Chapter 3 Transformations

An Introduction to Optimization

Spring, 2014

Wei-Ta Chu

Linear Transformations

- A function $\mathcal{L} : R^n \to R^m$ is called a linear transformation if
  - 1. $\mathcal{L}(a\boldsymbol{x}) = a\mathcal{L}(\boldsymbol{x})$ for every $\boldsymbol{x} \in R^n$ and $a \in R$
  - 2. $\mathcal{L}(\boldsymbol{x}+\boldsymbol{y}) = \mathcal{L}(\boldsymbol{x}) + \mathcal{L}(\boldsymbol{y})$ for every $\boldsymbol{x}, \boldsymbol{y} \in R^n$
- If we fix the bases for R^n and R^m, then the linear transformation can be represented by a matrix.
- Theorem 3.1: Suppose that $\boldsymbol{x} \in R^n$ is a given vector, and $\boldsymbol{x}'$ is the representation of $\boldsymbol{x}$ with respect to the given basis for R^n. If $\boldsymbol{y} = \mathcal{L}(\boldsymbol{x})$ and $\boldsymbol{y}'$ is the representation of $\boldsymbol{y}$ with respect to the given basis for R^m, then $\boldsymbol{y}' = \boldsymbol{A}\boldsymbol{x}'$, where $\boldsymbol{A} \in R^{m\times n}$ and is called the ***matrix representation*** of $\mathcal{L}$.
- Special case: with respect to natural bases for R^n and R^m

$$\boldsymbol{y} = \mathcal{L}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$$

Linear Transformations

- Let $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n\}$ and $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n\}$ be two bases for R^n. Define the matrix

$$\boldsymbol{T} = [\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n]^{-1}[\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n]$$

$$[\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n] = [\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n]\boldsymbol{T}$$

  that is, the *i*th column of $\boldsymbol{T}$ is the vector of coordinates of $\boldsymbol{e}_i$ with respect to the basis $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n\}$.

- Given a vector, let $\boldsymbol{x}$ be the coordinates of the vector with respect to $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n\}$ and $\boldsymbol{x}'$ be the coordinates of the same vector with respect to $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n\}$. Then, $\boldsymbol{x}' = \boldsymbol{T}\boldsymbol{x}$.

Example (Finding a Transition Matrix)

- Consider bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{u}_1', \mathbf{u}_2'\}$ for R^2, where

$$\mathbf{u}_1 = (1, 0), \mathbf{u}_2 = (0, 1);$$
$$\mathbf{u}_1' = (1, 1), \mathbf{u}_2' = (2, 1).$$

  Find the transition matrix from B' to B.

  Find $[\mathbf{v}]_B$ if $[\mathbf{v}]_{B'} = [-3\ 5]^T$.

- Solution:
  - First we must find the coordinate matrices for the new basis vectors $\mathbf{u}_1'$ and $\mathbf{u}_2'$ relative to the old basis B.
  - By inspection $\mathbf{u}'_1 = \mathbf{u}_1 + \mathbf{u}_2$ so that

$$[\mathbf{u}'_1]_B = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } [\mathbf{u}'_2]_B = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\boldsymbol{u}_1' = \boldsymbol{u}_1 + \boldsymbol{u}_2$$
$$\boldsymbol{u}_2' = 2\boldsymbol{u}_1 + \boldsymbol{u}_2$$

  - Thus, the transition matrix from B' to B is

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$

Example (Finding a Transition Matrix)

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$

- Using the transition matrix yields

$$[\boldsymbol{v}]_B = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 2 \end{bmatrix}$$

- As a check, we should be able to recover the vector **v** either from $[\mathbf{v}]_B$ or $[\mathbf{v}]_{B'}$.
- $-3\mathbf{u}_1' + 5\mathbf{u}_2' = 7\mathbf{u}_1 + 2\mathbf{u}_2 = \mathbf{v} = (7,2)$

Example (A Different Viewpoint)

$$\mathbf{u}_1 = (1, 0),\ \mathbf{u}_2 = (0, 1);\ \mathbf{u}_1' = (1, 1),\ \mathbf{u}_2' = (2, 1)$$

- In the previous example, we found the transition matrix from the basis *B*' to the basis *B*. However, we can just as well ask for the transition matrix from *B* to *B*'.
- We simply change our point of view and regard *B*' as the old basis and *B* as the new basis.
- As usual, the columns of the transition matrix will be the coordinates of the new basis vectors relative to the old basis.

