Graph/Matrix:

Spectral Theory, Computation & Applications

Xiaobai Sun

Initial draft: Aug. 2024

Contents

1 Preliminary

1.1 Basic notions & notations

Let $A \in \mathbb{C}^{n\times n}$, $n > 1$.

- The Hermitian transpose of A is A^{H}. A matrix is Hermitian if $A^{\mathrm{H}} = A$. A matrix is (complex) symmetric if $A^{\mathrm{T}} = A$.

- An orthonormal matrix $U_{n\times k}$ is composed of mutually orthogonal columns and each column normalized to 1, i.e., $e_i^{\mathrm{T}} U^{\mathrm{H}} U e_j = \delta_{ij}$, $1 \leq i, j \leq k$, or $U^{\mathrm{H}} U = I_k$. Here δ_{ij} is the Kronecker delta function. When $k = n$, matrix U is unitary if $U \in \mathbb{C}^{n\times n}$) or orthogonal if $U \in \mathbb{R}^{n\times n}$.

- Two matrices $X_{n\times k}$ and $Y_{n\times k}$ are bi-orthogonal in columns if $e_i^{\mathrm{T}} Y^{\mathrm{H}} X e_j = \delta_{ij}$, $1 \leq i, j \leq k$, i.e., $Y^{\mathrm{H}} X = I_k$. When $n = k$, $Y^{\mathrm{H}} = X^{-1}$.

1.2 Eigenvalue-eigenvector pairs

$$A x_j = \lambda_j x_j, \qquad x_j \neq 0, \quad x_j \in \mathbb{C}^{n\times 1}, \quad \lambda_j \in \mathbb{C}, \quad 1 \leq j \leq n \tag{1}$$

- The spectrum: $\Lambda(A) \triangleq \{\lambda_j\} \subset \mathbb{C}$

- The spectral radius: $\rho(A) \triangleq \max\{\, |\lambda_j| \,\}$

  Verify that $\rho(A) \leq \|A\|_p$ for any p-norm, including in particular $\|\cdot\|_1$ and $\|\cdot\|_\infty$, which can be computationally and easily obtained.

One can follow the definitions and statements item by item with ease:

☐ Any two eigenvectors associated with two distinct eigenvalues are linearly independents.

☐ Let p be the number of distinct eigenvalues. Then, $1 \leq p \leq n$, and the number of linearly independent eigenvectors is no less than p and no greater than n.

☐ $(A - \lambda_j I)x_j = 0$, $x_j \neq 0$

$\Longrightarrow$ when shifted by an eigenvalue λ_j, the matrix $A - \lambda_j I$ is singular

The geometric multiplicty of eigenvalue λ_j is the dimension of the null space of $A - \lambda_j I$, i.e., $m_{j,\mathrm{g}} \triangleq \dim(\mathrm{null}(A - \lambda_j I))$

☐ The characteristic polynomial: $\chi(\lambda) \triangleq \det(\lambda I - A)$

$\chi(\lambda)$ is of degree n and has n roots.

$\chi(\lambda_j) = \det(\lambda_j I - A) = 0.$

Every eigenvalye λ_j is a root of $\chi(\lambda)$; every root of $\chi(\lambda)$ is an eigenvalue of A

$\Longrightarrow \chi(\lambda) = \prod_{j=1:n}(\lambda - \lambda_j)$,

☐ The algebraic multiplicity of λ_j: A has p distinct eigenvalues, $1 \leq p \leq n$. Then, $\chi(\lambda) = \prod_{j=1:p}(\lambda - \lambda_j)^{m_j}$ where m_j is the algebraic multiplicity of λ_j, $\sum_j m_j = n$.

☐ The trace of A is the sum of the diagonal elements, which is readily obtainable. The trace is also equal to the sum of the eigenvalues,

$$\mathrm{trace}(A) \triangleq \sum_{i=1:n} A(i,i) = \sum_{j=1:p} m_j \lambda_j.$$

☐ Basic invariant properties of eigenvectors:

- Shifting: $(A - \mu I)x_j = (\lambda_j - \mu)x_j$, $\mu \in \mathbb{C}$

– Uniform scaling: $(\alpha A)x_j = (\alpha\lambda_j)x_j$, $\alpha \in \mathbb{C}$

– Polynomial of A: $A^k x_j = \lambda_j x_j \Longrightarrow p(A)x_j = p(\lambda_j)x_j$

☐ Basic invariant properties of eigenvalues:

- Transposition: $\Lambda(A^{\mathrm{T}}) = \Lambda(A)$

(Verification: $\det(\lambda_j I - A) = \det(\lambda_j I - A^{\mathrm{T}})$.)

- Transform T: matrix A undergoes a similarity transform

$\Longrightarrow \widehat{A} = TAT^{-1}$, and $\lambda(A) = \Lambda(\widehat{A})$

(Verification: $\widehat{A}(Tx_j) = \lambda_j Tx_j$.)

