1. Review of linear algebra

Notation. Vectors are denoted by boldface lowercase letters (such as $\mathbf{a}, \mathbf{b}$). To indicate their dimensions, we use notation like $\mathbf{a} \in \mathbb{R}^n$. The ith element of $\mathbf{a}$ is written as a_i or $\mathbf{a}(i)$. We denote the constant vector of one as $\mathbf{1}$ (with its dimension implied by the context).

Matrices are denoted by boldface uppercase letters (such as $\mathbf{A}, \mathbf{B}$). Similarly, we write $\mathbf{A} \in \mathbb{R}^{m \times n}$ to indicate its size. The (i, j) entry of $\mathbf{A}$ is denoted by a_{ij} or $\mathbf{A}(i, j)$. The ith row of $\mathbf{A}$ is denoted by $\mathbf{A}(i, :)$ while its columns are written as $\mathbf{A}(:, j)$, as in MATLAB. We use $\mathbf{I}$ to denote the identity matrix (with its dimension implied by the context).

1.1. Matrix multiplication.

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times k}$ be two real matrices. Their product is an $m \times k$ matrix $\mathbf{C} = (c_{ij})$ with entries

$$c_{ij} = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j} = \mathbf{A}(i, :) \cdot \mathbf{B}(:, j).$$

It is possible to obtain one full row (or column) of $\mathbf{C}$ at a time via matrix-vector multiplication:

$$\begin{aligned} \mathbf{C}(i, :) &= \mathbf{A}(i, :) \cdot \mathbf{B} \\ \mathbf{C}(:, j) &= \mathbf{A} \cdot \mathbf{B}(:, j) \end{aligned}$$

The full matrix $\mathbf{C}$ can be written as a sum of rank-1 matrices:

$$\mathbf{C} = \sum_{\ell=1}^{n} \mathbf{A}(:, \ell) \cdot \mathbf{B}(\ell, :).$$

When one of the matrices is a diagonal matrix, we have the following rules:

$$\underbrace{\mathbf{A}}_{\text{diagonal}} \mathbf{B} = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \begin{pmatrix} \mathbf{B}(1, :) \\ \vdots \\ \mathbf{B}(n, :) \end{pmatrix} = \begin{pmatrix} a_1 \mathbf{B}(1, :) \\ \vdots \\ a_n \mathbf{B}(n, :) \end{pmatrix}$$

$$\mathbf{A} \underbrace{\mathbf{B}}_{\text{diagonal}} = [\mathbf{A}(:, 1) \ldots \mathbf{A}(:, n)] \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} = [b_1 \mathbf{A}(:, 1) \ldots b_n \mathbf{A}(:, n)]$$

Finally, below are some identities involving the vector $\mathbf{1}$:

$$\mathbf{1}\mathbf{1}^T = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix}, \qquad \mathbf{1}^T \mathbf{1} = 1$$

$$\mathbf{A}\mathbf{1} = \sum_j \mathbf{A}(:, j), \qquad \mathbf{1}^T \mathbf{A} = \sum_i \mathbf{A}(i, :), \qquad \mathbf{1}^T \mathbf{A} \mathbf{1} = \sum_i \sum_j \mathbf{A}(i, j).$$

EXAMPLE 1.1. Let

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 2 & 3 \end{pmatrix}, \mathbf{\Lambda}_1 = \begin{pmatrix} 1 & & \\ & 0 & \\ & & -1 \end{pmatrix}, \mathbf{\Lambda}_2 = \begin{pmatrix} 2 & \\ & -3 \end{pmatrix}.$$

Find the products $\mathbf{AB}, \mathbf{\Lambda}_1 \mathbf{B}, \mathbf{B} \mathbf{\Lambda}_2, \mathbf{1}^T \mathbf{B}, \mathbf{B1}$ and verify the above rules.

Remark. Another way to multiply two matrices of the same size, say $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$, is through the Hadamard product, also called the entrywise product:

$$\mathbf{C} = \mathbf{A} \circ \mathbf{B} \in \mathbb{R}^{m\times n}, \qquad \text{with} \quad c_{ij} = a_{ij} b_{ij}.$$

For example,

$$\begin{pmatrix} 0 & 2 & -3 \\ -1 & 0 & -4 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & -3 \\ 2 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 9 \\ -2 & 0 & 4 \end{pmatrix}.$$

1.2. Rank, trace and determinant. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. The number of linearly independent rows (or columns) is called the rank of $\mathbf{A}$, and often denoted as $\mathrm{rank}(\mathbf{A})$. It is known that $\mathrm{rank}(\mathbf{A}) \le \min(m, n)$. A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to have full rank if $\mathrm{rank}(\mathbf{A}) = n$; otherwise, it is said to be rank deficient.