$$\mathbf{u}_1 = -\mathbf{u}_1' + \mathbf{u}_2';\mathbf{u}_2 = 2\mathbf{u}_1' - \mathbf{u}_2'$$

$$[\boldsymbol{u}_1]_{B'} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad [\boldsymbol{u}_2]_{B'} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad Q = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$$

Remarks

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \qquad Q = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$$

- If we multiply the transition matrix from *B'* to *B* and the transition matrix from *B* to *B'*, we find

$$PQ = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

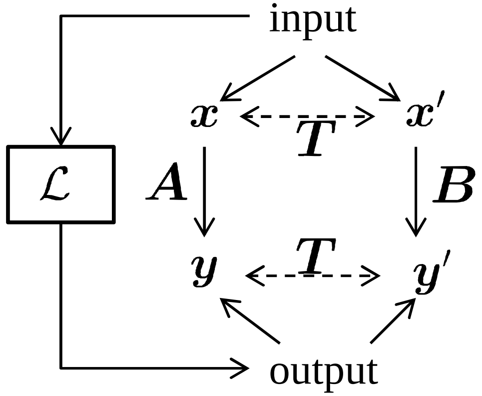
$$Q = P^{-1}$$

Linear Transformations

- Consider a linear transformation $\mathcal{L}: R^n \to R^n$ and let $\boldsymbol{A}$ be its representation with respect to $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ and $\boldsymbol{B}$ its representation with respect to $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, \ldots, \boldsymbol{e}'_n\}$.
- Let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{y}' = \boldsymbol{B}\boldsymbol{x}'$. Therefore,

$$\boldsymbol{y}' = \boldsymbol{T}\boldsymbol{y} = \boldsymbol{T}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x}' = \boldsymbol{B}\boldsymbol{T}\boldsymbol{x}$$

  and hence $\boldsymbol{T}\boldsymbol{A} = \boldsymbol{B}\boldsymbol{T}$, or $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$
- Two $n \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are ***similar*** if there exists a nonsingular matrix $\boldsymbol{T}$ such that $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$.
- In conclusion, similar matrices correspond to the same linear transformation with respect to difference bases.



Eigenvalues and Eigenvectors

- Let $\boldsymbol{A}$ be an $n \times n$ square matrix. A scalar λ and a nonzero vector $\boldsymbol{v}$ satisfying the equation $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ are said to be, respectively, an ***eigenvalue*** and an ***eigenvector*** of $\boldsymbol{A}$.
- The matrix $\lambda\boldsymbol{I} - \boldsymbol{A}$ must be singular; that is, $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0$
- This leads to an *n*th-order polynomial equation

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0$$

The polynomial $\det(\lambda\boldsymbol{I} - \boldsymbol{A})$ is called the ***characteristic polynomial***, and the equation is called the ***characteristic equation***.

Eigenvalues and Eigenvectors

- Suppose that the characteristic equation $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$ has *n* distinct roots $\lambda_1, \lambda_2, ..., \lambda_n$. Then, there exist *n* linearly independent vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ such that

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \qquad i = 1, 2, ..., n$$

- Consider a basis formed by a linearly independent set of eigenvectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$. With respect to this basis, the matrix $\boldsymbol{A}$ is ***diagonal***.
- Let $\boldsymbol{T} = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n]^{-1}$

$$\begin{aligned}
\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} &= \boldsymbol{T}\boldsymbol{A}[\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n] \\
&= \boldsymbol{T}[\boldsymbol{A}\boldsymbol{v}_1, \boldsymbol{A}\boldsymbol{v}_2, ..., \boldsymbol{A}\boldsymbol{v}_n] = \boldsymbol{T}[\lambda_1\boldsymbol{v}_1, \lambda_2\boldsymbol{v}_2, ..., \lambda_n\boldsymbol{v}_n] \\
&= \boldsymbol{T}\boldsymbol{T}^{-1}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}
\end{aligned}$$

Eigenvalues and Eigenvectors

- A matrix $\boldsymbol{A}$ is symmetric if $\boldsymbol{A} = \boldsymbol{A}^T$.
- Theorem 3.2: All eigenvalues of a real symmetric matrix are real.
- Theorem 3.3: Any real symmetric $n \times n$ matrix has a set of *n* eigenvectors that are mutually orthogonal. (i.e., this matrix can be orthogonally diagonalized)
- If $\boldsymbol{A}$ is symmetric, then a set of its eigenvectors forms an orthogonal basis for R^n. If the basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ is normalized so that each element has norm of unity, then defining the matrix $\boldsymbol{T} = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n]$
  we have $\boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{I}$, or $\boldsymbol{T}^T = \boldsymbol{T}^{-1}$
- A matrix whose transpose is its inverse is said to be an ***orthogonal matrix***.

