Chapter 6

Eigenvalues and eigenvectors

We review the basic properties of eigenvalues and eigenvectors of a matrix, and we present an algorithm for computational approximation in the case of real symmetric matrices.

6.1 Eigenvalues and eigenvectors

Complex vector spaces

In this section we change the focus from real vector spaces to complex vector spaces. We let $z \in \mathbb{C}$ denote a complex scalar, with $\bar{z}$ the complex conjugate of z.

The complex vector space $\mathbb{C}^n$ is defined by the basic operations of componentwise addition and scalar multiplication of complex numbers, and with the transpose of a vector $x \in \mathbb{C}^n$ replaced by the *adjoint* x^*, corresponding to the transpose with the imaginary parts negated. Similarly, the adjoint of a complex $m \times n$ matrix $A = (a_{ij})$ is the $n \times m$ matrix $A^* = (\bar{a}_{ji})$. If $A = A^*$ the matrix A is *Hermitian*, and if $AA^* = A^*A$ it is *normal*.

The inner product of $x, y \in \mathbb{C}^n$ is defined by

$$(x, y) = x^* y = \sum_{i=i}^{n} \bar{x}_i y_i, \tag{6.1}$$

with the associated norm for $x \in \mathbb{C}^n$,

$$\|x\| = (x, x)^{1/2}. \tag{6.2}$$

Matrix spectrum and eigenspaces

We now consider a square matrix $A \in \mathbb{C}^{n\times n}$ acting on a complex vector space $\mathbb{C}^n$. An *eigenvector* of A is a nonzero vector $x \in \mathbb{C}^n$, such that

$$Ax = \lambda x, \tag{6.3}$$

with $\lambda \in \mathbb{C}$ the corresponding *eigenvalue*. The subspace of $\mathbb{C}^n$ spanned by the eigenvectors corresponding to λ, together with the zero vector, is an *eigenspace* E_λ, and the set of all eigenvalues $\{\lambda_j\}_{j=1}^n$ is the *spectrum* of A, denoted by $\Lambda(A)$. The sum and the product of all eigenvalues are related to the trace and the determinant of A as

$$\det(A) = \prod_{j=1}^{n} \lambda_j \qquad \mathrm{tr}(A) = \sum_{j=1}^{n} \lambda_j. \tag{6.4}$$

The eigenspace E_λ is an *invariant subspace* under A, so that $AE_\lambda \subseteq E_\lambda$, and $\dim(E_\lambda)$ is the number of linearly independent eigenvectors corresponding to the eigenvalue λ, known as the *geometric multiplicity* of λ.

We have that the eigenspace $E_\lambda = \mathrm{null}(\lambda I - A)$, since $(\lambda I - A)x = 0$, and thus for a nonempty eigenspace E_λ, $\lambda I - A$ is a singular matrix, so that

$$\det(\lambda I - A) = 0. \tag{6.5}$$

Characteristic polynomial

The *characteristic polynomial* of the matrix $A \in \mathbb{C}^{n\times n}$, is the degree n polynomial

$$p_A(z) = \det(zI - A), \tag{6.6}$$

with $z \in \mathbb{C}$. For λ an eigenvalue of A, we thus have that

$$p_A(\lambda) = 0, \tag{6.7}$$

and by the fundamental theorem of algebra we can express $p_A(\lambda)$ as

$$p_A(\lambda) = (z - \lambda_1)(z - \lambda_2)\cdots(z - \lambda_n), \tag{6.8}$$

where each λ_j is an eigenvalue of A, not necessary unique. The multiplicity of each eigenvalue λ as a root to the equation $p_A(\lambda) = 0$ is the *algebraic multiplicity* of λ, where an eigenvalue is said to be *simple* if its algebraic multiplicity is 1. The algebraic multiplicity of an eigenvalue λ is at least as great as its geometric multiplicity.

Eigenvalue decompositions

A *defective matrix* is a matrix which has one or more *defective eigenvalues*, where a defective eigenvalue is an eigenvalue for which its algebraic multiplicity exceeds its geometric multiplicity.

Theorem 5 (Eigenvalue decomposition)**.** *Each nondefective matrix* $A \in \mathbb{C}^{n\times n}$ *has an eigenvalue decomposition*

$$A = X\Lambda X^{-1}, \tag{6.9}$$

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

We also say that a nondefective matrix is *diagonalizable*. Given that the factorization (6.9) exits, we have that

$$AX = X\Lambda, \tag{6.10}$$

which expresses (6.3) as

$$Ax_j = \lambda_j x_j, \tag{6.11}$$

with λ_j the jth diagonal entry of Λ, and x_j the jth column of X.