The trace of a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is defined as the sum of the entries in its diagonal:

$$\mathrm{trace}(\mathbf{A}) = \sum_i a_{ii}.$$

If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times m$ matrix, then

$$\mathrm{trace}(\mathbf{AB}) = \mathrm{trace}(\mathbf{BA}).$$

The matrix determinant is a rule to evaluate square matrices to numbers:

$$\det : \mathbf{A} \in \mathbb{R}^{n\times n} \mapsto \det(\mathbf{A}) \in \mathbb{R}.$$

Its general definition is quite complicated, but there are lots of different ways to evaluate matrix determinants (see `https://en.wikipedia.org/wiki/Determinant`). The matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be invertible or nonsingular if $\det(A) \neq 0$, which can be shown to be equivalent to being of full rank. An important property of matrix determinant is for two square matrices of the same size $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$,

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}).$$

Example 1.2. For the matrix $\mathbf{A}$ defined in Ex. 1.1, find its rank, trace and determinant.

1.3. Eigenvalues and eigenvectors. Let $\mathbf{A}$ be an $n \times n$ real matrix. The characteristic polynomial of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}).$$

The (complex) roots λ_i of the characteristic equation $p(\lambda) = 0$ are called the eigenvalues of $\mathbf{A}$. For a specific eigenvalue λ_i, any nonzero vector $\mathbf{v}_i$ satisfying

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v}_i = \mathbf{0}$$

or equivalently,

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

is called an eigenvector of $\mathbf{A}$ (associated to the eigenvalue λ_i). All eigenvectors associated to λ_i span a linear subspace, called the eigenspace. It is denoted as $\mathrm{E}(\lambda_i)$. The dimension g_i of $\mathrm{E}(\lambda_i)$ is called the geometric multiplicity of λ_i, while the degree a_i of the factor $(\lambda - \lambda_i)^{a_i}$ in $p(\lambda)$ is called the algebraic multiplicity of λ_i. Note that we must have $\sum a_i = n$ and for all i, $1 \le g_i \le a_i$.

EXAMPLE 1.3. For the matrix $\mathbf{A}$ in Ex. 1.1, find its eigenvalues and their multiplicities, as well as associated eigenvectors.

The following theorem indicates that the trace and determinant of a square matrix can both be computed from the eigenvalues of the matrix.

THEOREM 1.1. *Let $\mathbf{A}$ be a real square matrix whose eigenvalues are $\lambda_1, \ldots, \lambda_n$ (counting multiplicities). Then*

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i \qquad \textit{and} \quad \operatorname{trace}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i.$$

EXAMPLE 1.4. For the matrix $\mathbf{A}$ defined previously, verify the identities in the above theorem.

DEFINITION 1.1. A square matrix $\mathbf{A}$ is diagonalizable if it is similar to a diagonal matrix, i.e., there exist an invertible matrix $\mathbf{P}$ and a diagonal matrix Λ such that

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}.$$

Remark. If we write $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$ and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, then the above equation can be rewritten as $\mathbf{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i$, for all $1 \le i \le n$. This shows that the λ_i are the eigenvalues of $\mathbf{A}$ and $\mathbf{p}_i$ the associated eigenvectors. Thus, the above factorization is called the eigenvalue decomposition of $\mathbf{A}$.

EXAMPLE 1.5. The matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix}$$

is diagonalizable because

$$\begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 3 & \\ & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}^{-1}$$

but $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$ is not (how to determine this?).

The following theorem provides a way for checking the diagonalizability of a square matrix.

THEOREM 1.2. *A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if it has n linearly independent eigenvectors.*

This theorem immediately implies the following results.

