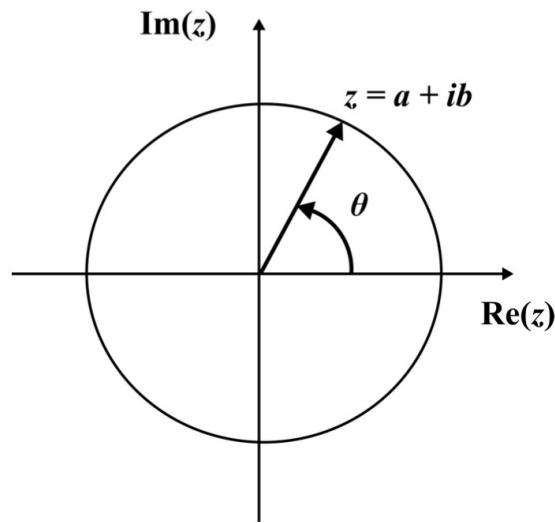
and the magnitude, or *modulus*, of z is defined as

$$|z| = \sqrt{\bar{z}z} = \sqrt{(a - ib)(a + ib)} = \sqrt{a^2 + b^2}.$$

The complex plane

$$z = r \exp(i\theta) = r(\cos(\theta) + i \sin(\theta))$$

with $r = |z|$, and $\theta = \text{Arg}(z)$ the counter-clockwise angle measured from the x_1 -axis in the interval $[0, 2\pi]$. The identity $\exp(ix) = \cos(x) + i \sin(x)$ is *Euler's formula*, which is valid also for $x \in C$.



The complex vector spaces $C^{m \times n}$ and C^n

The complex vector spaces $C^{m \times n}$ and C^n are defined by the vector space operations of component-wise addition and scalar multiplication of complex numbers, where the matrix and vector operations in $R^{m \times n}$ and R^n extend to complex vector spaces with natural modifications induced by the properties of the complex scalars. The analogue of the transpose of a column vector $x \in C^n$ is the *adjoint* x^* , defined as the row vector with components replaced by their complex conjugates, and the adjoint of a complex $m \times n$ matrix $A = (a_{ij})$ is the $n \times m$ matrix

$$A^* = (\bar{a}_{ji}).$$

Analogous to the transpose, the adjoint of a matrix product satisfies the following rule,

$$(AB)^* = B^* A^*,$$

for $A \in C^{m \times k}$ and $B \in C^{k \times n}$. If $A = A^*$ the matrix A is *Hermitian*, a generalization of symmetry for real matrices.

The complex vector spaces $C^{m \times n}$ and C^n

A matrix is *normal* if $AA^* = A^*A$, and clearly every Hermitian matrix is normal. The generalization of an orthogonal real matrix is a *unitary* matrix, a square complex matrix U which satisfies the property

$$U^*U = UU^* = I,$$

and consequently, that $U^{-1} = U^*$.

The vector space C^n is an inner product space with dimension $\dim(C^n) = n$, inner product

$$(x, y) = x^*y = \sum_{i=1}^n \bar{x}_i y_i, \quad x, y \in C^n,$$

and the associated norm

$$\|x\| = (x, x)^{1/2}, \quad x \in C^n.$$

The complex vector spaces $C^{m \times n}$ and C^n

Example 6.1. The complex vector space C^2 has dimension $\dim(C^2) = 2$. Hence, the vectors

$$b_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

constitute an orthonormal basis for C^2 , since

$$(b_1, b_2) = b_1^* b_2 = 0, \quad \|b_1\| = \|b_2\| = 1.$$

To express the complex vector $v = (1 + i, -i)^T$ in this basis we write

$$v = \begin{bmatrix} 1+i \\ -i \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

which gives that

$$\begin{cases} c_1 &= 1+i, \\ c_2 &= -i. \end{cases}$$

Eigenvalues and eigenvectors

An *eigenvector* of a square matrix $A \in C^{n \times n}$ is a nonzero vector $x \in C^n$, such that

$$Ax = \lambda x,$$

with $\lambda \in C$ the associated *eigenvalue*.

$$Ax = \lambda x \Leftrightarrow (A - \lambda I)x = 0,$$

$$\text{null}(A - \lambda I) \neq \{0\} \Leftrightarrow \det(A - \lambda I) = 0.$$

Spectrum and algebraic multiplicity

Since the determinant is an n degree polynomial in λ , by the *fundamental theorem of algebra* the *characteristic equation*

$$\det(A - \lambda I) = 0$$

has n complex roots $\lambda_j \in C$, each an eigenvalue of A , not necessarily unique, and

$$\det(A - \lambda I) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n).$$

Algebraic multiplicity is the multiplicity of each eigenvalue as a root of the characteristic equation. The set of all eigenvalues is the *spectrum*, and the maximal eigenvalue modulus the *spectral radius*.

$$\Lambda(A) = \{\lambda_j\}_{j=1}^n, \quad \rho(A) = \max_{\lambda_j \in \Lambda(A)} |\lambda_j|.$$

Eigenspace and geometric multiplicity

- The *eigenspace* E_λ is the subspace of C^n spanned by all the eigenvectors associated with the eigenvalue λ , together with the zero vector.
- *Geometric multiplicity* of λ is equal to the dimension of its eigenspace $\dim(E_\lambda)$, equal to the number of linearly independent eigenvectors associated with λ .
- The algebraic multiplicity of an eigenvalue λ is at least as great as its geometric multiplicity, and λ is said to be *simple* if its algebraic multiplicity is one.
- An eigenvalue is *defective* if its algebraic multiplicity exceeds its geometric multiplicity, and a matrix is defective if it has one or more defective eigenvalues.