☐ Conjugation: $\bar{A}\bar{x}_j = \bar{\lambda}_j\bar{x}_j$,

1.3 Invariant subspaces

A subspace $\mathcal{X}$ is an invariant subspace of A if for any $x \in \mathcal{X}$, we have $A\,x \in \mathcal{X}$.

Any subset of eigenvectors of A, with the same eigenvalue or multiple distinct eigenvalues, spans an invariant subspace, and vice versa.

In particular, an eigenvector of A spans a one-dimensional invariant subspace associated with λ_j.

- All eigenvectors of A associated with a distinct eigenvalue λ_j span an invariant subspace with dimension equal to the geometric multiplicity $m_{j,\mathrm{g}}/$
- The dimension of the largest invariant subspace associated with λ_j is the algebraic multiplicity m_j of λ_j.
- One can get a unitary basis U for any invariant subspace, by an orthogonalization procedure, such as the Gram-Schmidt procedure.

1.4 Gershgorin discs & theorem

- Gershgorin discs:

  $\{D_i(A) = D(a_{ii}, r_i) \mid \text{center } a_{ii}, \text{radius } r_i = \sum_{j \neq i} |a_{ij}|,\ i = 1:n\}$

- Gershgorin theorem:

  Any eigenvalue λ of A lies in one of the Gershgorin disks: i.e., $|\lambda - a_{ii}| \leq r_i$ for some i.

  Proof sketch. For every λ, there exists $x \neq 0$ such that $Ax = \lambda x$. Let $i = \arg\max_j |x_j|$. Then, $(a_{ii} - \lambda) = -\sum_{j \neq i} a_{ij} x_j / x_k$.

- If the discs are disjoint, then each disc contains one and only eigenvalue. If the union of k discs is disjoint from the rest, then the union contains exactly k eigenvalues.

  Proof. A sketch. We introduce a simple and useful *morphing* and variation technique. Let $C(t) = tA + (1-t)D$, $t \in [0, 1]$, morphing between A and its diagonal matrix D. We have $C_{ii}(t) = a_{ii}$, i.e., the disc centers are t-invariant. The radii change as $t \cdot r_i$. If the discs of A are disjoint, so are those of $C(t)$, $t \in [0, 1)$. Every eigenvalue $\lambda(C(t))$ is a continuous functions of t. If an eigenvalue migrates from one disk of $C(0) = D$ to join another eigenvalue in a disk of $C(1) = A$, it must go across the strip dividing the departure disk and the arrival disk, contrary to the continuity of each and every eigenvalue. A similar argument supports the second statement. □

Some simple ways to tighten the bounds:

- Use the Gershgorin disks of A^{T} as well $|\lambda - a_{ii}| \leq \min\{r_i, c_i\}$, where $c_i = \sum_{j \neq i} |a_{j,i}|$ is the sum of column-i without a_{ii}. That is, $|\lambda - a_{ii}| \leq \min\{r_i, c_i\}$.

- Scaling, $D^{-1}AD$, $|\lambda - a_{ii}| \leq \min\{\sum_{j \neq i} |d_i^{-1} a_{ij} d_j|,\ \sum_{j \neq i} |d_j^{-1} a_{ji} d_i|\}$

2 Eigenvalue decomposition: diagonalizable

2.1 Spectral transform by the eigenvector matrix

A matrix is diagonalizable if it has a complete system of eigenvectors and is transformed into diagonal form by the eigenvector matrix. Let $X = [x_1, \cdots, x_n]$ be the eigenvector matrix. Then X is invertible.

It is straightforward to verify the following:

- ▷ A sufficient condition. If the eigvalues of A are all distinct from each other, then A is diagonalizable
- ▷ The necessary and sufficient condition: A is diagonalizable if and only if for every eigenvalue λ_j, $m_{j,\mathrm{g}} = m_j$, the geometric multiplicity is equal to the algebraic multiplicity
- ▷ $AX = X\Lambda$, where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$.
- ▷ EVD of A: $A = X\Lambda Y^{\mathrm{H}}$. $Y^{\mathrm{H}} \triangleq X^{-1}$.

  Remark: A is the product of three factors, the right eigenvector matrix, the left eigenvector matrix, and the diagonal eigenvalue matrix. We use the term "decomposition", instead of "factorization", to indicate that the decomposition process in general can not be done in a finite number of factorization steps (with arithmetic operations and simple root extraction operations), according the Abel's theorem.

