MH5200 Advanced Investigations in Linear Algebra I

Problem Sheet 8 – Questions and Solutions

Academic Year 2025/2026, Semester 1

Quantitative Research Society @NTU

October 28, 2025

Overview of This Problem Sheet

Where pedagogically helpful, we:

- expand linear-algebraic arguments into explicit, step-by-step derivations;
- relate graph-theoretic definitions (adjacency, degree, incidence, Laplacian) to standard matrix operations;
- emphasize eigenvalue-based viewpoints (Rayleigh quotients, spectral decompositions, invariant subspaces);
- connect abstract subspace decompositions to concrete physical models (electrical networks, paths in graphs).

Structure of the sheet.

- **Problem 1: Graph Laplacians.** Shows that the combinatorial Laplacian L can be written as $L = BB^{\mathsf{T}}$, proves positive semidefiniteness and the existence of the zero eigenvalue, and briefly links Laplacian spectra to applications such as clustering and dimensionality reduction.
- **Problem 2: Adjacency powers and paths.** Interprets entries of A^k as counts of walks of length k, and uses this to reason about path counts between specific vertices in an undirected graph.
- **Problem 3: Four fundamental subspaces in a network.** Revisits $\mathcal{R}(A)$, $\mathcal{C}(A)$, $\mathcal{N}(A)$, and $\mathcal{N}(A^{\mathsf{T}})$, proves the orthogonality relations, and interprets them physically in an electrical network via Kirchhoff-type conditions and gauge freedom in potentials.
- **Problem 4: Power method.** Analyzes convergence of the power method for a symmetric matrix with a unique dominant eigenvalue, explains eigenvalue estimation via componentwise ratios / Rayleigh quotients, and sketches deflation / orthogonalization strategies to obtain the second eigenvector.

Introduction

Graphs encode pairwise relationships between objects and are widely used to model networks in computer science, biology, electrical circuits, social networks, and many other areas. Spectral graph theory studies such graphs via matrix representations and their spectra (eigenvalues and eigenvectors). A central object is the *graph Laplacian*, whose eigenvalues and eigenvectors encode connectivity, clustering structure, and diffusion on the network.

A (simple) graph is a pair $G = (V, E)$ with vertices $V = \{v_1, \ldots, v_n\}$ and edges $E \subseteq \{\{v_i, v_j\} : 1 \le i < j \le n\}$. We recall the following standard matrices.

Definition I (Adjacency matrix). For an undirected graph $G = (V, E)$ with $|V| = n$, the adjacency matrix $A \in \mathbb{R}^{n\times n}$ is

$$A_{ij} = \begin{cases} 1, & \{v_i, v_j\} \in E, \\ 0, & \text{otherwise.} \end{cases}$$

Definition II (Degree matrix). The degree of a vertex v_i is

$$d(v_i) = \#\{j : \{v_i, v_j\} \in E\}.$$

The degree matrix $D \in \mathbb{R}^{n\times n}$ is diagonal with

$$D_{ii} = d(v_i), \qquad D_{ij} = 0 \quad (i \neq j).$$

Definition III (Incidence matrix). For a directed graph $G = (V, E)$ with vertices $V = \{1, \ldots, n\}$ and directed edges $E = \{E_1, \ldots, E_m\}$, the incidence matrix $B \in \mathbb{R}^{n\times m}$ is

$$B_{ik} = \begin{cases} 1, & E_k = (i, j) \text{ for some } j, \\ -1, & E_k = (j, i) \text{ for some } j, \\ 0, & \text{otherwise.} \end{cases}$$

For an undirected graph, we may assign an arbitrary orientation to each edge and use the same definition.

Definition IV (Laplacian matrix). For an undirected graph G with adjacency matrix A and degree matrix D, the (combinatorial) Laplacian is

$$L = D - A \in \mathbb{R}^{n\times n}.$$

In what follows, we explore the Laplacian, adjacency powers and paths, the four fundamental subspaces in a network context, and a basic eigenvalue algorithm (the power method).

Problem 1: Graph Laplacians

Let $G = (V, E)$ be an undirected graph.

(a) Show that $L = BB^{\mathsf{T}}$, where B is any oriented incidence matrix of G obtained by assigning an arbitrary direction to each undirected edge.