For some matrices eigenvectors can be chosen to be pairwise orthogonal, so that a matrix A is *unitary diagonalizable*, that is

$$A = Q\Lambda Q^*, \tag{6.12}$$

with $Q \in \mathbb{C}^{n\times n}$ an orthogonal matrix with orthonormal eigenvectors of A as column vectors, and $\Lambda \in \mathbb{C}^{n\times n}$ a diagonal matrix with the eigenvalues of A on the diagonal.

Theorem 6. *A matrix is unitary diagonalizable if and only if it is normal.*

Hermitian matrices have real eigenvalues, and thus in the particular case of a real symmetric matrix, no complex vector spaces are needed to characterize the matrix spectrum and eigenspaces.

Theorem 7. *An Hermitian matrix is unitary diagonalizable with real eigenvalues.*

Irrespectively if the matrix is nondefective or Hermitian, any square matrix always has a Schur factorization, with the diagonal matrix replaced by an upper triangular matrix.

Theorem 8 (Schur factorization)**.** *For every square matrix A there exists a Schur factorization*

$$A = QTQ^*, \tag{6.13}$$

where Q is an orthogonal matrix, and T is an upper triangular matrix with the eigenvalues of A on the diagonal.

More generally, if $X \in \mathbb{C}^{n\times n}$ is nonsingular, the map $A \mapsto X^{-1}AX$ is a *similarity transformation* of A, and we say that two matrices A and B are *similar* if there exists a similarity transformation such that $B = X^{-1}AX$.

Theorem 9. *Two similar matrices have the same eigenvalues with the same algebraic and geometric multiplicity.*

6.2 Eigenvalue algorithms

QR algorithm

To compute the eigenvalues of a matrix A, one may seek the roots of the characteristic polynomial. Although, for a large matrix polynomial root finding is expensive and unstable. Instead the most efficient algorithms are based on computing eigenvalues and eigenvectors by constructing one of the factorizations (6.9), (6.12) or (6.13).

We now present the *QR algorithm*, in which a Schur factorization (6.13) of a matrix A is constructed from successive QR factorizations.

Algorithm 4: QR algorithm

$A^{(0)} = A$
for $k = 1, 2, ...$ **do**
  $Q^{(k)}R^{(k)} = A^{(k-1)}$
  $A^k = R^{(k)}Q^{(k)}$
end

We note that for each iteration $A^{(k)}$ of the algorithm, we have that

$$A^{(k)} = R^{(k)}Q^{(k)} = (Q^{(k)})^{-1}A^{(k-1)}Q^{(k)}, \tag{6.14}$$

so that $A^{(k)}$ and $A^{(k-1)}$ are similar, and thus have the same eigenvalues. Under suitable assumptions A^k will converge to an upper triangular matrix, or in the case of a Hermitian matrix a diagonal matrix, with the eigenvalues on the diagonal.

The basic QR algorithm can be accelerated: (i) by Householder reflectors to reduce the initial matrix $A^{(0)}$ to *Hessenberg form*, that is a matrix with zeros below the first subdiagonal (or in the case of an Hermitian matrix a *tridiagonal form*), (ii) by introducing a *shift* to instead of $A^{(k)}$ factorize the matrix $A^{(k)} - \mu^{(k)} I$, which has identical eigenvectors, and where $\mu^{(k)}$ an eigenvalue estimate, and (iii) if any off-diagonal element is close to zero, both off-diagonal elements are zeroed out to deflate the matrix $A^{(k)}$ into submatrices on which the QR algorithm is then applied.

Rayleigh quotient

To simplify the presentation, in the rest of this section we restrict attention to matrices that are real and symmetric, for which all eigenvalues λ_j are real and the corresponding eigenvectors q_j are orthonormal.

We now consider the question: given a vector $x \in \mathbb{R}^n$, what is the real number $\alpha \in \mathbb{R}$ that best approximate an eigenvalue of A in the sense that $\|Ax - \alpha x\|$ is minimized?

If $x = q_j$ is an eigenvector of A, then $\alpha = \lambda_j$ is the corresponding eigenvalue. If not, α is the solution to the $n \times 1$ least squares problem

$$\min_{\alpha \in \mathbb{R}} \|Ax - \alpha x\|, \tag{6.15}$$

for which the normal equations are given as

$$x^T A x = x^T \alpha x. \tag{6.16}$$

With $\alpha = r(x)$, we define the *Rayleigh quotient* as

$$r(x) = \frac{x^T A x}{x^T x}, \tag{6.17}$$

where $r(x)$ is an approximation of an eigenvalue λ_j, if x is close to the eigenvector q_j. In fact, $r(x)$ converges quadratically to $r(q_j) = \lambda_j$, that is

$$r(x) - r(q_j) = O(\|x - q_j\|^2), \tag{6.18}$$

as $x \to q_j$.