Algebraic and geometric multiplicity

Example 6.3. Consider a matrix A and its characteristic equation,

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}, \quad \det(A - \lambda I) = (1 - \lambda)(2 - \lambda) = 0.$$

A has two distinct eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$, and by equation (6.2) each eigenvector $x = (x_1, x_2)$ must satisfy the system of equations $Ax = \lambda x$, that is,

$$\begin{aligned} x_1 + 2x_2 &= \lambda x_1, \\ 2x_2 &= \lambda x_2. \end{aligned}$$

Therefore, $x_2^{(1)} = 0$ for the eigenvector $x^{(1)}$ associated with $\lambda_1 = 1$, and $x_1^{(2)} = 2x_2^{(2)}$ for the eigenvector $x^{(2)}$ associated with $\lambda_2 = 2$. For example, we can choose the two eigenvectors

$$x^{(1)} = (1, 0)^T, \quad x^{(2)} = (2, 1)^T.$$

Both λ_1 and λ_2 are simple eigenvalues with geometric multiplicity equal to one.

Algebraic and geometric multiplicity

The matrix power A^k represents repeated multiplication by the same matrix k times, which by equation (6.2) has the same eigenvector x as A , but the new eigenvalue λ^k ,

$$A^k x = A^{k-1} Ax = A^{k-1} \lambda x = \dots = \lambda^k x.$$

If we add a constant $c \in C$ to the diagonal of A we get the matrix $A + cI$, which has the same eigenvector as A but the new eigenvalue $\lambda + c$,

$$(A + cI)x = Ax + cx = \lambda x + cx = (\lambda + c)x.$$

If A is a nonsingular matrix, then λ^{-1} is an eigenvalue of the inverse matrix A^{-1} , with the same eigenvectors, which follows from multiplication of equation (6.2) by A^{-1} from the left,

$$A^{-1}Ax = A^{-1}\lambda x \Leftrightarrow x = A^{-1}\lambda x \Leftrightarrow \lambda^{-1}x = A^{-1}x.$$

If $\lambda \in \Lambda(A)$ is a complex eigenvalue and A is a real matrix, then also $\bar{\lambda} \in \Lambda(A)$.

Similarity transformation

Let $X \in C^{n \times n}$ be a nonsingular matrix, then the map

$$A \mapsto X^{-1}AX$$

is a *similarity transformation* of A , and we say that two square matrices A and B are *similar* if

$$B = X^{-1}AX.$$

Since X is nonsingular and $\dim(X) = n$, it follows that its column vectors span C^n . Therefore, X corresponds to a change of basis transformation, from the standard basis to the basis of column vectors of X . Hence, B can be interpreted as the linear transformation A expressed in the basis of the column vectors of X .

Similarity transformation

Let $X \in C^{n \times n}$ be a nonsingular matrix, then the map

$$A \mapsto X^{-1}AX$$

is a *similarity transformation* of A , and we say that two square matrices A and B are *similar* if

$$B = X^{-1}AX.$$

Theorem 6.4. *With $X \in C^{n \times n}$ a nonsingular matrix, the matrix $B = X^{-1}AX$ is similar to the matrix A , and the two similar matrices have the same eigenvalues with the same algebraic and geometric multiplicity, and for each eigenvector x of A the vector $X^{-1}x$ is an eigenvector of B .*

Similarity transformation

Proof. If $Ax = \lambda x$ and $B = X^{-1}AX$, it follows that

$$BX^{-1}x = X^{-1}AXX^{-1}x = X^{-1}Ax = \lambda X^{-1}x,$$

which means that B has the same eigenvalues as A , but the transformed eigenvectors $X^{-1}x$. The geometric multiplicity of each eigenvalue stays the same since linear independence of the eigenvectors x implies linear independence of the eigenvectors $X^{-1}x$, which follows from

$$\alpha_1 X^{-1}x_1 + \dots + \alpha_k X^{-1}x_k = X^{-1}(\alpha_1 x_1 + \dots + \alpha_k x_k).$$

The algebraic multiplicity of the eigenvalues of A and B are the same since the characteristic polynomials are identical,

$$\begin{aligned}\det(\lambda I - B) &= \det(\lambda X^{-1}X - X^{-1}AX) = \det(X^{-1}(\lambda I - A)X) \\ &= \det(X^{-1}) \det(\lambda I - A) \det(X) = \det(\lambda I - A).\end{aligned}$$

Spectral theorems

Theorem 6.5 (Spectral theorem for nondefective matrices). *Each matrix $A \in C^{n \times n}$ which is nondefective has an eigenvalue decomposition*

$$A = X\Lambda X^{-1},$$

where $X \in C^{n \times n}$ is a nonsingular matrix with the eigenvectors of A as column vectors, and $\Lambda \in C^{n \times n}$ is a diagonal matrix with the eigenvalues of A on the diagonal.

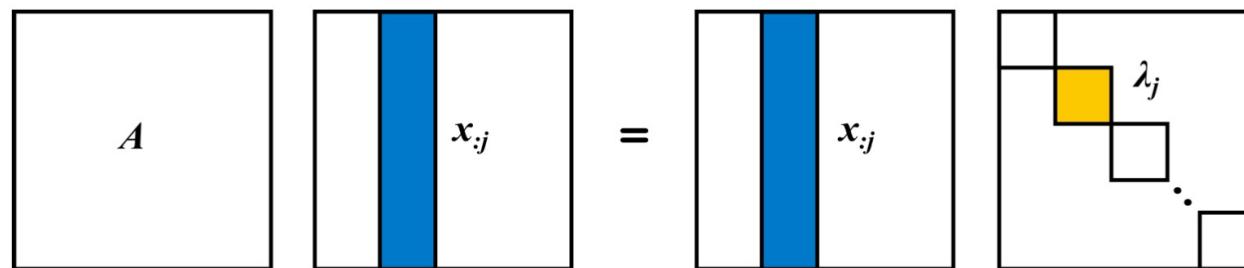


Figure 6.2. The matrix equation $AX = X\Lambda$ with the j th eigenvector (blue) and eigenvalue (yellow).

