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. $\mathcal{L}(ax) = a\mathcal{L}(x)$ for every $x \in R^n$ and $a \in R$
  2. $\mathcal{L}(x + y) = \mathcal{L}(x) + \mathcal{L}(y)$ for every $x, y \in R^n$

- 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 $x \in R^n$ is a given vector, and $x'$ is the representation of $x$ with respect to the given basis for $R^n$. If $y = \mathcal{L}(x)$ and $y'$ is the representation of $y$ with respect to the given basis for $R^m$, then $y' = Ax'$, where $A \in R^{m \times n}$ and is called the **matrix representation** of $\mathcal{L}$.

- Special case: with respect to natural bases for $R^n$ and $R^m$
  \[ y = \mathcal{L}(x) = Ax \]
Linear Transformations

Let \( \{e_1, e_2, ..., e_n\} \) and \( \{e'_1, e'_2, ..., e'_n\} \) be two bases for \( \mathbb{R}^n \). Define the matrix

\[
T = [e'_1, e'_2, ..., e'_n]^{-1}[e_1, e_2, ..., e_n]
\]

\[
[e_1, e_2, ..., e_n] = [e'_1, e'_2, ..., e'_n]T
\]

that is, the \( i \)th column of \( T \) is the vector of coordinates of \( e_i \) with respect to the basis \( \{e'_1, e'_2, ..., e'_n\} \).

Given a vector, let \( x \) be the coordinates of the vector with respect to \( \{e_1, e_2, ..., e_n\} \) and \( x' \) be the coordinates of the same vector with respect to \( \{e'_1, e'_2, ..., e'_n\} \). Then, \( x' = Tx \).
Example (Finding a Transition Matrix)

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

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

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

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

- 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

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

- As a check, we should be able to recover the vector \( v \) either from \([v]_B\) or \([v]_B'\).

- \(-3u_1' + 5u_2' = 7u_1 + 2u_2 = v = (7,2)\)
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.

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

\[
\begin{bmatrix}
\mathbf{u}_1 \\
\mathbf{u}_2
\end{bmatrix}_{B'} = \begin{bmatrix}
-1 \\
1
\end{bmatrix} \quad \begin{bmatrix}
\mathbf{u}_2 \\
\mathbf{u}_1
\end{bmatrix}_{B'} = \begin{bmatrix}
2 \\
-1
\end{bmatrix} \quad Q = \begin{bmatrix}
-1 & 2 \\
1 & -1
\end{bmatrix}
\]
Remarks

\[ P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \quad Q = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \]

- If we multiply the transition matrix from \( B' \) to \( B \) and the transition matrix from \( B \) to \( B' \), we find

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

\[ Q = P^{-1} \]
Consider a linear transformation \( \mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}^n \) and let \( A \) be its representation with respect to \( \{e_1, e_2, \ldots, e_n\} \) and \( B \) its representation with respect to \( \{e'_1, e'_2, \ldots, e'_n\} \).

Let \( y = Ax \) and \( y' = Bx' \). Therefore,
\[
y' = Ty = TAx = Bx' = BTx
\]
and hence \( TA = BT \), or \( A = T^{-1}BT \).

Two \( n \times n \) matrices \( A \) and \( B \) are *similar* if there exists a nonsingular matrix \( T \) such that \( A = T^{-1}BT \).

In conclusion, similar matrices correspond to the same linear transformation with respect to different bases.
Eigenvalues and Eigenvectors

- Let $A$ be an $n \times n$ square matrix. A scalar $\lambda$ and a nonzero vector $v$ satisfying the equation $Av = \lambda v$ are said to be, respectively, an eigenvalue and an eigenvector of $A$.
- The matrix $\lambda I - A$ must be singular; that is, $\det(\lambda I - A) = 0$.
- This leads to an $n$th-order polynomial equation
  \[ \det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0 \]
  The polynomial $\det(\lambda I - A)$ is called the characteristic polynomial, and the equation is called the characteristic equation.
Suppose that the characteristic equation $\det(\lambda I - A) = 0$ has $n$ distinct roots $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then, there exist $n$ linearly independent vectors $v_1, v_2, \ldots, v_n$ such that

$$Av_i = \lambda_i v_i \quad i = 1, 2, \ldots, n$$

Consider a basis formed by a linearly independent set of eigenvectors $\{v_1, v_2, \ldots, v_n\}$. With respect to this basis, the matrix $A$ is \textit{diagonal}.

