Bilinear Forms and Sesquilinear Forms

FEI LI[1]

March 2016

In this article, we give a detailed look at bilinear/sesquilinear forms in linear algebra. We're interested in conditions under which a form possesses an **orthonormal basis**. It turns out that this is true *if and only if it is (Hermitian) symmetric and positive definite* (Theorem 2.32). We show that the symmetry condition is due to symmetry of orthogonality. We then discuss operators and quadratic forms using our theory of bilinear forms.

[1]Email: fei.li.best@gmail.com.

1 Some Linear Algebra

Definition 1.1. A *vector space with basis* $\mathbf{w}$ is an ordered pair $(V, \mathbf{w})$ where V is a vector space over some field F and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ is a basis of V. If $v \in V, v = x_1 w_1 + x_2 w_2 + \cdots + x_n w_n$, then we use $(\mathbf{x}, \mathbf{w})$ to denote v.

We let $\mathscr{L}(V)$ denote the set of all linear operators on V. For $T \in \mathscr{L}(V)$ and $(V, \mathbf{w})$, we use $\mathcal{M}_T$ to denote the matrix of T with respect to $\mathbf{w}$. If say $(\mathbf{x}, \mathbf{w}) \mapsto (\mathbf{y}, \mathbf{w})$ by T, then $\mathbf{x} \mapsto \mathbf{y}$ by $\mathcal{M}_T$.

Definition 1.2. The *transpose* of a matrix $\mathcal{M}_T$, denoted as $\mathcal{M}_T^t$, is the matrix obtained by interchanging rows and columns of $\mathcal{M}_T$. The *complex transpose* of $\mathcal{M}_T$ is the matrix $\mathcal{M}_T^*$ obtained by taking complex conjugate of each entry of $\mathcal{M}_T^t$. It is also called the *adjoint* of $\mathcal{M}_T$.

Definition 1.3. Let $\mathcal{M}_T$ be a matrix with respect to a real or complex vector space $(V, \mathbf{w})$.

- It is *symmetric* if $\mathcal{M}_T = \mathcal{M}_T^t$;
- It is *Hermitian* (or *self-adjoint*) if $\mathcal{M}_T = \mathcal{M}_T^*$;
- It is *orthogonal* if $\mathcal{M}_T^t \mathcal{M}_T = I$;
- It is *unitary* if $\mathcal{M}_T^* \mathcal{M}_T = I$.

Change of Basis

For two basis $\mathbf{w}$ and $\mathbf{w}'$, let S be the linear operator that sends each w_i to w_i', i.e., $S w_i = w_i'$ for $i = 1, \ldots, n$. Let P be the matrix of S with respect to $(V, \mathbf{w})$. It is called the *base change matrix*.

Theorem 1.4. Let $T \in \mathscr{L}(V)$, and let $\mathbf{w}$ and $\mathbf{w}'$ be two basis of V. Suppose $\mathcal{M}_T$ is the matrix of T with respect to $(V, \mathbf{w})$. Then the matrix of T with respect to $(V, \mathbf{w}')$ is given by $\mathcal{M}_T' = P^{-1} \mathcal{M}_T P$.

Proof. Let $v \in V$ be arbitrary and suppose $Tv = z$, and $v = (\mathbf{x}, \mathbf{w}) = (\mathbf{x}', \mathbf{w}')$, while $z = (\mathbf{y}, \mathbf{w}) = (\mathbf{y}', \mathbf{w}')$. Then

$$\mathbf{y} = \mathcal{M}_T \mathbf{x} \tag{1}$$

$$\mathbf{y}' = \mathcal{M}_T' \mathbf{x}'. \tag{2}$$

We show $\mathbf{x} = P\mathbf{x}'$ and $\mathbf{y} = P\mathbf{y}'$. A demonstrative example of this should be enough. When V is 2-dimensional, $\mathbf{w} = (w_1, w_2)$, $\mathbf{w}' = (w_1', w_2')$, $w_1' = aw_1 + bw_2, w_2' = cw_1 + dw_2$,

$$\begin{aligned} v &= x_1 w_1 + x_2 w_2 \\ &= x_1' w_1' + x_2' w_2' \\ &= x_1'(aw_1 + bw_2) + x_2'(cw_1 + dw_2) \\ &= (ax_1' + cx_2')w_1 + (bx_2' + dx_2')w_2 \end{aligned} \tag{3}$$

Thus $x_1 = ax_1' + cx_2'$, $x_2 = bx_2' + dx_2'$, which is to say

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}.$$

The proof of the general case has the same format, but are more cumbersome to write out.