Spectral theorems

Example 6.3. Consider a matrix A and its characteristic equation,

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}, \quad \det(A - \lambda I) = (1 - \lambda)(2 - \lambda) = 0. \quad (6.5)$$

A has two distinct eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$, and by equation (6.2) each eigenvector $x = (x_1, x_2)$ must satisfy the system of equations $Ax = \lambda x$, that is,

$$\begin{aligned} x_1 + 2x_2 &= \lambda x_1, \\ 2x_2 &= \lambda x_2. \end{aligned}$$

Therefore, $x_2^{(1)} = 0$ for the eigenvector $x^{(1)}$ associated with $\lambda_1 = 1$, and $x_1^{(2)} = 2x_2^{(2)}$ for the eigenvector $x^{(2)}$ associated with $\lambda_2 = 2$. For example, we can choose the two eigenvectors

$$x^{(1)} = (1, 0)^T, \quad x^{(2)} = (2, 1)^T.$$

Spectral theorems

Example 6.6. The matrix defined in equation (6.5) is nondefective with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$, and therefore by Theorem 6.5 the following factorization exists,

$$\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}.$$

The eigenvectors $x^{(1)} = (1, 0)^T$ and $x^{(2)} = (2, 1)^T$ are linearly independent, and hence they span the vector space C^2 .

Spectral theorems

The matrix A is *unitary diagonalizable* if there exists a factorization

$$A = U\Lambda U^*,$$

with $U \in C^{n \times n}$ a unitary matrix with unit eigenvectors of A as column vectors, and $\Lambda \in C^{n \times n}$ a diagonal matrix with the eigenvalues of A on the diagonal. It turns out that for a matrix to be

Theorem 6.7 (Spectral theorem for normal matrices). *A matrix is unitary diagonalizable if and only if it is normal.*

Spectral theorems

Example 6.8. The normal matrix

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

has the purely imaginary eigenvalues $\lambda_1 = i$ and $\lambda_2 = -i$, for which an orthonormal basis of eigenvectors can be chosen,

$$x^{(1)} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}, \quad x^{(2)} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix},$$

which leads to the factorization

$$\begin{aligned} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{-1} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}. \end{aligned}$$

Spectral theorems

Theorem 6.9 (Spectral theorem for Hermitian matrices). *An Hermitian matrix is unitary diagonalizable with real eigenvalues. If the matrix is positive semi-definite the eigenvalues are non-negative, and if the matrix is positive definite all eigenvalues are positive.*

Proof.

$$0 = (Ax, x) - (Ax, x) = (Ax, x) - (x, Ax) = (\lambda x, x) - (x, \lambda x) = (\bar{\lambda} - \lambda)(x, x).$$

Hence, $\bar{\lambda} = \lambda$ because $(x, x) = \|x\|^2 > 0$. For a Hermitian positive definite matrix A , no eigenvalue can be zero or negative since

$$0 < x^* Ax = x^* \lambda x = \lambda \|x\|^2,$$

and analogously if A is Hermitian positive semi-definite all eigenvalues are non-negative.

Spectral theorems

As for the transpose in R^n , the adjoint in C^n satisfies the rule $(Ax, y) = (x, A^*y)$, since

$$(Ax, y) = (Ax)^*y = x^*A^*y = (x, A^*y).$$

Proof.

$$0 = (Ax, x) - (Ax, x) = (Ax, x) - (x, Ax) = (\lambda x, x) - (x, \lambda x) = (\bar{\lambda} - \lambda)(x, x).$$

Hence, $\bar{\lambda} = \lambda$ because $(x, x) = \|x\|^2 > 0$. For a Hermitian positive definite matrix A , no eigenvalue can be zero or negative since

$$0 < x^*Ax = x^*\lambda x = \lambda\|x\|^2,$$

and analogously if A is Hermitian positive semi-definite all eigenvalues are non-negative.

Spectral theorems

Example 6.10. The real symmetric positive definite matrix

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

has real positive eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 1$, for which we can choose the orthonormal eigenvector basis

$$x^{(1)} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad x^{(2)} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The resulting factorization is

$$\begin{aligned} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}^{-1} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \end{aligned}$$

Schur factorization

Theorem 6.11 (Schur factorization). *Every square matrix A has a Schur factorization*

$$A = UTU^*,$$

where U is a unitary $n \times n$ matrix, and T is an upper triangular $n \times n$ matrix with the eigenvalues of A on the diagonal.

From Theorem 6.11 it follows that the sum and the product of all eigenvalues $\{\lambda_i\}_{i=1}^n$ of a general $n \times n$ matrix A are related to the trace and the determinant of A as

$$\det(A) = \prod_{j=1}^n \lambda_j, \quad \text{trace}(A) = \sum_{j=1}^n \lambda_j$$

Singular value decomposition (SVD)

A non-negative real number $\sigma \geq 0$ is a *singular value* of A , if

$$Av = \sigma u,$$

$$A^*u = \sigma v,$$

for $u \in C^m$ the *left-singular vector* and $v \in C^n$ the *right-singular vector*.

Singular value decomposition (SVD)

Theorem 6.14 (Singular value decomposition). *For any $m \times n$ real or complex matrix A , there exists a singular value decomposition*

$$A = U\Sigma V^*$$

where U is an $m \times m$ unitary matrix and V an $n \times n$ unitary matrix, and where Σ is an $m \times n$ diagonal matrix with non-negative real numbers σ_i on the diagonal, ordered from the highest numbers to the lowest. The singular values σ_i are unique, whereas the matrices U and V are unique only if A is a square matrix.