- ▷ The left eigenvector matrix : $Y^{\mathrm{H}}A = \Lambda Y^{\mathrm{H}}$,. $Y^{\mathrm{H}} = X^{-1}$
- ▷ EVD of a matrix polynomial: $A^k = X\Lambda^k Y^{\mathrm{H}} \Longrightarrow p(A) = Xp(\Lambda)Y^{\mathrm{H}}$
- ▷ EVE of analytic matrix functions: e.g. $\exp(A) = X\exp(\Lambda)Y^{\mathrm{H}}$
- ▷ If A is nonsingular in addition, $A^{-k} = X\Lambda^{-k}Y^{\mathrm{H}}$, $k \geq 0$

2.2 Information propagation or modulation

An alternative view of the EVD is the expansion of A into additive spectral components:

$$A^k = \sum_{j=1}^{n} \lambda_j^k x_j y_j^{\mathrm{H}}, \qquad k \geq 1. \tag{2}$$

Every rank-1 term represents the spectral triple: (λ_j^k, x_j, y_j), only the eigenvalue changes with k.

Scale A so that $\rho(A) = 1$. Then,

$$\begin{aligned} A^k &= B^k + E^k \\ B^k &= \sum_{|\lambda_j|=1} \lambda_j^k \, x_j \, y_j^{\mathrm{H}}, \\ E^k &= \sum_{|\lambda_j|<1} \lambda_j^k \, x_j \, y_j^{\mathrm{H}} \to 0 \qquad \text{as } k \to \infty. \end{aligned} \tag{3}$$

where the eigenvalues are indexed from the largest (most dominant) in magnitude to the smallest (least dominant). REMARK: B^k is indeed the k-th power of B, not a notation. The same with E^k. By the spectral split, $B^j \times E^{j'} = 0$, $j, j \geq 1$.

- Matrix E^k converges to zero, $E^k \to 0$, the components with smaller eigenvalues in magnitude decay faster

- Matrix B^k remains of rank q, q is the multiplicity of λ_1

  If $\lambda_1 = 1$, then B^k converges to the rank-q matrix

  If $\lambda_1 = -1$, then the subsequence B^{2k} converges to the rank-q matrix

  if $\lambda_1 = e^{i2\pi/m}$, $m \in \mathbb{N}$, then the subsequence B^{mk} converges to the rank-q matrix, including the two previous cases

Diagonalizable matrices are further categorized as normal matrices or otherwise. The eigenvector matrix of a normal matrix is unitary or orthogonal. Two subclasses of normal matrices are Hermitian matrices and circulant (periodic) convolution matrices.

2.3 Symmetric eigenvalue decomposition

Let A be a Hermitian matrix, $A^{\mathrm{H}} = A$. If A is real-valued, it is symmetric $A^{\mathrm{T}} = A$. The adjacency matrix for an undirected graph is symmetric.

- ▷ The eigenvalues of A are real valued: $\lambda(A) \in \mathbb{R}$.

  Consequently, the eigenvalues can be put in non-descending (or not ascending order), $\lambda_j \leq \lambda_{j+1}$

- ▷ The eigenvectors associated with two different eigenvalues are orthogonal to each other. Recall that the invariant subspace associated with each eigenvalue has orthogonal basis.

- ▷ A simple proof can be found in the Schur decomposition, by inductive construction.

- ▷ EVD of A: $A = U\Lambda U^{\mathrm{H}}$, where U is unitary if A is complex-valued and orthogonal if A is real-valued

2.3.1 The Rayleigh quotient & Courant-Fischer theorem

$$R(x, A) \triangleq \frac{x^{\mathrm{H}} A x}{x^{\mathrm{Hx}} x}, \quad x \neq 0 \tag{4}$$

is a continuous variation embedding of the spectrum

$$\lambda_{\min} \leq R(x, A) \leq \lambda_{\max} \tag{5}$$

and that for any $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, there exist $x \neq 0$ such that $R(x, A) = \lambda$.

Suppose we index the eigenvalues as follows,

$$\lambda_{\min} = \lambda_1 \leq \lambda_1 \leq \cdots \lambda_{n-1} \leq \lambda_n = \lambda_{\max}$$

By the direct use of A's EVD, one can get the Courant-Fischer's theorem on min-max or max-min representation of the eigenvalues,

$$\begin{aligned}
\lambda_1 &= \min_{x \neq 0} R(x, A), \\
\lambda_j &= \min_{x^{\mathrm{H}}[x_1, \cdots, x_{j-1}] = 0} R(x, A), \quad j > 1 \\
&= \max_{x^{\mathrm{H}}[x_{j+1}, \cdots, x_n] = 0} R(x, A) \\
&= \min_{X : \dim(X) = j} \; \max_{x \in X} R(x, A) \\
&= \max_{X : \dim(X) = 1+n-j} \; \min_{x \in X} R(x, A)
\end{aligned} \tag{6}$$

3 Eigenvalue decomposition: general

We consider the eigenvalue decomposition (EVD) of a square matrix in general. In particular, the adjacency matrices for digraphs may not be diagonalizable. The EVD may take different forms and structures, including the Schur decomposition, the Jordan decomposition, and some in between.