COROLLARY 1.3. *The following matrices are diagonalizable:*

- *Any matrix whose eigenvalues all have identical geometric and algebraic multiplicities, i.e., $g_i = a_i$ for all i;*
- *Any matrix with n distinct eigenvalues;*

1.4. Symmetric matrices. A symmetric matrix is a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ whose transpose coincides with itself: $\mathbf{A}^T = \mathbf{A}$. Recall also that an orthogonal matrix is a square matrix whose columns and rows are both orthogonal unit vectors (i.e., orthonormal vectors):

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I},$$

or equivalently,

$$\mathbf{Q}^{-1} = \mathbf{Q}^T.$$

THEOREM 1.4. *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be a symmetric matrix. Then*

- *All the eigenvalues of* $\mathbf{A}$ *are real;*
- $\mathbf{A}$ *is orthogonally diagonalizable, i.e., there exists an orthogonal matrix* $\mathbf{Q}$ *and a diagonal matrix* Λ *such that*

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T.$$

Remark.

- For symmetric matrices, the eigenvalue decomposition is also called the spectral decomposition.
- The converse is also true. Therefore, a matrix is symmetric if and only if it is orthogonally diagonalizable.
- Write $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$. Then

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{q}_i \mathbf{q}_i^T.$$

- We often sort the diagonals of Λ in decreasing order:

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = \lambda_{\min}.$$

EXAMPLE 1.6. Find the spectral decomposition of the following matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix}$$

Answer.

$$\mathbf{A} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & \\ & -1 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^T$$

THEOREM 1.5. *For a given symmetric matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$*, then*

$$\max_{\mathbf{v}\in\mathbb{R}^n:\, \mathbf{v}\neq\mathbf{0}} \frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \lambda_{\max} \quad (\textit{when } \mathbf{v} = \textit{largest eigenvector of } \mathbf{A})$$

$$\min_{\mathbf{v}\in\mathbb{R}^n:\, \mathbf{v}\neq\mathbf{0}} \frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \lambda_{\min} \quad (\textit{when } \mathbf{v} = \textit{smallest eigenvector of } \mathbf{A})$$

Remark. The quantity $\frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\mathbf{v}^T\mathbf{v}}$ is called the *Rayleigh quotient*.

EXAMPLE 1.7. For the matrix $\mathbf{A}$ in the preceding example, the maximum of the Rayleigh quotient is 4, achieved when $\mathbf{v} = \frac{1}{\sqrt{5}}(1, 2)^T$.

DEFINITION 1.2. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is positive semidefinite if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. It is positive definite if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ whenever $\mathbf{x} \neq \mathbf{0}$.

THEOREM 1.6. *A symmetric matrix* $\mathbf{A}$ *is positive definite (semidefinite) if and only if all the eigenvalues are positive (nonnegative).*

EXAMPLE 1.8. Determine if the following matrix is positive definite (or semidefinite):

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

HW0: Linear algebra review

Due: *Wed., August 29, in class.*

- Redo all examples in the lecture notes (Ex. 1.1-1.8) by coding in one of the software (MATLAB, R, or Python) to verify the answers.

  Note that your must show all your codes and output in order to receive full credit. If you chose MATLAB, below are some functions you might need:
  - *.* (entrywise product)*
  - *diag*
  - *trace*
  - *det*
  - *eig, eigs*
  - *repmat*
  - *ones, zeros*
  - *eye*
  - *rand*
- Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $\mathbf{D} \in \mathbb{R}^{n\times n}$ a diagonal matrix with positive diagonal entries (i.e., $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ with $d_1, \ldots, d_n > 0$). Show that the solution of

  $$\max_{\mathbf{v}\in\mathbb{R}^n:\, \mathbf{v}\neq\mathbf{0}} \frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\mathbf{v}^T\mathbf{D}\mathbf{v}}$$

  is given by the largest eigenvector of $\mathbf{D}^{-1}\mathbf{A}$ and the maximum of the quotient is the largest eigenvalue of $\mathbf{D}^{-1}\mathbf{A}$ (which is the same as the largest eigenvalue of $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. Why?).

  Hint: Define

  $$\mathbf{D}^{1/2} = \mathrm{diag}(d_1^{1/2}, \ldots, d_n^{1/2})$$

  and

  $$\mathbf{D}^{-1/2} = \mathrm{diag}(d_1^{-1/2}, \ldots, d_n^{-1/2}).$$

  Change variables by letting

  $$\mathbf{u} = \mathbf{D}^{1/2}\mathbf{v}$$

  to transform the above problem to the one in the lecture notes.