Singular value decomposition (SVD)

$$A = \begin{matrix} & \\ & \\ & \end{matrix} = \begin{matrix} & & \\ & \textcolor{violet}{|} & \\ & & U \end{matrix} \begin{matrix} & & \\ & \textcolor{black}{\textcolor{white}{\square}} & \\ & \textcolor{black}{\textcolor{white}{\square}} & \textcolor{black}{\textcolor{white}{\square}} & \\ & \textcolor{black}{\textcolor{white}{\square}} & \textcolor{black}{\textcolor{white}{\square}} & \textcolor{black}{\textcolor{white}{\square}} & \\ & \textcolor{black}{\textcolor{white}{\square}} & \textcolor{black}{\textcolor{white}{\square}} & \textcolor{black}{\textcolor{white}{\square}} & \textcolor{black}{\textcolor{white}{\square}} & \ddots & \\ & & & & & & \textcolor{black}{\textcolor{white}{\square}} \end{matrix} \begin{matrix} & \\ & \\ & \textcolor{blue}{|} & \\ & & V^* \end{matrix}$$

Singular value decomposition (SVD)

$$\begin{aligned} A^*A &= (U\Sigma V^*)^* U\Sigma V^* = V\Sigma^* U^* U\Sigma V^* = V(\Sigma^* \Sigma) V^*, \\ AA^* &= U\Sigma V^* (U\Sigma V^*)^* = U\Sigma V^* V\Sigma^* U^* = U(\Sigma \Sigma^*) U^*. \end{aligned}$$

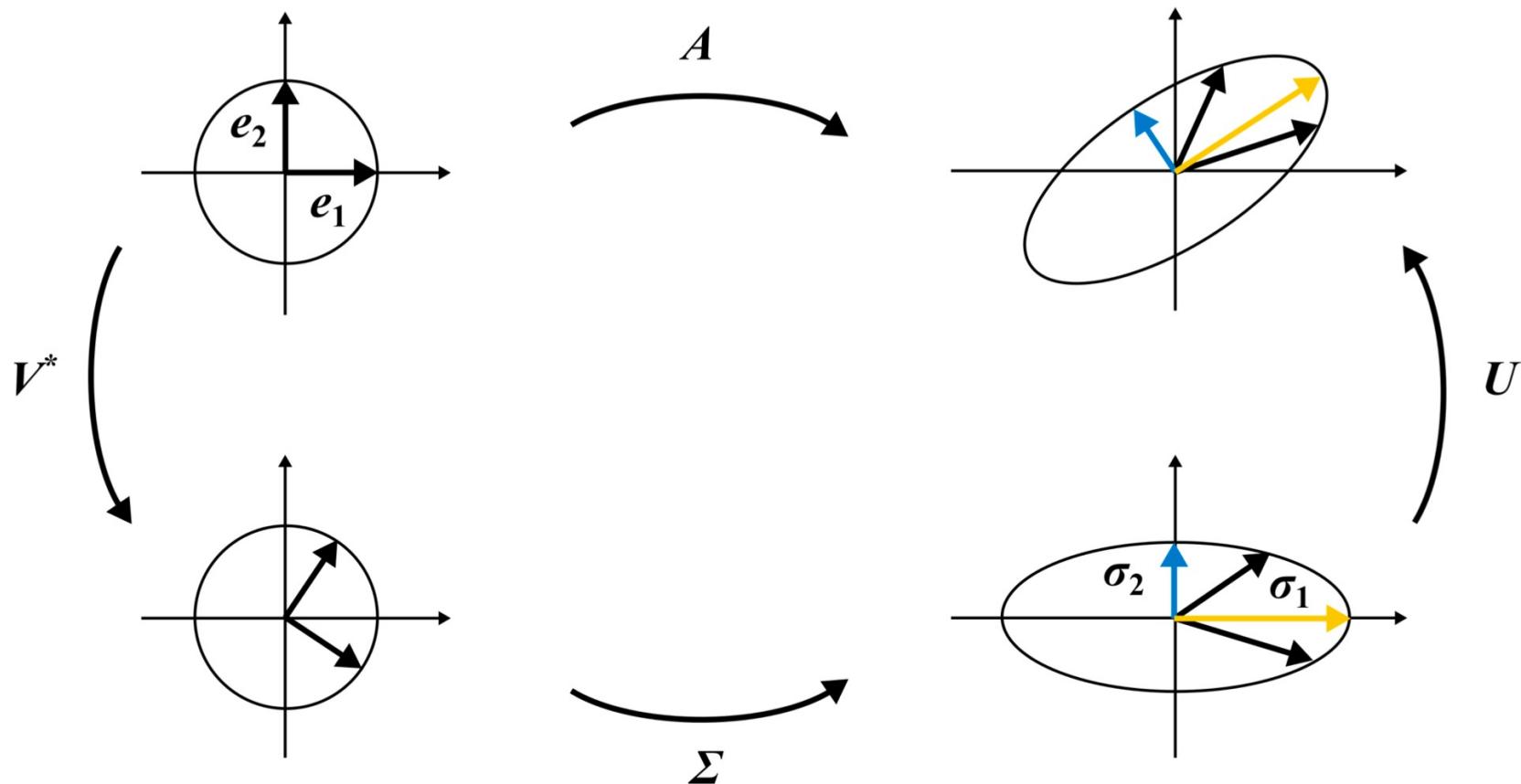
Since A^*A and AA^* are Hermitian positive semi-definite, we find that the columns of V form an orthonormal set of eigenvectors of A^*A , the columns of U an orthonormal set of eigenvectors of AA^* , and the non-negative singular values on the diagonal of Σ are the square roots of the non-negative eigenvalues of AA^* and A^*A . Hence, the existence of a SVD for any matrix A is given by the spectral theorem for Hermitian matrices, and if A is a square matrix the SVD is unique. If A is a rectangular $m \times n$ matrix with $m > n$, then the last $m - n$ columns of U is a set of unit vectors defined such that the columns of U form an orthonormal basis for C^m , equivalent to U being a unitary matrix.