Example

- Find an orthogonal matrix P that diagonalizes
$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}$$
- Solution:
  - The characteristic equation of A is
$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda - 4 & -2 & -2 \\ -2 & \lambda - 4 & -2 \\ -2 & -2 & \lambda - 4 \end{bmatrix} = (\lambda - 2)^2 (\lambda - 8) = 0$$
  - The basis of the eigenspace corresponding to $\lambda = 2$ is $\mathbf{u}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$
  - Applying the Gram-Schmidt process to $\{\mathbf{u}_1, \mathbf{u}_2\}$ yields the following orthonormal eigenvectors:
$$\mathbf{v}_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \text{ and } \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}$$

Example

- The basis of the eigenspace corresponding to $\lambda = 8$ is $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
- Applying the Gram-Schmidt process to $\{\mathbf{u}_3\}$ yields:

$$\mathbf{v}_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

- Thus,

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}$$

orthogonally diagonalizes A.

Orthogonal Projections

- If $\mathcal{V}$ is a subspace of R^n, then the ***orthogonal complement*** of $\mathcal{V}$, denoted by $\mathcal{V}^\perp$, consists of all vectors that are orthogonal to every vector in $\mathcal{V}$, i.e. $\mathcal{V}^\perp = \{\boldsymbol{x} : \boldsymbol{v}^T\boldsymbol{x} = 0 \text{ for all } \boldsymbol{v} \in \mathcal{V}\}$
- The orthogonal complement of $\mathcal{V}$ is also a subspace.
- Together, $\mathcal{V}$ and $\mathcal{V}^\perp$ span R^n in the sense that every vector $\boldsymbol{x} \in R^n$ can be represented uniquely as $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \in \mathcal{V}$ and $\boldsymbol{x}_2 \in \mathcal{V}^\perp$
- The representation above is the ***orthogonal decomposition*** of $\boldsymbol{x}$
- We say that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are ***orthogonal projections*** of $\boldsymbol{x}$ onto the subspaces $\mathcal{V}$ and $\mathcal{V}^\perp$, respectively. We write $R^n = \mathcal{V} \oplus \mathcal{V}^\perp$ and say that R^n is a ***direct sum*** of $\mathcal{V}$ and $\mathcal{V}^\perp$. We say that a linear transformation $\boldsymbol{P}$ is an ***orthogonal projector*** onto $\mathcal{V}$ if for all $\boldsymbol{x} \in R^n$ we have $\boldsymbol{P}\boldsymbol{x} \in \mathcal{V}$ and $\boldsymbol{x} - \boldsymbol{P}\boldsymbol{x} \in \mathcal{V}^\perp$

Orthogonal Projections

- Theorem 3.4: Let $\boldsymbol{A} \in R^{m \times n}$, the ***range*** or ***image*** of $\boldsymbol{A}$ can be denoted

$$\mathcal{R}(\boldsymbol{A}) \triangleq \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in R^n\}$$

Column space

- The ***nullspace*** or ***kernel*** of $\boldsymbol{A}$ can be denoted

$$\mathcal{N}(\boldsymbol{A}) \triangleq \{\boldsymbol{x} \in R^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$$

- $\mathcal{R}(\boldsymbol{A})$ and $\mathcal{N}(\boldsymbol{A})$ are subspaces.
- $\mathcal{R}(\boldsymbol{A})^{\perp} = \mathcal{N}(\boldsymbol{A}^T)$ and $\mathcal{N}(\boldsymbol{A})^{\perp} = \mathcal{R}(\boldsymbol{A}^T)$ (four ***fundamental spaces*** in Linear Algebra)

  Row space

- If $\boldsymbol{P}$ is an orthogonal projector onto $\mathcal{V}$, then $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{V}$, and $\mathcal{R}(\boldsymbol{P}) = \mathcal{V}$
- Theorem 3.5: A matrix $\boldsymbol{P}$ is an orthogonal projector if and only if $\boldsymbol{P}^2 = \boldsymbol{P} = \boldsymbol{P}^T$