Substitute $\mathbf{x} = P\mathbf{x}'$ and $\mathbf{y} = P\mathbf{y}'$ into Eq. (1) we have

$$P\mathbf{y}' = \mathcal{M}_T P\mathbf{x}' \tag{4}$$

or

$$\mathbf{y}' = P^{-1}\mathcal{M}_T P\mathbf{x}'. \tag{5}$$

Compare this to Eq. (2) we get the desired result. □

2 Bilinear Forms and Sesquilinear Forms

Bilinear Forms

Definition 2.1. Let V be a real vector space. A *bilinear form* $\langle\cdot,\cdot\rangle_{\mathcal{F}} : V \times V \to \mathbb{R}$ is a real-valued function defined on $V \times V$ that is linear in each variable:

$$\begin{cases} \langle v_1 + v_2, w\rangle_{\mathcal{F}} = \langle v_1, w\rangle_{\mathcal{F}} + \langle v_2, w\rangle_{\mathcal{F}} \\ \langle rv, w\rangle_{\mathcal{F}} = r\langle v, w\rangle_{\mathcal{F}} \end{cases} \qquad \begin{cases} \langle v, w_1 + w_2\rangle_{\mathcal{F}} = \langle v, w_1\rangle_{\mathcal{F}} + \langle v, w_2\rangle_{\mathcal{F}} \\ \langle v, rw\rangle_{\mathcal{F}} = r\langle v, w\rangle_{\mathcal{F}} \end{cases}$$

On Notation. The subscript notation is to remind readers that our definition of bilinear forms is more general than dot product. We are going to investigate under what conditions a bilinear form is indeed a dot product (i.e., has an orthonormal basis).

Definition 2.2. Given $(V, \mathbf{w})$, the *matrix of the form* $\mathcal{M}_{\mathcal{F}}$ is defined to be $\big(\langle w_i, w_j\rangle_{\mathcal{F}}\big)$.

Example 2.3. If V is 2-dimensional, $\mathbf{w} = (w_1, w_2)$, then the matrix of the form is

$$\begin{pmatrix} \langle w_1, w_1\rangle_{\mathcal{F}} & \langle w_1, w_2\rangle_{\mathcal{F}} \\ \\ \langle w_2, w_1\rangle_{\mathcal{F}} & \langle w_2, w_2\rangle_{\mathcal{F}} \end{pmatrix}. \tag{6}$$

Proposition 2.4. Given $v_1, v_2 \in (V, \mathbf{w})$, $v_1 = (\mathbf{x}, \mathbf{w})$, $v_2 = (\mathbf{y}, \mathbf{w})$, then

$$\langle v_1, v_2\rangle_{\mathcal{F}} = \mathbf{x}^t \mathcal{M}_{\mathcal{F}} \mathbf{y}. \tag{7}$$

Proof. 2-dimensional case should give readers more insights than using obscure sigma summations for a general proof. If $v_1 = x_1 w_1 + x_2 w_2$, $v_2 = y_1 w_1 + y_2 w_2$, then

$$\begin{aligned} \langle v_1, v_2\rangle_{\mathcal{F}} &= \langle x_1 w_1 + x_2 w_2, y_1 w_1 + y_2 w_2\rangle_{\mathcal{F}} \\ &= \langle x_1 w_1, y_1 w_1 + y_2 w_2\rangle_{\mathcal{F}} + \langle x_2 w_2, y_1 w_1 + y_2 w_2\rangle_{\mathcal{F}} \\ &= x_1 y_1\langle w_1, w_1\rangle_{\mathcal{F}} + x_1 y_2\langle w_1, w_2\rangle_{\mathcal{F}} + x_2 y_1\langle w_2, w_1\rangle_{\mathcal{F}} + x_2 y_2\langle w_2, w_2\rangle_{\mathcal{F}}. \end{aligned} \tag{8}$$