Singular value decomposition (SVD)

Theorem 6.15 (Fundamental theorem of linear algebra part III). *Let A be an $m \times n$ matrix of rank r , with singular value decomposition $A = U\Sigma V^*$. Then the four fundamental subspaces are spanned by the columns of the unitary matrices $U = (u_{:j})$ and $V = (v_{:j})$,*

- (i) $\text{range}(A) = \text{span}\{u_{:j}\}_{j=1}^r$,
- (ii) $\text{null}(A^*) = \text{span}\{u_{:j}\}_{j=r+1}^m$,
- (iii) $\text{range}(A^*) = \text{span}\{v_{:j}\}_{j=1}^r$,
- (iv) $\text{null}(A) = \text{span}\{v_{:j}\}_{j=r+1}^n$.

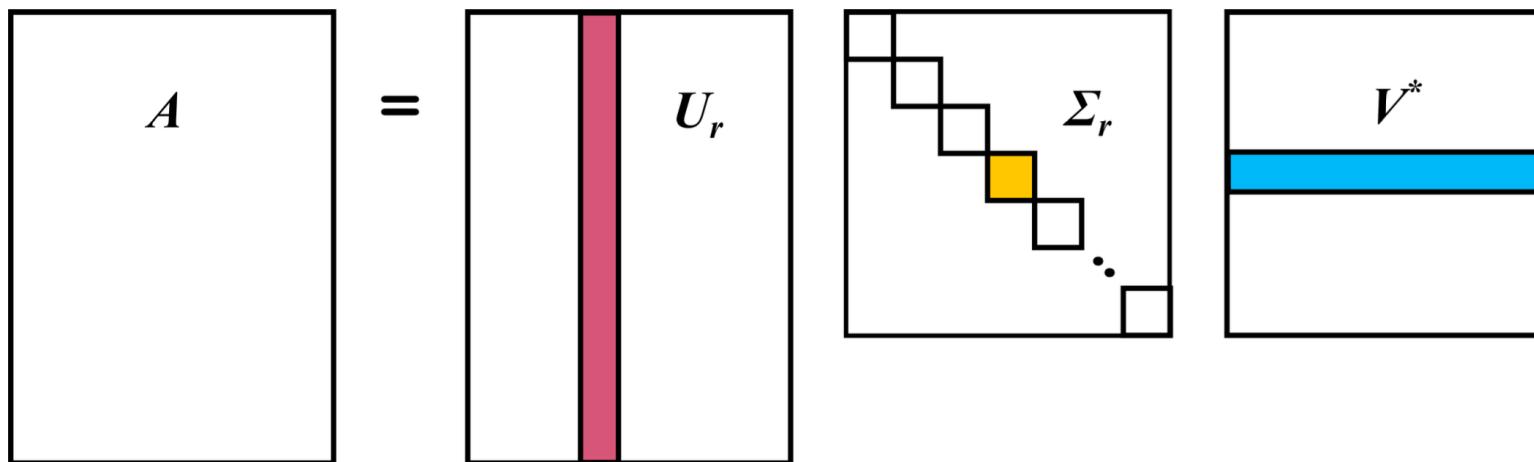
Singular value decomposition (SVD)



Reduced SVD

$$A = U_r \Sigma_r V^*,$$

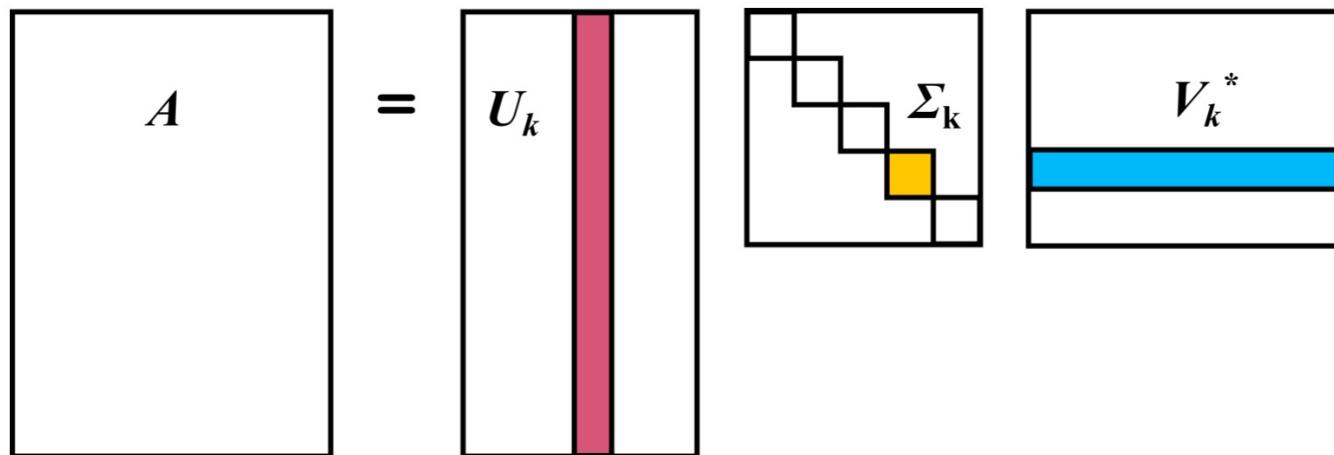
with the matrices $U_r \in C^{m \times n}$, $\Sigma_r \in C^{n \times n}$ and $V \in C^{n \times n}$. Note that the reduced matrix U_r is not a unitary matrix, and that it does not provide the full basis for C^m .



Truncated SVD

$$A \approx A_k = U_k \Sigma_k V_k^*,$$

with $U_k \in C^{m \times k}$, $\Sigma_k \in C^{k \times k}$ and $V_k \in C^{k \times n}$, where now neither of the reduced matrices U_k and V_k are unitary matrices. We refer to A_k as a low rank approximation of A .