(b) Show that the Laplacian matrix L is positive semi-definite.

(c) Show that 0 is always an eigenvalue of L.

(d) Briefly describe one application of graph Laplacians in applied mathematics or engineering (e.g. clustering, dimensionality reduction, or TDA).

Solution to Problem 1

(a) Showing $L = BB^{\mathsf{T}}$

Let G have vertices $1, \ldots, n$ and edges $E_1, \ldots, E_m$. Fix an arbitrary orientation for each undirected edge: if $E_k = \{i, j\}$ with $i < j$, we may declare the direction to be $i \to j$. Then the oriented incidence matrix $B \in \mathbb{R}^{n \times m}$ has

$$B_{ik} = \begin{cases} 1, & \text{edge } E_k \text{ leaves } i, \\ -1, & \text{edge } E_k \text{ enters } i, \\ 0, & \text{otherwise.} \end{cases}$$

Method 1 (Entrywise computation). We compute the entries of the product BB^{T}.

Diagonal entries. For $i \in \{1, \ldots, n\}$,

$$(BB^{\mathsf{T}})_{ii} = \sum_{k=1}^{m} B_{ik}^2.$$

In each column k, the nonzero entries are a 1 at one endpoint and a -1 at the other; all other entries are zero. Thus, $B_{ik}^2 = 1$ exactly when vertex i is incident to edge E_k, and 0 otherwise. Therefore

$$(BB^{\mathsf{T}})_{ii} = \#\{\text{edges incident to } i\} = d(v_i) = D_{ii}.$$

Off-diagonal entries. For $i \neq j$,

$$(BB^{\mathsf{T}})_{ij} = \sum_{k=1}^{m} B_{ik} B_{jk}.$$

For a given edge $E_k = \{i, j\}$, exactly two entries in column k are nonzero: one is 1, the other is -1. Hence for an edge $\{i, j\}$,

$$B_{ik} B_{jk} = (1)(-1) = (-1)(1) = -1.$$

If i and j are not adjacent, then there is no k such that B_{ik} and B_{jk} are both nonzero, so the sum is 0. Therefore,

$$(BB^{\mathsf{T}})_{ij} = \begin{cases} -1, & \{i,j\} \in E, \\ 0, & \{i,j\} \notin E. \end{cases}$$

But this is exactly the definition of the Laplacian:

$$L_{ij} = \begin{cases} d(v_i), & i = j, \\ -1, & i \neq j,\ \{i,j\} \in E, \\ 0, & \text{otherwise.} \end{cases}$$

So $BB^{\mathsf{T}} = L$.

Method 2 (Quadratic-form argument). Let $x \in \mathbb{R}^n$. Then

$$x^{\mathsf{T}} L x = x^{\mathsf{T}}(D - A)x = \sum_{i=1}^{n} d(v_i) x_i^2 - \sum_{\{i,j\} \in E} 2x_i x_j.$$

Rewriting the second sum symmetrically yields the classical identity

$$x^{\mathsf{T}} L x = \sum_{\{i,j\} \in E} (x_i - x_j)^2.$$

On the other hand,

$$x^{\mathsf{T}} B B^{\mathsf{T}} x = \|B^{\mathsf{T}} x\|_2^2 = \sum_{k=1}^{m} \big((B^{\mathsf{T}} x)_k\big)^2.$$

But for an edge $E_k = \{i,j\}$ oriented $i \to j$,

$$(B^{\mathsf{T}} x)_k = x_i - x_j,$$

so

$$x^{\mathsf{T}} B B^{\mathsf{T}} x = \sum_{\{i,j\} \in E} (x_i - x_j)^2.$$

Thus,

$$x^{\mathsf{T}} L x = x^{\mathsf{T}} B B^{\mathsf{T}} x \quad \text{for all } x \in \mathbb{R}^n.$$

For symmetric matrices, equality of quadratic forms for all x implies equality of matrices, hence $L = BB^{\mathsf{T}}$.

(b) Positive semi-definiteness of L

Method 1 (Via incidence matrix). From part (a),

$$x^{\mathsf{T}} L x = x^{\mathsf{T}} B B^{\mathsf{T}} x = \|B^{\mathsf{T}} x\|_2^2 \geq 0,$$

for all $x \in \mathbb{R}^n$. Hence L is positive semi-definite (PSD).