In Schur decomposition, $A = QRQ^{\rm H}$, where Q is unitary, R is (upper) triangular. The diagonal elements of R are the eigenvalues of R and A, they may be in arbitrary order. We describe the Jordan decomposition, $A = X\Lambda_J X^{-1}$, where X is the Jordan vector matrix, and Λ_J is of the most compact triangular form, it is bi-diagonal. The eigenvalues of Λ_J are on the diagonal, clustered by their algebraic multiplicities.

3.1 The Schure decomposition

The Schur decomposition is $A = QRQ^{\rm T}$, where Q is unitary and R is upper triangular. The eigenvalues of A are on the diagonal of R. This decomposition is easy to prove by inductive construction. It is true for $n = 1$. In the induction step for $n > 1$, let $A_n = A$. We get an eigenpair of A_n and make a reduction in size via a similarity unitary transform. Let $(\lambda_1, q_1$ be an (arbitrary) eigenpair of A_n, with $q_n^{\rm H} q_n = 1$. Let H_n be an Householder matrix (unitary) with $H_n e_1 = q_1$. Then,

$$A_n = H_n \begin{pmatrix} \lambda_1 & r_1^{\rm T} \\ 0 & A_{n-1} \end{pmatrix} H_n^{\rm H} = H_n\,(1 \oplus Q_{n-1}) \begin{pmatrix} \lambda_1 & r_1^{\rm T} \\ 0 & R_{n-1} \end{pmatrix} H_n^{\rm H}\left(1 \oplus Q_{n-1}^{\rm H}\right) = Q_n\, R_n\, Q_n^{\rm H} \tag{7}$$

where by induction assumption that $A_{n-1} = Q_{n-1}R_{n-1}Q^{\rm n-1}$ with U_{n-1} being unitary and R_{n-1} being upper triangular.

- ☐ When A is normal, R is diagonal.

- ☐ When A is non-normal but diagonalizable, R is not diagonal. In detail, $A =$

$X\Lambda X^{-1}$. Let $X = QR_x$ be the QR factorization of X. Then, R_X is nonsingular, and $A = QRQ^{\mathrm{H}}$ with $R = (R_x \Lambda R_x^{-1})$ indicating the deviation of A from a normal matrix. In detail, when R is known, we have $RR_x = R_x\Lambda$, i.e., R_x is the eigenvector matrix of R when A is diagonalizable.

□ The computational procudre is fine grained.

3.2 Spectral analysis via the distinct invariant subspaces

In general, A has $p \leq n$ distinct eigenvalues. Every distinct eigenvalue λ_j is associated with an unique invariant subspace χ_j of dimension m_j, which is the algebraic multiplicity of λ_j. Let X_j be a particular set of the basis vectors. Then

$$A\,X_j = X_j A_j, \quad \sum_{j=1:p} m_j = n. \tag{8}$$

Let $X = [X_i, X_2, \cdots, X_p]$. Then, X is nonsigular and

$$A\,X = X\,\mathrm{diag}(A_j, j = 1:p) \tag{9}$$

That is, X transforms A into a block diagonal matrix.

The eigenvalues of A_j are equal to λ_j. Specifically, $\lambda_j = \mathrm{trace}(A_j)/m_j$. The eigenvalues of A_j may or may not be explicit as matrix elements. Here, A_j is diagonalizable if the geometric multiplicity of λ_j is equal to its algebraic multiplicity. When A_j is not diagonalizable, we have more detailed view into the structure of the invariant subspace.

3.3 The Schur basis

Let $A_j = U_j \widehat{R}_j U_j^{\mathrm{T}}$ be a Schur decomposition of A_j. Let $X_j U_j = Q_j R_{x,j}$ be a QR factorization of $X_j U_j$. Then, $A Q_j = Q_j R_j Q_j^{\mathrm{H}}$ with $R_j = R_{x,j} \widehat{R}_j R_{x,j}^{-1}$. We may refer Q_j as a Schur basis for the inviant subspace span(X_j).

Let $X = [Q_1, \cdots, Q_p]$. Then, X transforms A into a block diagonal matrix diag$(R_j, j = 1:p)$, with the upper triangular submatrices R_j on the diagonal. Matrix Q may or may not be unitary, although each block column Q_j has orthonormal columns.

3.4 Similarity reduction to the Jordan form

Every (upper) $R_j - \lambda_j I$ matrix can be transformed into a Jordan form.

3.5 The Jordan matrix

We introduce a non-diagonalizable matrix in a very simple form. The Jordan matrix J_n is the shift matrix, $J_n = [0, I_n(:, 1:n-1)]$. It is upper triangular with nonzero on the supper diagonal only. For example,

$$J_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad J(i', j') = \delta_{i'(j'-1)}$$

Matrix J_n has the zero eigenvalue of algebraic multiplicty n, $J_n^n = 0$. It has only one eigenvector, which is e_1, and therefore the geometric multiplicity of the eigenvalue is 1. It can be verified that $J_n^k [e_1. \cdots, e_k] = 0$, $1 \le k < n$, e_k is the k-th column of the identity matrix I.