Let $T = [v_1, v_2, \ldots, v_n]^{-1}$

$$TAT^{-1} = TA[v_1, v_2, \ldots, v_n]$$

$$= T[A v_1, A v_2, \ldots, A v_n] = T[\lambda_1 v_1, \lambda_2 v_2, \ldots, \lambda_n v_n]$$

$$= TT^{-1} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$
Eigenvalues and Eigenvectors

- A matrix $A$ is symmetric if $A = A^T$.
- 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 $A$ is symmetric, then a set of its eigenvectors forms an orthogonal basis for $\mathbb{R}^n$. If the basis $\{v_1, v_2, ..., v_n\}$ is normalized so that each element has norm of unity, then defining the matrix $T = [v_1, v_2, ..., v_n]$ we have $T^T T = I$, or $T^T = T^{-1}$
- 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:

$$\mathbf{v}_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}$$
Example

- The basis of the eigenspace corresponding to $\lambda = 8$ is $u_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- Applying the Gram-Schmidt process to $\{u_3\}$ yields:

$$v_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

- Thus,

$$P = [v_1 \quad v_2 \quad v_3] = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}$$

orthogonally diagonalizes $A$. 
Orthogonal Projections

- If $\mathcal{V}$ is a subspace of $\mathbb{R}^n$, 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 $\mathbb{R}^n$ in the sense that every vector $\boldsymbol{x} \in \mathbb{R}^n$ 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 $\mathbb{R}^n = \mathcal{V} \oplus \mathcal{V}^\perp$ and say that $\mathbb{R}^n$ is a **direct sum** of $\mathcal{V}$ and $\mathcal{V}^\perp$. We say that a linear transformation $P$ is an **orthogonal projector** onto $\mathcal{V}$ if for all $\boldsymbol{x} \in \mathbb{R}^n$ we have $P\boldsymbol{x} \in \mathcal{V}$ and $\boldsymbol{x} - P\boldsymbol{x} \in \mathcal{V}^\perp$. 
Orthogonal Projections

- Theorem 3.4: Let $A \in \mathbb{R}^{m \times n}$, the range or image of $A$ can be denoted
  \[ \mathcal{R}(A) \triangleq \{ Ax : x \in \mathbb{R}^n \} \]
  \textit{Column space}

- The nullspace or kernel of $A$ can be denoted
  \[ \mathcal{N}(A) \triangleq \{ x \in \mathbb{R}^n : Ax = 0 \} \]

- $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are subspaces.

- $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ and $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$ \textit{(four fundamental spaces in Linear Algebra)}
  \textit{Row space}

- If $P$ is an orthogonal projector onto $\mathcal{V}$, then $P x = x$ for all $x \in \mathcal{V}$, and $\mathcal{R}(P) = \mathcal{V}$

- Theorem 3.5: A matrix $P$ is an orthogonal projector if and only if $P^2 = P = P^T$
Quadratic Forms

\[ a_1x_1^2 + a_2x_2^2 + a_3x_1x_2 \quad \Rightarrow \quad [x_1 \quad x_2] \begin{bmatrix} a_1 & a_3/2 \\ a_3/2 & a_2 \end{bmatrix} [x_1 \\ x_2] \]

\[ a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + a_4x_1x_2 + a_5x_1x_3 + a_6x_2x_3 \]

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

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

\[ 2x^2 + 6xy - 7y^2 = [x \quad y] \begin{bmatrix} 2 & 5 \\ 1 & -7 \end{bmatrix} [x \\ y] \]
For if the matrix $Q$ is not symmetric, we can always replace it with the symmetric

$$Q_0 = Q_0^T = \frac{1}{2} (Q + Q^T)$$

$$x^T Q x = x^T Q_0 x = x^T \left( \frac{1}{2} Q + \frac{1}{2} Q^T \right) x$$

A quadratic form $x^T Q x$ is said to be **positive definite** if $x^T Q x > 0$ for all nonzero vectors $x$. It is **positive semidefinite** if $x^T Q x \geq 0$ for all $x$. Similarly, we define the quadratic form to be **negative definite**, or **negative semidefinite**, if $x^T Q x < 0$ or $x^T Q x \leq 0$.
The principal minors for a matrix $Q$ are $\det(Q)$ itself and the determinants of matrices obtained by successively removing an $i$th row and an $i$th column.

The leading principal minors are $\det(Q)$ and the minors obtained by successive removing the last row and the last column.

$$
\begin{align*}
\Delta_1 & = q_{11} \\
\Delta_2 & = \det \begin{bmatrix} q_{11} & q_{12} \\
q_{21} & q_{22} \end{bmatrix} \\
\Delta_3 & = \det \begin{bmatrix} q_{11} & q_{12} & q_{13} \\
q_{21} & q_{22} & q_{23} \\
q_{31} & q_{32} & q_{33} \end{bmatrix} \\
& \vdots \\
\Delta_n & = \det(Q)
\end{align*}
$$
Theorem 3.6 Sylvester’s Criterion: A quadratic form $x^T Q x$, $Q = Q^T$, is positive definite if and only if the leading principal minors of $Q$ are positive.