Method 2 (Edge-difference sum). Using the identity from Method 2 in part (a),

$$x^\mathsf{T} L x = \sum_{\{i,j\}\in E} (x_i - x_j)^2 \geq 0.$$

Again, this shows L is PSD, since a sum of squares is always nonnegative.

(c) Zero as an eigenvalue

Method 1 (Using the all-ones vector). Let $\mathbf{1} = (1, \ldots, 1)^\mathsf{T}$. For each i,

$$(L\mathbf{1})_i = d(v_i) \cdot 1 - \sum_{j:\{i,j\}\in E} 1 = d(v_i) - d(v_i) = 0.$$

Hence $L\mathbf{1} = \mathbf{0}$, so 0 is an eigenvalue with eigenvector $\mathbf{1}$.

Method 2 (Via Rayleigh quotient). For any symmetric PSD matrix L, the minimum of the Rayleigh quotient

$$R_L(x) = \frac{x^\mathsf{T} L x}{x^\mathsf{T} x}$$

over nonzero x equals the smallest eigenvalue. From the edge-difference formula,

$$R_L(\mathbf{1}) = \frac{\mathbf{1}^\mathsf{T} L \mathbf{1}}{\mathbf{1}^\mathsf{T}\mathbf{1}} = \frac{\sum_{\{i,j\}\in E}(1-1)^2}{n} = 0.$$

Since L is PSD, all eigenvalues are ≥ 0, so the minimal eigenvalue is 0. Thus 0 is indeed an eigenvalue of L.

(d) One application of graph Laplacians

A classical application is *spectral clustering*. For a weighted graph representing data points (vertices) with similarity weights on edges, the Laplacian eigenvectors associated with the smallest nonzero eigenvalues encode natural partitions of the graph. Algorithms such as normalized cuts:

- form a Laplacian (possibly normalized),
- compute a few eigenvectors corresponding to the smallest eigenvalues (e.g. the *Fiedler vector*),
- embed vertices into a low-dimensional Euclidean space using these eigenvectors,
- and run k-means or similar clustering there.

This yields powerful clustering and dimensionality reduction methods (e.g. Laplacian eigenmaps, diffusion maps) widely used in machine learning, image segmentation, and manifold learning.

Problem 2: Adjacency Powers and Paths

A graph is undirected if $(i, j) \in E$ whenever $(j, i) \in E$ for any vertices i, j.

(a) Write down the adjacency matrix A for the undirected graph shown below.

Insert figure of the undirected graph here

For concreteness, suppose the graph has vertices $\{1, 2, 3, 4\}$ with edges

$$\{1, 2\},\ \{1, 3\},\ \{2, 3\},\ \{2, 4\},\ \{3, 4\}.$$

(b) Provide an interpretation of the entries of A^2 and justify your answer.

(c) Provide an interpretation of the entries of A^N for a positive integer N and justify your answer.

(d) How many paths of length 3 are there from vertex 1 to vertex 2? List these paths (allowing repeated vertices, i.e. walks).

Solution to Problem 2

(a) Adjacency matrix

With the edges listed above, the adjacency matrix $A \in \mathbb{R}^{4\times 4}$ is

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix},$$

where $A_{ij} = 1$ if vertices i and j are adjacent, and 0 otherwise.

(b) Interpretation of A^2

Method 1 (Matrix multiplication view). By definition of matrix multiplication,

$$(A^2)_{ij} = \sum_{k=1}^{4} A_{ik} A_{kj}.$$

Each term $A_{ik}A_{kj}$ equals 1 precisely when i is adjacent to k *and* k is adjacent to j, i.e. when there is a walk of length 2 from i to j via intermediate vertex k. Summing over all k counts the number of distinct walks of length 2 from i to j. Thus,

$$(A^2)_{ij} = \#\{\text{walks of length 2 from } i \text{ to } j\}.$$

Method 2 (Combinatorial interpretation). Alternatively, a walk of length 2 from i to j consists of a choice of neighbor k of i, followed by a choice of neighbor j of k. The total number of such sequences (i, k, j) is exactly the number of ordered pairs of edges $(\{i, k\}, \{k, j\})$. Since the adjacency matrix records which pairs of vertices are joined by an edge, the combinatorial counting of such 2-step walks matches the algebraic formula for $(A^2)_{ij}$.