3.6 The Jordan decomposition

The Jordan decomposition captures more of the connection and difference between diagonalizable and non-diagonalizable matrices.

Next, we consider the Jordan basis for each of the invariant subspaces.

3.7 The Jordan bases

With the Jordan basis for the invariant subspace associated with λ_j, A_j is of the most compact form. It is block diagonal, each block on the diagonal is a Jordan block. Specifically, a Jordan basis is composed of $q_j \geq 1$ subsets, $X_j = [X_{j_k} \mid k = 1 : q_j]$. Each subset X_{j_k} spans an invariatn subspace associated with λ_j,

$$AX_{j_k} = X_{j_k} \left(\lambda_j I_{j_k} + J_{j_k}\right), \qquad \sum_{k=1:q_j} j_k = m_j, \quad n = \sum_j m_j \tag{10}$$

where J_k is of the standardized Jordan form.

- □ For each Jordan subset X_{j_k}, only the leading vector is an eigenvector, the rest follow the Jordan chain,

$$Ax_1 - \lambda_j x_1 = 0, \qquad (A - \lambda_j)x_q = x_{q-1}, \qquad 1 < q \leq j_k \tag{11}$$

or,

$$(A - \lambda_j I)^q \left[x_1, \cdots, x_q\right] = 0, \qquad 1 \leq q \leq j_k. \tag{12}$$

The equations of (12) sugget a more robust routine for obtaining a basis for the invariant subspace associated with λ_j.

- □ For each distinct eigenvalue λ_j, its geometric multiplicity $m_{j,\mathrm{g}}$ is the number of the Jordan blocks associated with λ_j, or equivalently, the number of eigenvectors associated with λ_j, $1 \leq m_{j,\mathrm{g}} \leq m_j$

- ☐ The algebraic multiplicity m_j of eigenvalue λ_j is equal to the sum of the Jordan block sizes j_k, $m_j = \sum_{k=1:q_j} j_k$.
- ☐ With all Jordan vectors associated with a distinct eigenvalue λ_j,

$$AX_j = X_j\Lambda_j, \quad \Lambda_j \triangleq (\lambda_j I_{m_j} + \mathrm{diag}(J_{j_k})), \qquad j = 1, \cdots, p \tag{13}$$

- ☐ Matrix A is diagonalizable if and only if all Jordan blocks J_{j_l} are of size 1.

We give a simple proof of Jordan form at the end of the section.

3.8 The spectral transform by the Jordan matrix

The Jordan vector matrix is $X = [X_i \mid j = 1 : p]$, where $p \geq 1$ is the number of distinct eigenvalues λ_j. It transforms A into the bidiagonal form: $\Lambda_J = \mathrm{diag}(\Lambda_j, j = 1 : p)$. We have the following,

$$\begin{aligned}
& A\,X = X\,\Lambda_J \\
\Longrightarrow \quad & A = X\Lambda_J X^{-1} = X\Lambda_J Y^{\mathrm{H}}, \qquad Y^{\mathrm{H}} = X^{-1} \\
\Longrightarrow \quad & A^k = X\Lambda_J^k Y^{\mathrm{H}}, \qquad k \geq 1 \\
\Longrightarrow \quad & p(A) = Xp(\Lambda_J)Y^{\mathrm{H}} \\
\Longrightarrow \quad & \exp(A) = X\exp(\Lambda_J)Y^{\mathrm{H}}
\end{aligned} \tag{14}$$

3.8.1 Information propagation or modulation

Assume that $\rho(A) = 1$.

$$\begin{aligned}
A^k &= B^k + E^k \\
B^k &= \sum_{|\lambda_j|=1} X_j\Lambda_j^k\, Y_j^{\mathrm{H}} \\
E^k &= \sum_{|\lambda_j|<1} X_j\Lambda_j^k\, Y_j^{\mathrm{H}} \to 0 \qquad \text{as } k \to \infty.
\end{aligned} \tag{15}$$

If A is nonnegative, irreducible and aperiodic, then $\lambda_1 = 1$ is simple, and A^k converges to $x_1 y_1^{\mathrm{T}} > 0$.

4 Graph Laplacian spetral

4.1 The Laplacian matrices

☐ $G(V, E, W)$ undirected, nonnegative weights on edges

$n = |V|$, $m = |E|$, : $W : E \to \mathbb{R}_{\geq 0}$

☐ A: adjacency matrix, symmetric

$A(i, j) > 0 \iff (i, j) \in E$

$A^{\mathrm{T}} = A$, $d = A\, e$ (the degree vector)