On the other hand,

$$\mathcal{M}_{\mathcal{F}}\mathbf{y} = \begin{pmatrix} \langle w_1, w_1\rangle_{\mathcal{F}} & \langle w_1, w_2\rangle_{\mathcal{F}} \\ \\ \langle w_2, w_1\rangle_{\mathcal{F}} & \langle w_2, w_2\rangle_{\mathcal{F}} \end{pmatrix} \begin{pmatrix} y_1 \\ \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1\langle w_1, w_1\rangle_{\mathcal{F}} + y_2\langle w_1, w_2\rangle_{\mathcal{F}} \\ \\ y_1\langle w_2, w_1\rangle_{\mathcal{F}} + y_2\langle w_2, w_2\rangle_{\mathcal{F}} \end{pmatrix}, \tag{9}$$

so

$$\begin{aligned} \mathbf{x}^t\mathcal{M}_{\mathcal{F}}\mathbf{y} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} y_1\langle w_1, w_1\rangle_{\mathcal{F}} + y_2\langle w_1, w_2\rangle_{\mathcal{F}} \\ \\ y_1\langle w_2, w_1\rangle_{\mathcal{F}} + y_2\langle w_2, w_2\rangle_{\mathcal{F}} \end{pmatrix} \\ &= x_1y_1\langle w_1, w_1\rangle_{\mathcal{F}} + x_1y_2\langle w_1, w_2\rangle_{\mathcal{F}} + x_2y_1\langle w_2, w_1\rangle_{\mathcal{F}} + x_2y_2\langle w_2, w_2\rangle_{\mathcal{F}}. \end{aligned} \tag{10}$$

□

Proposition 2.5 (Effect of Changing Basis on Matrix of the Form). If the matrix of the form is $\mathcal{M}_{\mathcal{F}}$ with respect to $(V, \mathbf{w})$, then the matrix of the form $\mathcal{M}'_{\mathcal{F}}$ with respect to $(V, \mathbf{w}')$ is equal to $P^t\mathcal{M}_{\mathcal{F}}P$, where P is the base change matrix.

Proof. Let $v_1 = (\mathbf{x}, \mathbf{w}) = (\mathbf{x}', \mathbf{w}')$ and $v_2 = (\mathbf{y}, \mathbf{w}) = (\mathbf{y}', \mathbf{w}')$. Then $\mathbf{x} = P\mathbf{x}'$ and $\mathbf{y} = P\mathbf{y}'$. According to Proposition 2.4,

$$\langle v_1, v_2\rangle_{\mathcal{F}} = \mathbf{x}^t\mathcal{M}_{\mathcal{F}}\mathbf{y} = (P\mathbf{x}')^t\mathcal{M}_{\mathcal{F}}(P\mathbf{y}') = \mathbf{x}'^t(P^t\mathcal{M}_{\mathcal{F}}P)\mathbf{y}'. \tag{11}$$

□

Definition 2.6. The form is said to be *symmetric* if

$$\langle v_1, v_2\rangle_{\mathcal{F}} = \langle v_2, v_1\rangle_{\mathcal{F}} \tag{12}$$

for any $v_1, v_2 \in V$. It is said to be *skew-symmetric* if $\langle v_1, v_2\rangle_{\mathcal{F}} = -\langle v_2, v_1\rangle_{\mathcal{F}}$ for any $v_1, v_2 \in V$.

Proposition 2.7. A bilinear form $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ is symmetric if and only if its matrix $\mathcal{M}_{\mathcal{F}}$ is symmetric with respect to $(V, \mathbf{w})$ for any basis $\mathbf{w}$.

Proof. First suppose $\mathcal{M}_{\mathcal{F}}$ is symmetric with respect to $(V, \mathbf{w})$, where $\mathbf{w}$ can be any basis of V. Then for any $v_1 = (\mathbf{x}_1, \mathbf{w}), v_2 = (\mathbf{y}, \mathbf{w})$,

$$\begin{aligned} \langle v_1, v_2\rangle_{\mathcal{F}} &= \mathbf{x}_1^t\mathcal{M}_{\mathcal{F}}\mathbf{y} \\ &= \left(\mathbf{x}_1^t\mathcal{M}_{\mathcal{F}}\mathbf{y}\right)^t \\ &= \mathbf{y}^t\mathcal{M}_{\mathcal{F}}^t\mathbf{x}_1 \\ &= \mathbf{y}^t\mathcal{M}_{\mathcal{F}}\mathbf{x}_1 \\ &= \langle v_2, v_1\rangle_{\mathcal{F}}. \end{aligned} \tag{13}$$