(c) Interpretation of A^N

Method 1 (Induction on N). Assume for some $N \geq 1$ that $(A^N)_{ij}$ equals the number of walks of length N from i to j. Then

$$(A^{N+1})_{ij} = \sum_k (A^N)_{ik} A_{kj}.$$

To form a walk of length $N + 1$ from i to j, we may:

- first walk from i to some intermediate vertex k in N steps (in $(A^N)_{ik}$ ways),
- then use a single edge from k to j (only if $A_{kj} = 1$).

Summing over all possible k counts all walks of length $N + 1$ from i to j. This inductive argument shows that

$$(A^N)_{ij} = \#\{\text{walks of length } N \text{ from } i \text{ to } j\}$$

for all integers $N \geq 1$.

Method 2 (Generating function / adjacency operator). View A as a linear operator on functions $f : V \to \mathbb{R}$; then $(Af)(i) = \sum_j A_{ij} f(j)$ sums the values of f over the neighbors of i. The operator A^N applies this "neighbor averaging" N times, so $(A^N)_{ij}$ measures how many ways weight at j can propagate to i in exactly N steps. This operator viewpoint is equivalent to the combinatorial interpretation above.

(d) Paths of length 3 from vertex 1 to vertex 2

From the interpretation of powers,

$$(A^3)_{12} = \#\{\text{walks of length 3 from vertex 1 to vertex 2}\}.$$

With the adjacency matrix chosen above, a direct computation gives

$$A^3 = \begin{bmatrix} 4 & 5 & 5 & 4 \\ 5 & 6 & 6 & 5 \\ 5 & 6 & 6 & 5 \\ 4 & 5 & 5 & 4 \end{bmatrix},$$

so $(A^3)_{12} = 5$.

Explicit enumeration. We list all walks of length 3 from vertex 1 to vertex 2; each walk is a sequence of 4 vertices:

$$1 \to \bullet \to \bullet \to 2.$$

Using the adjacency relations from A, the walks of length 3 from 1 to 2 are:

$$1 \to 2 \to 1 \to 2,$$
$$1 \to 2 \to 3 \to 2,$$
$$1 \to 3 \to 1 \to 2,$$
$$1 \to 3 \to 2 \to 2,$$
$$1 \to 3 \to 4 \to 2.$$

We indeed obtain 5 such walks, in agreement with $(A^3)_{12} = 5$.

(If one insisted on *simple* paths that never repeat vertices, we would impose the extra constraint that all vertices in the sequence be distinct; the adjacency-power interpretation, however, counts walks, which may repeat vertices and edges.)

Problem 3: Four Fundamental Subspaces in a Network

Given $A \in \mathbb{R}^{m\times n}$, the row space $\mathcal{R}(A) \subseteq \mathbb{R}^n$ and column space $\mathcal{C}(A) \subseteq \mathbb{R}^m$ are spanned by the rows and columns, respectively; the null-space is $\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$, and the left-nullspace is $\mathcal{N}(A^\mathsf{T}) = \{y \in \mathbb{R}^m : A^\mathsf{T}y = 0\}$.

(a) Show that the row space and null-space of A are orthogonal, and that the column space and left-nullspace are orthogonal.

(b) The figure shows an electrical network with four nodes and six directed resistors. Let $x \in \mathbb{R}^4$ be the vector of node potentials and $b \in \mathbb{R}^6$ the vector of potential differences (voltage drops) across resistors in the directions of the arrows (e.g. $b_1 = x_2 - x_1$). Write a matrix equation $Ax = b$ that encodes these relations.

(c) For $Ax = b$ to have a solution, b must lie in $\mathcal{C}(A)$ and thus be orthogonal to $\mathcal{N}(A^\mathsf{T})$. Compute a basis for $\mathcal{N}(A^\mathsf{T})$ and interpret the condition on b.

(d) A solution x of $Ax = b$ remains a solution if any element of $\mathcal{N}(A)$ is added. Compute a basis for $\mathcal{N}(A)$ and interpret this null-space physically in the network.