☐ B: incidence matrix, $n \times m$

$B(:, \ell) = \pm(e_i - e_j)$ with $\ell = (i, j) \in E$

$e^{\mathrm{T}} B = 0$, $B^{\mathrm{T}} e = 0$

need a modification to admit self-loops $j = i$

☐ B_+: incidence matrix with different edge encoding

$B_+(:, \ell) = (e_i + e_j)$ with $\ell = (i, j) \in E$

This is used in some textbooks or literature articles

☐ D: $D = \mathrm{diag}(d)$

☐ L: the (plain) Laplacian matrix

- $L = B\, \mathrm{diag}(W)\, B^{\mathrm{T}}$
  It is the accumulation of edge Laplacians

$$L = \sum_{(i,j) \in E} (e_i - e_j)(e_i - e_j)^{\mathrm{T}} \tag{16}$$

It can be split in different ways

- $L = D - A$

  invariant to self-loops: $L = (D + D_o) - (A + D_o)$
- L is symmetric, non-negative definite
- EVD: $L = Q\Lambda Q^{\mathrm{T}}$

  a complete set of orthogonal eigenvectors with axial weights $\lambda_j \geq 0$

☐ With B_+, we may define $L_+ = B_+ B_+^{\mathrm{T}} = D + A$.

☐ Then, $D = (L_+ + L_{-1})/2$ and $A = (L_+ - L_{-1})/2$, where $L_- = D - A$.

☐ When G is d-regular (a constant degree d)

- $L = dI - A$
- $\lambda_j(A) = \lambda_j(L) - d$, a uniform shifting

Example of regular graphs: cliques, k-dimensional torus graphs, k-dimensional hypercube, buckyball (C_{60}), or constructed by a random graph generator with constant degrees

☐ The Rayleigh quotient with L

$$
\begin{aligned}
R(x \mid \widehat{L}) &= x^{\mathrm{T}} x, \quad x^{\mathrm{T}} x = 1 \\
&= \sum_{(i,j)\in E} x^{\mathrm{T}}(e_i - e_j)\cdot(e_i - e_j)^{\mathrm{T}} x \\
&= \sum_{(i,j)\in E} (x(i) - x(j))^2 \\
&= \frac{1}{2}\sum_{i\in V}\sum_{j\in\mathcal{N}(i)} (x(i) - x(j))^2 \\
R(x_1 \mid \widehat{L}) &= 0, \quad x_1 = e
\end{aligned}
\tag{17}
$$

The last two expressions: the variation of x over any edge, the variation of x over the neighborhood graph $\mathcal{N}(i)$ of any vertex

4.2 Invariance subspaces

4.2.1 The null space

- $B^{\mathrm{T}}e = 0 \to Le = 0$ and $0 \in \Lambda(L)$
- The null space of L is one dimensional $\iff$ G is connected $\iff$ $\lambda_2 > 0$
- G has exactly p connected components $\iff$ $p = \dim(\mathrm{null}(L))$

4.2.2 The Fiedler value & subspace

Assume G is connected, i.e., $\lambda_2(L) > 0$.

- The Rayleigh quotient with L
- The Fiedler value is the least variation among all vertex functions, $x : V \to \mathbb{R}$ subject to the zero sum and the unit length,

$$\begin{aligned} \lambda_2 = \min_x x^{\mathrm{T}} L x = \sum_{(i,j)\in E} (x(i) - x(j))^2, \\ \text{s.t. } x^{\mathrm{T}} e = 0, \quad x^{\mathrm{T}} x = 1, \end{aligned} \tag{18}$$

The Fiedler subspace is the space spanned by all Fiedler vectors (the minima):

$$\begin{aligned} x_2 = \arg \min_{\substack{x^{\mathrm{T}} e = 0 \\ x^{\mathrm{T}} x = 1}} x^{\mathrm{T}} L x \\ \text{s.t. } x^{\mathrm{T}} e = 0, \quad x^{\mathrm{T}} x = 1, \end{aligned} \tag{19}$$

- Edge partition and vertex partition by a Fiedler vector

$$
\begin{aligned}
x^{\mathrm{T}}Lx = & \sum_{\substack{(i,j)\in E \\ i,j\in V_+}} (x(i)-x(j))^2 \\
& + \sum_{\substack{(i,j)\in E \\ i,j\in V_-}} (x(i)-x(j))^2 \\
& + \sum_{\substack{(i,j)\in E \\ x(i)\cdot x(j)<0}} (x(i)-x(j))^2
\end{aligned}
\tag{20}
$$

A simple construction for mental experiment (not to blind to an optimization process): Let G be a graph of two cliques K_{n_1} and K_{n_2} (or two cycles C_{n_1} and C_{n_2}) connected by a single edge, $n = n_1 + n_2$, all edge unweighted. Each subgraph is a community. Consider two particular cases: case (i) $n_1 = n_2 = n/2$; case (ii) $n_1 = n - 2$, $n_2 = 2$.

- Why the values of x_2 at the border/boundart nodes of V_+ and V_- are closer to the zero ?
- how $x(i)$ is affected by the difference between $|V_+|$ and $|V_-|$?
- how many zero elements in x_2?