Quadratic Forms

$$a_1 x_1^2 + a_2 x_2^2 + a_3 x_1 x_2 \quad \Rightarrow \quad \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_1 & a_3/2 \\ a_3/2 & a_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2 + a_4 x_1 x_2 + a_5 x_1 x_3 + a_6 x_2 x_3$$

$$\Rightarrow \quad \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_1 & a_4/2 & a_5/2 \\ a_4/2 & a_2 & a_6/2 \\ a_5/2 & a_6/2 & a_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

- A quadratic form $f : R^n \to R^n$ is a function $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$, where $\boldsymbol{Q}$ is an $n \times n$ real matrix. There is no loss of generality in assuming $\boldsymbol{Q}$ to be symmetric: $\boldsymbol{Q} = \boldsymbol{Q}^T$

$$2x^2 + 6xy - 7y^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 1 & -7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Quadratic Forms

$$2x^2 + 6xy - 7y^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 1 & -7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & -7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

- For if the matrix $\boldsymbol{Q}$ is not symmetric, we can always replace it with the symmetric

$$\boldsymbol{Q}_0 = \boldsymbol{Q}_0^T = \frac{1}{2}\left(\boldsymbol{Q} + \boldsymbol{Q}^T\right)$$

$$\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{Q}_0 \boldsymbol{x} = \boldsymbol{x}^T \left(\frac{1}{2}\boldsymbol{Q} + \frac{1}{2}\boldsymbol{Q}^T\right) \boldsymbol{x}$$

- A quadratic form $\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$ is said to be ***positive definite*** if $\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} > 0$ for all nonzero vectors $\boldsymbol{x}$. It is ***positive semidefinite*** if $\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$. Similarly, we define the quadratic form to be ***negative definite***, or ***negative semidefinite***, if $\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} < 0$ or $\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \leq 0$

Quadratic Forms

- The ***principal minors*** for a matrix $\boldsymbol{Q}$ are $\det(\boldsymbol{Q})$ itself and the determinants of matrices obtained by successively removing an *i*th row and an *i*th column.
- The ***leading principal minors*** are $\det(\boldsymbol{Q})$ and the minors obtained by successive removing the last row and the last column.

$$\Delta_1 = q_{11} \qquad \Delta_2 = \det \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

$$\Delta_3 = \det \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \qquad \cdots \qquad \Delta_n = \det(\boldsymbol{Q})$$

Quadratic Forms

- **Theorem 3.6 Sylvester's Criterion**: A quadratic form $\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q} = \boldsymbol{Q}^T$, is positive definite if and only if the leading principal minors of $\boldsymbol{Q}$ are positive.
- Note that if $\boldsymbol{Q}$ is not symmetric, Sylvester's criterion cannot be used.
- A *necessary* condition for a real quadratic form to be positive semidefinite is that the leading principal minors be nonnegative. However, it is *not* a *sufficient* condition. In fact, a real quadratic form is positive semidefinite if and only if all *principal minors* are nonnegative.

Quadratic Forms

- A symmetric matrix $\boldsymbol{Q}$ is said to be ***positive definite*** if the quadratic form $\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$ is positive definite.
- If $\boldsymbol{Q}$ is positive definite, we write $\boldsymbol{Q} > 0$
- Positive semidefinite, negative definite, negative semidefinite properties are defined similarly.
- The symmetric matrix $\boldsymbol{Q}$ is ***indefinite*** if it is neither positive semidefinite nor negative semidefinite.
- Theorem 3.7: A symmetric matrix $\boldsymbol{Q}$ is positive definite (or positive semidefinite) if and only if all eigenvalues of $\boldsymbol{Q}$ are positive (or nonnegative)

Matrix Norms

- The norm of a matrix $\boldsymbol{A}$, denoted by $\|\boldsymbol{A}\|$, is any function that satisfies the following conditions:
  - $\|\boldsymbol{A}\| > 0$ if $\boldsymbol{A} \neq \boldsymbol{O}$, and $\|\boldsymbol{O}\| = 0$, where $\boldsymbol{O}$ is a matrix with all entries equal to zero.
  - $\|c\boldsymbol{A}\| = |c|\|\boldsymbol{A}\|$, for any $c \in R$
  - $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}|$
- An example of a matrix norm is the ***Frobenius norm***, defined as $\|\boldsymbol{A}\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij})^2\right)^{1/2}$
- Note that the Frobenius norm is equivalent to the Euclidean norm on R^{mn}.
- For our purpose, we consider only matrix norms satisfying the addition condition: $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$