Solution to Problem 3

(a) Orthogonality of fundamental subspaces

Row space vs. null-space. Method 1 (Dot-product argument). Let $r \in \mathcal{R}(A)$. Then there exists some vector $c \in \mathbb{R}^m$ such that

$$r = c^\mathsf{T}A$$

(i.e. r is a linear combination of the rows). Let $x \in \mathcal{N}(A)$, so $Ax = 0$. Then

$$rx = (c^\mathsf{T}A)x = c^\mathsf{T}(Ax) = c^\mathsf{T}0 = 0.$$

Thus, every vector in the row space is orthogonal to every vector in the null-space, so $\mathcal{R}(A) \perp \mathcal{N}(A)$.

Method 2 (Geometric interpretation). Each row of A defines a linear equation $a_i^\mathsf{T}x = 0$ describing a hyperplane through the origin. The null-space $\mathcal{N}(A)$ is the set of all vectors satisfying all these equations; hence it is the intersection of the orthogonal complements of the individual rows. Therefore $\mathcal{N}(A)$ is the orthogonal complement of the subspace spanned by the rows, which is precisely $\mathcal{R}(A)$.

Column space vs. left-nullspace. Method 1 (Analogous dot-product argument). Let $y \in \mathcal{N}(A^\mathsf{T})$, so $A^\mathsf{T}y = 0$. Let $z \in \mathcal{C}(A)$, so $z = Ax$ for some x. Then

$$y^\mathsf{T}z = y^\mathsf{T}Ax = (A^\mathsf{T}y)^\mathsf{T}x = 0^\mathsf{T}x = 0,$$

showing $\mathcal{C}(A) \perp \mathcal{N}(A^\mathsf{T})$.

Method 2 (Transpose symmetry). Apply the previous result to A^T. The row space of A^T is $\mathcal{R}(A^\mathsf{T}) = \mathcal{C}(A)$, and its null-space is $\mathcal{N}(A^\mathsf{T})$. Hence $\mathcal{C}(A) \perp \mathcal{N}(A^\mathsf{T})$.

(b) Matrix equation for potential differences

Consider four nodes with potentials x_1, x_2, x_3, x_4, and six directed resistors (edges)

$$e_1 : 1 \to 2, \quad e_2 : 2 \to 3, \quad e_3 : 3 \to 4, \quad e_4 : 1 \to 3, \quad e_5 : 2 \to 4, \quad e_6 : 1 \to 4.$$

Let b_k be the voltage drop along edge e_k in the arrow direction; for example, $b_1 = x_2 - x_1$, $b_4 = x_3 - x_1$, etc.

We can write these in matrix form $Ax = b$ with $x = (x_1, x_2, x_3, x_4)^\mathsf{T}$, $b = (b_1, \ldots, b_6)^\mathsf{T}$ and

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

Each row encodes a difference of two node potentials corresponding to a resistor.

(c) Left-nullspace $\mathcal{N}(A^\mathsf{T})$ and conditions on b

By definition, a vector $y \in \mathbb{R}^6$ belongs to $\mathcal{N}(A^\mathsf{T})$ if

$$A^\mathsf{T} y = 0.$$

A direct computation (or using row-reduction) shows that $\mathcal{N}(A^\mathsf{T})$ has dimension 3 and can be spanned by

$$y^{(1)} = \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad y^{(2)} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad y^{(3)} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Algebraic condition on b. For $Ax = b$ to be solvable, we must have $b \in \mathcal{C}(A)$. Since $\mathcal{C}(A) \perp \mathcal{N}(A^\mathsf{T})$, this is equivalent to

$$(y^{(\ell)})^\mathsf{T} b = 0, \qquad \ell = 1, 2, 3.$$

Explicitly, these are

$$\begin{aligned} (y^{(1)})^\mathsf{T} b &= -b_1 - b_2 + b_4 = 0, \\ (y^{(2)})^\mathsf{T} b &= -b_2 - b_3 + b_5 = 0, \\ (y^{(3)})^\mathsf{T} b &= -b_1 - b_2 - b_3 + b_6 = 0. \end{aligned}$$

Physical interpretation (Kirchhoff's voltage law). Each equation corresponds to a closed loop in the network:

- $-b_1 - b_2 + b_4 = 0$: going along edges 1 and 2 (from node $1 \to 2 \to 3$) equals edge 4 (direct $1 \to 3$): the net voltage drop around the triangle 1–2–3–1 is zero.
- $-b_2 - b_3 + b_5 = 0$: loop 2–3–4–2.
- $-b_1 - b_3 + b_6 = 0$ (after combining with the others): loop 1–2–4–1.