Power iteration

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

$$v = \sum_{j=1}^{n} \alpha_j q_j, \tag{6.19}$$

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

$$Av = \sum_{j=1}^{n} \alpha_j A q_j = \sum_{j=1}^{n} \alpha_j \lambda_j q_j, \tag{6.20}$$

which if iterated gives

$$A^k v = \sum_{j=1}^{n} \alpha_j A^k q_j = \sum_{j=1}^{n} \alpha_j \lambda_j^k q_j = \lambda_1^k (\alpha_1 q_1 + \sum_{j=2}^{n} \alpha_j (\lambda_j/\lambda_1)^k q_j). \tag{6.21}$$

Now assume that $\alpha_1 = (v, q_1) \neq 0$, and that the eigenvalues of A are ordered such that

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|, \tag{6.22}$$

where we say that λ_1 is the *dominant eigenvalue*, and q_1 the *dominant eigenvector*. Thus $|\lambda_j/\lambda_1| < 1$ for all j, which implies that

$$(\lambda_j/\lambda_1)^k \to 0, \tag{6.23}$$

as $k \to \infty$, and thus that the approximation

$$A^k v \approx \lambda_1^k \alpha_1 q_1, \tag{6.24}$$

improves as k increases. That is, $A^k v$ approaches a multiple of the eigenvector q_j, which can be obtained by normalizating, so that

$$v^{(k)} \equiv A^k v / \|A^k v\| \approx q_j, \tag{6.25}$$

from which an approximation of the corresponding eigenvalue can be obtained by the Rayleigh quotient, which leads us to *power iteration*.

Algorithm 5: Power iteration

$v^{(0)}$ such that $\|v^{(0)}\| = 1$
for $k = 1, 2, ...$ **do**
  $w = Av^{(k-1)}$ ▷ `apply` A
  $v^{(k)} = w/\|w\|$ ▷ `normalize`
  $\lambda^{(k)} = (v^{(k)})^T A v^{(k)}$ ▷ `Raylegh quotient`
end

The convergence of the Power iteration to the dominant eigenvector is linear by a constant factor $|\lambda_2/\lambda_1|$, whereas the convergence to the dominant eigenvalue is quadratic in the same factor, due to the convergence of Rayleigh quotient.

Efficiency of the algorithm thus depends on the size of the factor $|\lambda_2/\lambda_1|$. The idea of *inverse iteration*, is to apply power iteration to the matrix $(A-\mu I)^{-1}$, with eigenvalues $\{(\lambda_j-\mu)^{-1}\}$ and with the same eigenvectors as A, since

$$Av = \lambda v \Leftrightarrow (A-\mu I)v = (\lambda-\mu)v \Leftrightarrow (\lambda-\mu)^{-1}v = (A-\mu I)^{-1}v. \quad (6.26)$$

With 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 of power iteration.

Rayleigh quotient iteration is inverse iteration where μ is updated to the eigenvalue approximation of the previous step of the iteration, and the convergence to an eigenvalue/eigenvector pair is cubic.

Algorithm 6: Rayleigh quotient iteration

$v^{(0)}$ such that $\|v^{(0)}\| = 1$
$\lambda^{(0)} = (v^{(0)})^T A v^{(0)}$
for $k = 1, 2, \ldots$ **do**
  Solve $(A-\lambda^{(k-1)})w = v^{(k-1)}$ for w ▷ `apply` $(A-\lambda^{(k-1)}I)^{-1}$
  $v^{(k)} = w/\|w\|$ ▷ `normalize`
  $\lambda^{(k)} = (v^{(k)})^T A v^{(k)}$ ▷ `Raylegh quotient`
end

The QR algorithm as a power iteration

We now revisit the QR algorithm. Let $Q^{(k)}$ and $R^{(k)}$ be the matrices generated from the (unshifted) QR algorithm, then the matrix products

$$\underline{Q}^{(k)} = Q^{(1)}Q^{(2)}\cdots Q^{(k)}, \quad (6.27)$$

and

$$\underline{R}^{(k)} = R^{(k)}R^{(k-1)}\cdots R^{(1)}, \quad (6.28)$$

correspond to a QR factorization of the kth power of A,

$$A^k = \underline{Q}^{(k)}\underline{R}^{(k)}, \quad (6.29)$$

which can be proven by induction.

That is, the QR algorithm constructs successive orthonormal bases for the powers A^k, thus functioning like a power iteration that simultaneously iterates on the whole set of approximate eigenvectors.

Further, the diagonal elements of the kth iterate $A^{(k)}$ are the Rayleigh quotients of A corresponding to the column vectors of $\underline{Q}^{(k)}$,

$$A^{(k)} = (\underline{Q}^{(k)})^T A \underline{Q}^{(k)}, \tag{6.30}$$

and thus the diagonal elements of $A^{(k)}$ converges (quadratically) to the eigenvalues of A.

With the accelerations (i)-(iii) the QR algorithm exhibit cubic converge rate in both eigenvalues and eigenvectors.