Note that if $Q$ 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.
A symmetric matrix $Q$ is said to be **positive definite** if the quadratic form $x^TQx$ is positive definite.

If $Q$ is positive definite, we write $Q > 0$.

Positive semidefinite, negative definite, negative semidefinite properties are defined similarly.

The symmetric matrix $Q$ is **indefinite** if it is neither positive semidefinite nor negative semidefinite.

Theorem 3.7: A symmetric matrix $Q$ is positive definite (or positive semidefinite) if and only if all eigenvalues of $Q$ are positive (or nonnegative).
Matrix Norms

- The norm of a matrix $A$, denoted by $\|A\|$, is any function that satisfies the following conditions:
  - $\|A\| > 0$ if $A \neq O$, and $\|O\| = 0$, where $O$ is a matrix with all entries equal to zero.
  - $\|cA\| = |c|\|A\|$, for any $c \in R$
  - $\|A + B\| \leq \|A\| + \|B\|$

- An example of a matrix norm is the **Frobenius norm**, defined as
  $$\|A\|_F = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij})^2 \right)^{1/2}$$

- 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: $\|AB\| \leq \|A\|\|B\|$
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 $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. We say that the matrix norm is **induced** by, or is **compatible** with, the given vector norms if for any matrix $A \in \mathbb{R}^{m \times n}$ and any vector $x \in \mathbb{R}^n$, the following inequality is satisfied:

$$\|Ax\|_m \leq \|A\| \|x\|_n$$
Matrix Norms

- We can define an induced matrix norm as
  \[ \| A \| = \max_{\| x \| = 1} \| Ax \| \]
  that is, \( \| A \| \) is the maximum of the norms of the vectors \( Ax \)
  where the vector \( x \) runs over the set of all vectors with unit norm.
  We may omit the subscripts in the following.

- For each matrix \( A \) the maximum \( \max_{\| x \| = 1} \| Ax \| \) is attainable;
  that is, a vector \( x_0 \) exists such that \( \| x_0 \| = 1 \) and \( \| Ax_0 \| = \| A \| \)
Matrix Norms

- **Theorem 3.8**: Let
  \[
  \|x\| = \left( \sum_{k=1}^{n} |x_k|^2 \right)^{1/2} = \sqrt{\langle x, x \rangle}
  \]

  the matrix norm induced by this vector norm is
  \[
  \|A\| = \sqrt{\lambda_1}
  \]

  where \(\lambda_1\) is the largest eigenvalue of the matrix \(A^T A\)

- **Rayleigh’s Inequality**: If an \(n \times n\) matrix \(P\) is real symmetric positive definite, then
  \[
  \lambda_{\text{min}}(P)\|x\|^2 \leq x^T P x \leq \lambda_{\text{max}}(P)\|x\|^2
  \]

  where \(\lambda_{\text{min}}(P)\) denotes the smallest eigenvalue of \(P\), and \(\lambda_{\text{max}}(P)\)
  denotes the largest eigenvalue of \(P\).
Consider the matrix and let the norm in $\mathbb{R}^2$ be given by

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

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

and $\det(\lambda I_2 - A^T A) = \lambda^2 - 10\lambda + 9 = (\lambda - 1)(\lambda - 9)$

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

The eigenvector of $A^T A$ corresponding to $\lambda_1 = 9$ is $x_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Note that $\|Ax_1\| = \|A\|$

$$\|Ax_1\| = \left\| \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \frac{1}{\sqrt{2}} \left\| \begin{bmatrix} 3 \\ 3 \end{bmatrix} \right\| = 3$$

Because $A = A^T$ in this example, we have $\|A\| = \max_{1 \leq i \leq n} |\lambda_i(A)|$.

However, in general $\|A\| \neq \max_{1 \leq i \leq n} |\lambda_i(A)|$. Indeed, we have $\|A\| \geq \max_{1 \leq i \leq n} |\lambda_i(A)|$.
Example

\[
A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad A^T A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}
\]

\[
\det[\lambda I_2 - A^T A] = \det \begin{bmatrix} \lambda & 0 \\ 0 & \lambda - 1 \end{bmatrix} = \lambda(\lambda - 1)
\]

- Note that 0 is the only eigenvalue of A. Thus, for \( i = 1, 2, \)

\[
\|A\| = 1 > |\lambda_i(A)| = 0
\]