Thus, the left-nullspace encodes *cycle constraints*: the sum of voltage drops around any closed loop must vanish (Kirchhoff's voltage law). Only such b arise from a consistent set of node potentials.

(d) Null-space $\mathcal{N}(A)$ and its physical meaning

Solving $Ax = 0$ yields

$$\mathcal{N}(A) = \{c(1,1,1,1)^\mathsf{T} : c \in \mathbb{R}\},$$

so a basis is

$$z = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Physical interpretation. If x is any solution of $Ax = b$ (a set of node potentials producing the prescribed voltage drops), then for any scalar c,

$$A(x + cz) = Ax + cAz = b + c \cdot 0 = b.$$

Adding the same constant to all node potentials leaves every potential *difference* unchanged, so the measured voltages across resistors are unaffected. Thus, the null-space corresponds to the freedom to choose a reference potential (ground). Only potential *differences* are physically meaningful.

Problem 4: Power Method

Let $A \in \mathbb{R}^{n\times n}$ be real symmetric with real eigenvalues

$$\lambda_1, \ldots, \lambda_n \in \mathbb{R}$$

and an orthonormal eigenbasis $u_1, \ldots, u_n$.

(a) Suppose $|\lambda_n| > |\lambda_{n-1}| \geq \cdots \geq |\lambda_1| \geq 0$ and $u_n^{\mathsf{T}} x^{(0)} \neq 0$. Show that the sequence

$$x^{(k+1)} = \frac{1}{\sqrt{(x^{(k)})^{\mathsf{T}} x^{(k)}}} A\, x^{(k)}, \qquad k = 0, 1, 2, \ldots,$$

converges to an eigenvector of A (namely u_n, up to sign), and that entrywise ratios between $Ax^{(k)}$ and $x^{(k)}$ converge to λ_n, for indices where the limit vector has nonzero components.

(b) Describe how to modify the power method to compute the eigenvector corresponding to the second largest eigenvalue λ_{n-1}, assuming that the eigenspace for λ_{n-1} is one-dimensional.

Solution to Problem 4

(a) Convergence of the power method to the dominant eigenvector

Method 1 (Spectral decomposition). Expand the initial vector $x^{(0)}$ in the eigenbasis:

$$x^{(0)} = \sum_{i=1}^{n} \alpha_i u_i, \qquad \alpha_i = u_i^{\mathsf{T}} x^{(0)}.$$

By assumption, $\alpha_n \neq 0$. Then

$$A^k x^{(0)} = \sum_{i=1}^{n} \alpha_i \lambda_i^k u_i = \lambda_n^k \left(\alpha_n u_n + \sum_{i=1}^{n-1} \alpha_i \left(\frac{\lambda_i}{\lambda_n} \right)^k u_i \right).$$

Since $|\lambda_i/\lambda_n| < 1$ for all $i < n$, the terms with $i < n$ decay to zero as $k \to \infty$. Hence, for large k,

$$A^k x^{(0)} \approx \lambda_n^k \alpha_n u_n.$$

The normalized iterate in the ideal (unnormalized) power method is

$$\tilde{x}^{(k)} = \frac{A^k x^{(0)}}{\|A^k x^{(0)}\|}.$$

From the above asymptotics,

$$\tilde{x}^{(k)} \to \pm u_n \quad \text{as } k \to \infty,$$

with sign determined by $\alpha_n \lambda_n^k$.