∘ The Fiedler partion is unique $\iff$ $\lambda_2 < \lambda_3$

∘ If G is the hypercube of k-dimensions, then $n = 2^k$, $\lambda_2(L) = 2$, with multiplicty k.

∘ The extension to the weighted case is straightforward.

4.2.3 The normalized Laplacian

The normalized Laplacian is related to the probabilistic random-walk transition matrix and the (geometrix) similarity matrix of the local neighborhoods.

▷ $\widehat{B}$:

$$\widehat{B} = D_v^{-1/2} B D_e^{1/2} \tag{21}$$

where $D_e = \text{diag}(W)$, with the edge weights on the diagonal. Each row of $\widehat{B}$ is normalized to the unit length.

By the same scalings, we have $\widehat{B}_+$.

▷ The normalized Laplacian:

$$\begin{aligned} \widehat{L} &\triangleq \widehat{B}\widehat{B}^{\mathrm{T}} \\ &= D^{-1/2} L D^{-1/2} \\ &= I - \widehat{A}, \quad \widehat{A} = D^{-1/2} A D^{-1/2} \end{aligned} \tag{22}$$

▷ Similarly, we have $\widehat{L}_+$, which is elementwise nonnegative and spectrally nonnegative definite.

We recognize that $\widehat{L}_+(i,j)$ is a geometric measure of the similarity between the neighborhoods $\mathcal{N}(i)$ and $\mathcal{N}(j)$. Equivalently, $ee^{\mathrm{T}} - \widehat{L}_+ = ee^{\mathrm{T}} - \widehat{A}$ is the matrix of pairwise distances among all neighborhoods.

This property makes the following simple connection important:

$$\widehat{L}_+ = 2I - \widehat{L} \tag{23}$$

The relationship is a negative flip followed by a shift. When we know one of them, we know the other.

▷ The spectral connection to the random-walk transition matrix

$A_w \sim \widehat{A}$, where $A_w = AD^{-1}$ and $\sim$ denotes the similar transform relationship/

$\lambda_j(A_w) = \lambda_j(\widehat{A}) = \lambda_j \in [-1, 1], \qquad j = 1 : n$

$\lambda_j(\widehat{L}) = 1 - \lambda_j(\widehat{A}) \in [0, 2]$

The Perron value of A_w is mapped to the smallext eigenvalue of $\widehat{L}$

$\lambda_j(\widehat{L}_+) = 1 + \lambda_j(\widehat{A}) \in [0, 2]$.

The Perron value of A_w (equal to 1) is shifted to the Perron value of $\widehat{L}_+$ (equal to 2); the smallest eigenvalues of $\widehat{L}_+$ are the largest eigenvalue of $\widehat{L}_+$.

EVDs:

$$\widehat{A} = Q\Lambda Q^{\mathrm{T}}, \qquad A_w = X\Lambda X^{-1}, \quad X = D^{1/2}Q$$

When G is regular with a constant degree d, $A_w = A/d$, a uniform scaling.

Variational analysis of the normalized Laplacians

The Rayleigh quotient (uniform wedge weights)

$$\begin{aligned} R(x \mid \widehat{L}) &= x^{\mathrm{T}} D^{-1/2} L D^{-1/2} x, \quad x^{\mathrm{T}} x = 1 \\ &= \sum_{(i,j)\in E} x^{\mathrm{T}} D^{-1/2}(e_i - e_j)\cdot(e_i - e_j)^{\mathrm{T}} D^{-1/2} x \\ &= \sum_{(i,j)\in E} \left(\frac{x(i)}{\sqrt{d(i)}} - \frac{x(j)}{\sqrt{d(j)}} \right)^2 \\ &= \frac{1}{2}\sum_{i\in V} \frac{1}{d(i)} \sum_{j\in\mathcal{N}(i)} \left(x(i) - \sqrt{\frac{d(i)}{d(j)}}\, x(j) \right)^2 \\ R(x_1 \mid \widehat{L}) &= 0, \quad x_1 = d^{1/2} \end{aligned} \tag{24}$$

The relationship between low-degree nodes and high-degree neighbors is changed by the vertex scaling.

It is straightforward to get the generalized expression of the Rayleigh quotient with ununiform edge weights.

The normalized cut

The Fiedler value is the least variation among all vertex functions, $x : V \to \mathbb{R}$ subject to the orthogonality to x_1 and the unit length

$$\begin{aligned} \lambda_2 = \min_y x^{\mathrm{T}} \widehat{L} x \\ \text{s.t. } x^{\mathrm{T}} x_1 = 0, \quad x^{\mathrm{T}} x = 1 \end{aligned} \tag{25}$$

The Fiedler subspace is the space spanned by all Fiedler vectors (the minima):

$$\begin{aligned} x_2 = \arg\min_y x^{\mathrm{T}} \widehat{L} x \\ \text{s.t. } x^{\mathrm{T}} x_1 = 0, \quad x^{\mathrm{T}} x = 1 \end{aligned} \tag{26}$$

Edge partition and vertex partition by a Fielder vector of the normalized Laplacian is similar to that without normalization.