QR algorithm

$$Q^{(k)} R^{(k)} = A^{(k-1)},$$

where in each iteration k the order of the factors is reversed,

$$A^{(k)} = R^{(k)} Q^{(k)} = (Q^{(k)})^{-1} A^{(k-1)} Q^{(k)} = Q^{(k)*} A^{(k-1)} Q^{(k)},$$

which means that $A^{(k)}$ and $A^{(k-1)}$ are similar, and therefore have the same eigenvalues. In fact, for any k , $A^{(k)}$ is similar to the input matrix A , since

$$\begin{aligned} A^{(k)} &= Q^{(k)*} A^{(k-1)} Q^{(k)} = Q^{(k)*} Q^{(k-1)*} A^{(k-2)} Q^{(k-1)} Q^{(k)} \\ &= Q^{(k)*} Q^{(k-1)*} \dots Q^{(1)*} A Q^{(1)} \dots Q^{(k-1)} Q^{(k)} \equiv U^{(k)*} A U^{(k)} \end{aligned}$$

QR algorithm

Hence, after k iterations we have the approximate Schur factorization

$$A = U^{(k)} A^{(k)} U^{(k)*},$$

where under suitable conditions $A^{(k)}$ will converge to an upper triangular matrix which has the eigenvalues of A on the diagonal. If A is a normal matrix, $A^{(k)}$ instead converges to the unitary diagonal form with the column vectors of the unitary matrix $U^{(k)}$ approximate eigenvectors associated to each eigenvalue.

QR algorithm

ALGORITHM 6.1. $(A, U) = \text{qr_algorithm}(A)$.

Input: a general $n \times n$ matrix A .

Output: approximate Schur factorization $n \times n$ matrices A and U .

```
1:  $U = I$ 
2: while stopping_criterion == false do
3:    $(Q, R) = \text{qr\_factorization}(A)$ 
4:    $A = \text{matrix\_matrix\_product}(R, Q)$ 
5:    $U = \text{matrix\_matrix\_product}(U, Q)$ 
6: end while
7: return  $A, U$ 
```

Rayleigh quotient (real & symmetric matrix)

We first consider the question: given a vector $x \in R^n$, what is the real number $\xi \in R$ that best approximates an eigenvalue of a symmetric matrix $A \in R^{n \times n}$ in the sense that $\|Ax - \xi x\|$ is minimized? If $x = q_j$ is an eigenvector of A , then $\xi = \lambda_j$ is the corresponding eigenvalue. If not, ξ is the solution to the $n \times 1$ least squares problem

$$\min_{\xi \in R} \|Ax - \xi x\|,$$

for which the normal equations take the form $x^T Ax = x^T \xi x$. Hence, we can view ξ as a function of x , which we define as the *Rayleigh quotient*

$$r(x) = \frac{x^T Ax}{x^T x}$$

Rayleigh quotient (real & symmetric matrix)

$$r(x) = \frac{x^T Ax}{x^T x},$$

where $r(x)$ is an approximation of the eigenvalue λ_j if x is close to the eigenvector q_j . If we assume that $x \rightarrow q_j$, then $r(x)$ converges quadratically to $r(q_j) = \lambda_j$,

$$r(x) - r(q_j) = \mathcal{O}(\|x - q_j\|^2)$$

Power iteration (real & symmetric matrix)

For a real symmetric $n \times n$ matrix A , the eigenvectors $\{q_j\}_{j=1}^n$ form an orthonormal basis for R^n , so that we can write any vector $v \in R^n$ as

$$v = \sum_{j=1}^n \alpha_j q_j,$$

with the coordinates $\alpha_j = (v, q_j)$. Furthermore, we can express the map $v \mapsto Av$ in terms of the eigenvalues λ_j of the matrix A ,

$$Av = \sum_{j=1}^n \alpha_j Aq_j = \sum_{j=1}^n \alpha_j \lambda_j q_j.$$

Power iteration (real & symmetric matrix)

The map amounts to a scaling of each eigenvector q_j by λ_j , and if iterated,

$$A^k v = \sum_{j=1}^n \alpha_j A^k q_j = \sum_{j=1}^n \alpha_j \lambda_j^k q_j.$$

Therefore, as $k \rightarrow \infty$, each A^k eigenvector $\lambda_j^k q_j$ converges to zero if $|\lambda_j| < 1$, or infinity if $|\lambda_j| > 1$. Now assume that

$$\alpha_1 = (v, q_1) \neq 0,$$

and that the eigenvalues of A are ordered such that

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|.$$

Power iteration (real & symmetric matrix)

We say that λ_1 is the *dominant eigenvalue* and q_1 the *dominant eigenvector*. It follows that $|\lambda_j/\lambda_1| < 1$ for all $j > 1$, which implies that $(\lambda_j/\lambda_1)^k \rightarrow 0$, as $k \rightarrow \infty$. By equation (6.21) we can write

$$A^k v = \lambda_1^k (\alpha_1 q_1 + \sum_{j=2}^n \alpha_j (\lambda_j/\lambda_1)^k q_j),$$

and therefore the approximation

$$A^k v \approx \lambda_1^k \alpha_1 q_1$$

improves for each iteration k . Hence, $A^k v$ approaches a multiple of the eigenvector q_1 , and we obtain a unit approximation by normalization,

$$v^{(k)} = A^k v / \|A^k v\| \approx q_1.$$

Power iteration (real & symmetric matrix)

ALGORITHM 6.2. **(lambda, v) = power_iteration(A).**

Input: an $n \times n$ matrix **A**.