Our given iteration

$$x^{(k+1)} = \frac{Ax^{(k)}}{\|x^{(k)}\|}$$

is essentially a re-normalized version of this process (up to scaling at each step instead of at the end). One can show by induction that

$$x^{(k)} = \frac{A^k x^{(0)}}{\|A^{k-1}x^{(0)}\| \cdots \|x^{(0)}\|}$$

so $x^{(k)}$ and $A^k x^{(0)}$ differ only by a positive scalar multiple. Thus the direction of $x^{(k)}$ converges to that of u_n, and

$$x^{(k)} \to \pm u_n, \qquad k \to \infty.$$

Componentwise eigenvalue estimate. Once $x^{(k)}$ is close to u_n, we have

$$Ax^{(k)} \approx A(\pm u_n) = \lambda_n(\pm u_n),$$

so for any component index j such that $(u_n)_j \neq 0$,

$$\frac{(Ax^{(k)})_j}{x_j^{(k)}} \to \lambda_n.$$

This is the "entrywise ratio" convergence. In practice, it is common (and numerically more stable) to use the Rayleigh quotient

$$\rho^{(k)} = \frac{(x^{(k)})^\mathsf{T} A x^{(k)}}{(x^{(k)})^\mathsf{T} x^{(k)}},$$

which also converges to λ_n.

Method 2 (Rayleigh quotient viewpoint). Since A is symmetric, the Rayleigh quotient $R_A(x) = \frac{x^\mathsf{T} A x}{x^\mathsf{T} x}$ has the variational characterization

$$\lambda_n = \max_{x \neq 0} R_A(x), \qquad \lambda_1 = \min_{x \neq 0} R_A(x),$$

with maximizers (respectively minimizers) being eigenvectors corresponding to λ_n (respectively λ_1). The power method can be interpreted as a greedy iteration that steers $x^{(k)}$ in the direction that increases $R_A(x^{(k)})$ towards its maximum λ_n. Once $x^{(k)}$ converges in direction to u_n, the Rayleigh quotient and the various componentwise ratios converge to λ_n.

(b) Modifying the power method to obtain the second eigenvector

Assume $|\lambda_n| > |\lambda_{n-1}| > |\lambda_{n-2}| \geq \cdots$ and that the eigenspace of λ_{n-1} is one-dimensional, spanned by u_{n-1}.

Method 1 (Deflation). First, use the power method to approximate the dominant eigenpair (λ_n, u_n). Then define the *deflated* matrix

$$\tilde{A} = A - \lambda_n u_n u_n^{\mathsf{T}}.$$

Because $u_n u_n^{\mathsf{T}}$ projects onto the span of u_n, we have

$$\tilde{A}u_n = Au_n - \lambda_n u_n (u_n^{\mathsf{T}} u_n) = \lambda_n u_n - \lambda_n u_n = 0,$$

so 0 is now the eigenvalue associated with u_n. For all $i < n$,

$$\tilde{A}u_i = Au_i - \lambda_n u_n (u_n^{\mathsf{T}} u_i) = \lambda_i u_i - \lambda_n (0) = \lambda_i u_i,$$

since the eigenvectors are orthonormal. Thus the eigenvalues of $\tilde{A}$ are

$$0,\ \lambda_{n-1}, \ldots, \lambda_1,$$

with $|\lambda_{n-1}|$ now strictly largest among the nonzero ones. Applying the power method to $\tilde{A}$ yields convergence to u_{n-1} (up to sign).

Method 2 (Orthogonalized power method). Another common approach is to enforce orthogonality to u_n at each iteration:

1. Run the power method on A to obtain an approximation $\hat{u}_n$ of the dominant eigenvector.
2. Choose an initial vector $x^{(0)}$ with $x^{(0)} \perp \hat{u}_n$.
3. Iterate

   $y^{(k+1)} = Ax^{(k)}, \qquad y^{(k+1)} \leftarrow y^{(k+1)} - (\hat{u}_n^{\mathsf{T}} y^{(k+1)})\, \hat{u}_n$ (orthogonal projection removing the u_n-compone

$$x^{(k+1)} = \frac{y^{(k+1)}}{\|y^{(k+1)}\|}.$$

Because we continually project away the component along $\hat{u}_n$, the iteration remains in (or close to) the subspace orthogonal to u_n. In that subspace, λ_{n-1} is the eigenvalue of largest magnitude, so the power method now converges to u_{n-1}.

In practice, one often combines deflation and orthogonalization to compute multiple leading eigenpairs robustly (the basis of subspace iteration and the Lanczos method).