4.2.4 Laplacian spectral embedding

A Laplacian spectral embedding serve a few objectives

☐ vertex embedding: every vertex gets an encded vector in a geometric space (for down-stream tasks)

☐ relatively low dimension, in comparison to the vertex size

☐ relatively high accuracy in maintaining the pairwise neighborhood similarities, i.e., accurate reconstruction of $\widehat{L}_+$
(an unsupervised approach)

The basic approach is simple:

$$
\begin{aligned}
\widehat{L}_{+} &= Q_{\mathrm{k}}\Lambda_k Q_k + Q_{\mathrm{n-k}}\Lambda_{n-k}Q_{n-k} \\
\widehat{L}_{+,k} &= Q_{\mathrm{k}}\Lambda_k Q_k
\end{aligned}
\tag{27}
$$

where Λ_k is composed of the largest k eigenvalues of $\widehat{L}_{+}$, related to the smallest k eigenvalues of $\widehat{L}$

With a spectral approximation $\widehat{L}_{+,k}$,

- the code for vertex i is $Q(i, 1:k)\sqrt{\Lambda_k}$
- the embedding dimension, i.e., the code length, is k
- The residual error in the approximate reconstruction is $R_k = Q_{\mathrm{n-k}}\Lambda_{n-k}Q_{n-k}$ with $\|R_k\|_2 = \max(\Lambda_{n-k})$ and $\|R_k\|_F = \|\Lambda_{n-k}\|_F$

Algorithm prototype

Let τ be a specified threshold on residual error

- compute $\mu = \|\widehat{L}_{+}\|_F^2$
- initialization: $\Lambda = \emptyset$, $Q = \emptyset$, $\Lambda = 0$, $k = 0$
- while $(\mu - \gamma)/\mu > \tau$
  - advance the index: $k \mathrel{+}= 1$
  - compute the next eigenpair: (λ_k, q_k)
    subject to the conditions: $q_k^{\mathrm{T}} Q = 0$, $q_k^{\mathrm{T}} q_k = 1$
  - update: $\Lambda \mathrel{+}= \lambda_k$; $Q \mathrel{+}= q_k$; $\Gamma \mathrel{+}= \lambda_k^2$;
- return: k, $Q_k := Q$, $\Lambda_k := \Lambda$

4.2.5 Laplacians of bipartites & digraphs

There are ongoing efforts to extend Laplacians to directed graphs (digraphs). We exclude any symmetrization method that ignores edge orientation in a digraph.

In this section, we introduce a simple way with regard to some important downstream graph analysis tasks. A digraph $G(V, E)$ may be viewed as a bipartite $G(V_s, V_t, E)$, $E \in V_s \times V_t$. Every vertex $v \in V$ corresponds to a source node v_s in V_s and a target node v_t in V_t.

The incidence matrix:

$$B_{\mathrm{ts}}(:, \ell) = \pm(e_{it} - e_{js}), \quad \ell = (it, js), \quad it \in V_t, \quad js \in V_s \tag{28}$$

The adjacency matrix:

$$A_{\mathrm{ts}} \triangleq \begin{array}{c} \text{target} \\ \text{source} \end{array} \begin{pmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{pmatrix}, \qquad d_{\mathrm{ts}} = A_{\mathrm{ts}} e = \begin{pmatrix} d_{\mathrm{in}} \\ d_{\mathrm{out}} \end{pmatrix} \tag{29}$$

The plain Laplacian:

$$L_{\mathrm{ts}} = B_{\mathrm{ts}} B_{\mathrm{ts}}^{\mathrm{T}} = \begin{pmatrix} D_{\mathrm{in}} & -A \\ -A^{\mathrm{T}} & D_{\mathrm{out}} \end{pmatrix}, \qquad L_{\mathrm{ts}} e = 0 \tag{30}$$

The normalized Laplacian:

$$\begin{aligned} \widehat{L}_{\mathrm{ts}} &= \widehat{B}_{\mathrm{ts}} \widehat{B}_{\mathrm{ts}}^{\mathrm{T}} = I - \widehat{A}_{\mathrm{ts}} \\ \widehat{A}_{\mathrm{ts}} &= \begin{pmatrix} 0 & D_{\mathrm{in}}^{-1/2} A D_{\mathrm{out}}^{-1/2} \\ \left(D_{\mathrm{in}}^{-1/2} A D_{\mathrm{out}}^{-1/2}\right)^{\mathrm{T}} & 0 \end{pmatrix} \\ \widehat{L}_{\mathrm{ts}} &\begin{pmatrix} d_{\mathrm{in}}^{1/2} \\ d_{\mathrm{out}}^{1/2} \end{pmatrix} = 0 \end{aligned} \tag{31}$$