Output: the largest eigenvalue **lambda** and the associated unit eigenvector **v**.

```
1: v = random_vector()
2: while stopping_criterion == false do
3:   v = matrix_vector_product(A, v)
4:   v[:] = v[:]/norm(v)
5: end while
6: v0 = v
7: v = matrix_vector_product(A, v)
8: lambda = dot(v0, Av)
9: return lambda, v
```

Power iteration (real & symmetric matrix)

- Power iteration converges to the dominant eigenvector at a linear rate with the constant factor $|\lambda_2 / \lambda_1|$.
- The idea of *inverse iteration* is to apply power iteration to the matrix $(A - \mu I)^{-1}$, with the eigenvalues $(\lambda_j - \mu)^{-1}$ and with the same eigenvectors as A.
- With the *shift* μ an approximation of λ_j , the eigenvalue $(\lambda_j - \mu)^{-1}$ can be expected to be dominant and much larger than the other eigenvalues, which results in an accelerated convergence.

Rayleigh quotient iteration

ALGORITHM 6.3. $(\lambda, v) = \text{rayleigh_quotient_iteration}(A)$.

Input: an $n \times n$ matrix A .

Output: the largest eigenvalue λ and an associated unit eigenvector v .

- 1: $(\lambda, v) = \text{power_iteration}(A)$
- 2: **while** stopping_criterion == false **do**
- 3: $B = \text{compute_inverse_matrix}(A - \lambda I)$
- 4: $(\lambda, v) = \text{power_iteration}(B)$
- 5: **end while**
- 6: **return** λ, v

QR algorithm as power iteration

We now revisit the QR algorithm. Let $Q^{(k)}$ and $R^{(k)}$ be a sequence of matrices generated by the QR algorithm. Then the matrix products

$$\hat{Q}^{(k)} = Q^{(1)}Q^{(2)} \dots Q^{(k)}, \quad \hat{R}^{(k)} = R^{(k)}R^{(k-1)} \dots R^{(1)},$$

correspond to a QR factorization of the k th power of the matrix A , $A^k = \hat{Q}^{(k)}\hat{R}^{(k)}$, which is proved by induction, using the property that $R^{(k)}Q^{(k)} = Q^{(k+1)}R^{(k+1)}$,

$$A = QR = Q^{(1)}R^{(1)},$$

$$A^2 = AA = Q^{(1)}R^{(1)}Q^{(1)}R^{(1)} = Q^{(1)}Q^{(2)}R^{(2)}R^{(1)},$$

$$\begin{aligned} A^3 &= A^2A = Q^{(1)}Q^{(2)}R^{(2)}R^{(1)}Q^{(1)}R^{(1)} = Q^{(1)}Q^{(2)}R^{(2)}Q^{(2)}R^{(2)}R^{(1)} \\ &= Q^{(1)}Q^{(2)}Q^{(3)}R^{(3)}R^{(2)}R^{(1)}, \end{aligned}$$

...

$$A^k = A^{k-1}A = \hat{Q}^{(k-1)}\hat{R}^{(k-1)}Q^{(1)}R^{(1)} = \hat{Q}^{(k)}\hat{R}^{(k)}.$$

QR algorithm as power iteration

That is, the QR algorithm constructs successive orthonormal bases for the powers A^k , functioning like a power iteration that simultaneously iterates on the whole set of approximate eigenvectors. Furthermore, the diagonal components of the k th iterate $A^{(k)}$ are the Rayleigh quotients of A corresponding to the column vectors of $\hat{Q}^{(k)}$,

$$A^{(k)} = \hat{Q}^{(k)T} A \hat{Q}^{(k)}.$$

Hence, the diagonal components of $A^{(k)}$ converges (quadratically) to the eigenvalues of A .

QR algorithm with shifts

The QR algorithm can be accelerated by introducing a shift $A - \mu I$. For multiple shifts $\mu^{(k)}$, each shifted matrix is factorized according to

$$A^{(k)} - \mu^{(k)}I = Q^{(k+1)}R^{(k+1)},$$

and the new iterate is chosen as $A^{(k+1)} = R^{(k+1)}Q^{(k+1)} + \mu^{(k)}I$. Analogous to equation (6.24),

$$\begin{aligned} p_m(A) &= (A - \mu^{(1)}I)\dots(A - \mu^{(m)}I) \\ &= (Q^{(1)}\dots Q^{(m)})(R^{(m)}\dots R^{(1)}) = \hat{Q}^{(m)}\hat{R}^{(m)}, \end{aligned} \quad (6.26)$$

and as in equation (6.25), the iterate $A^{(m)}$ corresponds to a similarity transformation of A ,

$$A^{(m)} = R^{(m)}Q^{(m)} + \mu^{(m-1)}I = (Q^{(1)}\dots Q^{(m)})^T A (Q^{(1)}\dots Q^{(m)}) = \hat{Q}^{(m)T} A \hat{Q}^{(m)}. \quad (6.27)$$

Therefore, from the QR factorization (6.26) we can compute $A^{(m)}$ from equation (6.27).

Reduction to Hessenberg form $H = U^T A U$

ALGORITHM 6.4. $\mathbf{H} = \text{reduction_to_hessenberg}(\mathbf{A})$.

Input: a real $n \times n$ matrix \mathbf{A} .

Output: Hessenberg matrix \mathbf{H} .

```
1: for  $k=1:n-2$  do
2:    $x[:] = A[k:n-1,k]$ 
3:    $v_k[:] = x[:]$ 
4:    $v_k[0] = v_k[0] - \text{sign}(x[0]) * \text{norm}(x)$ 
5:    $v_k[:] = v_k[:] / \text{norm}(v_k)$ 
6:   for  $m=k:n-1$  do
7:      $A[k:n-1,m] = A[k:n-1,m] - 2 * v_k[:] * \text{dot}(v_k, A[k:n-1,m])$ 
8:      $A[0:n-1,m] = A[0:n-1,m] - 2 * \text{dot}(v_k, A[0:n-1,m]) * v_k[:]$ 
9:   end for
10: end for
11: return  $A$ 
```

Reduction to Hessenberg form $H = U^T A U$

ALGORITHM 6.4. $\mathbf{H} = \text{reduction_to_hessenberg}(\mathbf{A})$.