Matrix Norms

- In many problems, both matrices and vectors appear simultaneously. Therefore, it is convenient to construct the matrix norm in such a way that it will be related to vector norms.
- To this end we consider a special class of matrix norms, called ***induced norms***.
- Let $\|\cdot\|_{(n)}$ and $\|\cdot\|_{(m)}$ be vector norms on R^n and R^m, respectively. We say that the matrix norm is ***induced*** by, or is ***compatible*** with, the given vector norms if for any matrix $\boldsymbol{A} \in R^{m\times n}$ and any vector $\boldsymbol{x} \in R^n$, the following inequality is satisfied:

$$\|\boldsymbol{A}\boldsymbol{x}\|_{(m)} \leq \|\boldsymbol{A}\|\|\boldsymbol{x}\|_{(n)}$$

Matrix Norms

- We can define an induced matrix norm as
$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|_{(n)}=1} \|\boldsymbol{A}\boldsymbol{x}\|_{(m)}$$
that is, $\|\boldsymbol{A}\|$ is the maximum of the norms of the vectors $\boldsymbol{A}\boldsymbol{x}$ where the vector $\boldsymbol{x}$ runs over the set of all vectors with unit norm. We may omit the subscripts in the following.
- For each matrix $\boldsymbol{A}$ the maximum $\max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|$ is attainable; that is, a vector $\boldsymbol{x}_0$ exists such that $\|\boldsymbol{x}_0\| = 1$ and $\|\boldsymbol{A}\boldsymbol{x}_0\| = \|\boldsymbol{A}\|$

Matrix Norms

- Theorem 3.8: Let
$$\|\boldsymbol{x}\| = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$$
the matrix norm induced by this vector norm is
$$\|\boldsymbol{A}\| = \sqrt{\lambda_1}$$
where λ_1 is the largest eigenvalue of the matrix $\boldsymbol{A}^T\boldsymbol{A}$
- **Rayleigh's Inequality**: If an $n \times n$ matrix $\boldsymbol{P}$ is real symmetric positive definite, then
$$\lambda_{min}(\boldsymbol{P})\|\boldsymbol{x}\|^2 \leq \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} \leq \lambda_{max}(\boldsymbol{P})\|\boldsymbol{x}\|^2$$
where $\lambda_{min}(\boldsymbol{P})$ denotes the smallest eigenvalue of $\boldsymbol{P}$, and $\lambda_{max}(\boldsymbol{P})$ denotes the largest eigenvalue of $\boldsymbol{P}$.

Matrix Norms

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

- Consider the matrix and let the norm in R^2 be given by

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2}$$

Then, $\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$

and $\det(\lambda \boldsymbol{I}_2 - \boldsymbol{A}^T\boldsymbol{A}) = \lambda^2 - 10\lambda + 9 = (\lambda - 1)(\lambda - 9)$

Thus, $\|\boldsymbol{A}\| = \sqrt{9} = 3$

- The eigenvector of $\boldsymbol{A}^T\boldsymbol{A}$ corresponding to $\lambda_1 = 9$ is $\boldsymbol{x}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- Note that $\|\boldsymbol{A}\boldsymbol{x}_1\| = \|\boldsymbol{A}\|$

$$\|\boldsymbol{A}\boldsymbol{x}_1\| = \left\| \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \frac{1}{\sqrt{2}} \left\| \begin{bmatrix} 3 \\ 3 \end{bmatrix} \right\| = 3$$

- Because $\boldsymbol{A} = \boldsymbol{A}^T$ in this example, we have $\|\boldsymbol{A}\| = \max_{1 \le i \le n} |\lambda_i(\boldsymbol{A})|$. However, in general $\|\boldsymbol{A}\| \neq \max_{1 \le i \le n} |\lambda_i(\boldsymbol{A})|$. Indeed, we have $\|\boldsymbol{A}\| \geq \max_{1 \le i \le n} |\lambda_i(\boldsymbol{A})|$

Example

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$\det[\lambda \boldsymbol{I}_2 - \boldsymbol{A}^T\boldsymbol{A}] = \det \begin{bmatrix} \lambda & 0 \\ 0 & \lambda - 1 \end{bmatrix} = \lambda(\lambda - 1)$$

- Note that 0 is the only eigenvalue of $\boldsymbol{A}$. Thus, for $i = 1, 2$, $\|\boldsymbol{A}\| = 1 > |\lambda_i(\boldsymbol{A})| = 0$