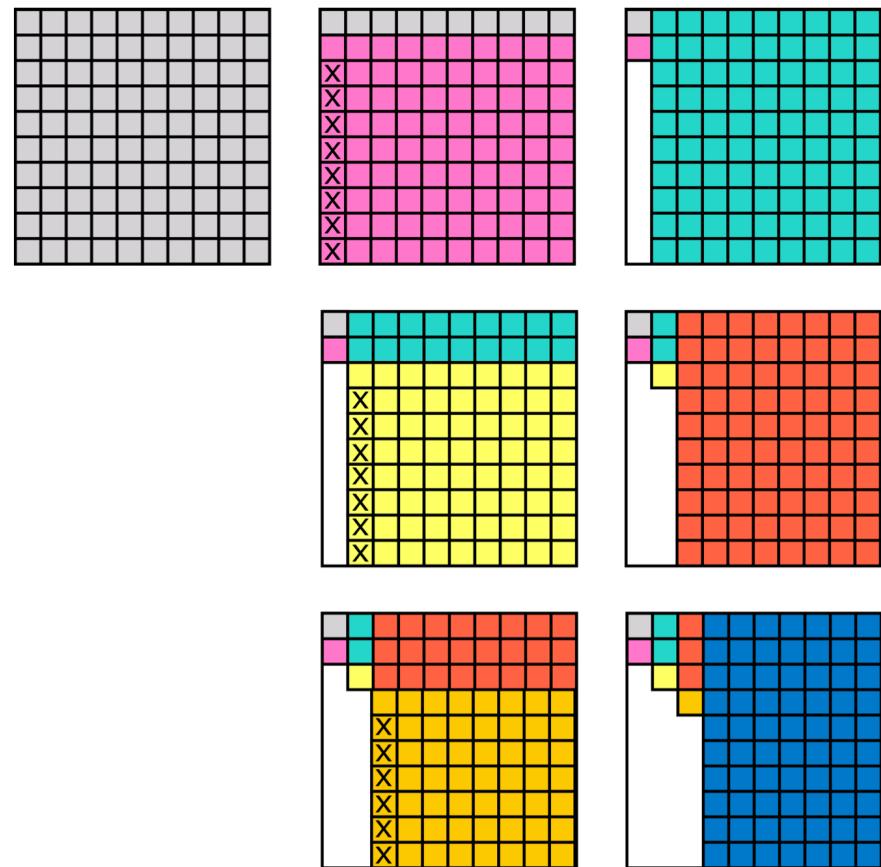
Input: a real $n \times n$ matrix \mathbf{A} .

Output: Hessenberg matrix \mathbf{H} .

```

1: for  $k=1:n-2$  do
2:    $x[:] = A[k:n-1,k]$ 
3:    $v_k[:] = x[:]$ 
4:    $v_k[0] = v_k[0] - \text{sign}(x[0]) * \text{norm}(x)$ 
5:    $v_k[:] = v_k[:] / \text{norm}(v_k)$ 
6:   for  $m=k:n-1$  do
7:      $A[k:n-1,m] = A[k:n-1,m] - 2*v_k[:] * \text{dot}(v_k, A[k:n-1,m])$ 
8:      $A[0:n-1,m] = A[0:n-1,m] - 2*\text{dot}(v_k, A[0:n-1,m]) * v_k[:]$ 
9:   end for
10: end for
11: return  $A$ 

```



Implicit QR algorithm

Theorem 6.16 (Implicit Q theorem). *For an $n \times n$ matrix A , real or complex, let $H = U^*AU$ be a Hessenberg matrix and with U a unitary matrix. If all subdiagonal components of H are nonzero, then up to scaling of the columns in U , the matrices H and U are uniquely determined by the first or last column of U .*

If the previous iterate $H^{(k-1)}$ is a Hessenberg matrix with nonzero subdiagonal components, then by the implicit Q theorem $\hat{Q}^{(m)T}H^{(k-1)}\hat{Q}^{(m)}$ it is uniquely determined by the first column of $\hat{Q}^{(m)}$ from the QR factorization

$$p_m(H^{(k-1)}) = \hat{Q}^{(m)}\hat{R}^{(m)}.$$

The first column of this equation states that

$$p_m(H^{(k-1)})e_1 = \hat{Q}^{(m)}\hat{R}^{(m)}e_1 = \hat{Q}^{(m)}\alpha e_1$$

Implicit QR algorithm

ALGORITHM 6.5. $H = \text{implicit_qr_algorithm}(A)$.

Input: a general $n \times n$ matrix A .

Output: approximate Schur factorization $n \times n$ matrix H .

- 1: $H = \text{reduction_to_hessenberg}(A)$
- 2: **while** stopping_criterion == false **do**
- 3: Select shifts.
- 4: Compute first column of $p(H)$.
- 5: Construct Householder transformation matrix Q_0 based on first column of $p(H)$.
- 6: Compute similarity transformation of H by matrix Q_0 .
- 7: Chase the bulge by the Householder transformation matrix Q_1 .
- 8: **end while**
- 9: **return** H

Implicit QR algorithm

ALGORITHM 6.5. $\mathbf{H} = \text{implicit_qr_algorithm}(\mathbf{A})$.

Input: a general $n \times n$ matrix \mathbf{A} .

Output: approximate Schur factorization $n \times n$ matrix \mathbf{H} .

```
1:  $\mathbf{H} = \text{reduction\_to\_hessenberg}(\mathbf{A})$ 
2: while stopping_criterion == false do
3:   Select shifts.
4:   Compute first column of  $p(\mathbf{H})$ .
5:   Construct Householder transformation matrix  $\mathbf{Q}_0$  based on first column of  $p(\mathbf{H})$ .
6:   Compute similarity transformation of  $\mathbf{H}$  by matrix  $\mathbf{Q}_0$ .
7:   Chase the bulge by the Householder transformation matrix  $\mathbf{Q}_1$ .
8: end while
9: return  $\mathbf{H}$ 
```