The proof of the other direction is almost the same. Suppose the form is symmetric. Given $(V, \mathbf{w})$, let $v_1 = (\mathbf{x}_1, \mathbf{w}), v_2 = (\mathbf{y}, \mathbf{w})$. Then

$$\langle v_1, v_2\rangle_{\mathcal{F}} = \mathbf{x}_1^t\mathcal{M}_{\mathcal{F}}\mathbf{y} = \mathbf{y}^t\mathcal{M}_{\mathcal{F}}\mathbf{x}_1 = \langle v_2, v_1\rangle_{\mathcal{F}}. \tag{14}$$

Since $\mathbf{x}_1^t\mathcal{M}_{\mathcal{F}}\mathbf{y}$ is a real number, it is equal to its own transpose, so $\mathbf{x}_1^t\mathcal{M}_{\mathcal{F}}\mathbf{y} = \left(\mathbf{x}_1^t\mathcal{M}_{\mathcal{F}}\mathbf{y}\right)^t$. The later is equal to $\mathbf{y}^t\mathcal{M}_{\mathcal{F}}^t\mathbf{x}_1$. We see from Eq. (14) that $\mathbf{y}^t\mathcal{M}_{\mathcal{F}}^t\mathbf{x}_1 = \mathbf{y}^t\mathcal{M}_{\mathcal{F}}\mathbf{x}_1$, which implies $\mathcal{M}_{\mathcal{F}}^t = \mathcal{M}_{\mathcal{F}}$. □

Sesquilinear Forms

Definition 2.8. Let V be a complex vector space. A *sesquilinear form* $\langle\cdot,\cdot\rangle_{\mathcal{F}} : V \times V \to \mathbb{C}$ is a complex-valued function defined on $V \times V$ that is conjugate linear in the first variable and linear in the second variable:

$$\begin{cases}\langle v_1 + v_2, w\rangle_{\mathcal{F}} = \langle v_1, w\rangle_{\mathcal{F}} + \langle v_2, w\rangle_{\mathcal{F}} \\ \langle cv, w\rangle_{\mathcal{F}} = \bar{c}\langle v, w\rangle_{\mathcal{F}}\end{cases} \qquad \begin{cases}\langle v, w_1 + w_2\rangle_{\mathcal{F}} = \langle v, w_1\rangle_{\mathcal{F}} + \langle v, w_2\rangle_{\mathcal{F}} \\ \langle v, cw\rangle_{\mathcal{F}} = c\langle v, w\rangle_{\mathcal{F}}\end{cases}$$

Given $(V, \mathbf{w})$, the matrix of the form $\mathcal{M}_{\mathcal{F}} = \big(\langle w_i, w_j\rangle_{\mathcal{F}}\big)$ is defined the same way as in the real case.

Proposition 2.9. Given $v_1, v_2 \in (V, \mathbf{w})$, $v_1 = (\mathbf{x}, \mathbf{w})$, $v_2 = (\mathbf{y}, \mathbf{w})$,

$$\langle v_1, v_2\rangle_{\mathcal{F}} = \mathbf{x}^* \mathcal{M}_{\mathcal{F}} \mathbf{y}. \tag{15}$$

Proposition 2.10 (Effect of Changing Basis on Matrix of the Form). If the matrix of the form is $\mathcal{M}_{\mathcal{F}}$ with respect to $(V, \mathbf{w})$, then the matrix of the form $\mathcal{M}'_{\mathcal{F}}$ with respect to $(V, \mathbf{w}')$ is equal to $P^* \mathcal{M}_{\mathcal{F}} P$, where P is the base change matrix.

Definition 2.11. A sesquilinear form is called *Hermitian symmetric* if

$$\langle v_1, v_2\rangle_{\mathcal{F}} = \overline{\langle v_2, v_1\rangle}_{\mathcal{F}} \tag{16}$$

for any $v_1, v_2 \in V$.

Proposition 2.12. A sesquilinear form $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ is Hermitian symmetric if and only if its matrix $\mathcal{M}_{\mathcal{F}}$ is Hermitian with respect to $(V, \mathbf{w})$ for any basis $\mathbf{w}$.

Exercise 2.13. Prove Proposition 2.9, Proposition 2.10, Proposition 2.12.

Corollary 2.14. $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ is Hermitian symmetric if and only if $\langle v, v\rangle_{\mathcal{F}} \in \mathbb{R}$ for all $v \in V$.

Proof. If the form is Hermitian symmetric, then by definition $\langle v, v\rangle_{\mathcal{F}} = \overline{\langle v, v\rangle}_{\mathcal{F}}$, thus the number $\langle v, v\rangle_{\mathcal{F}}$ is real. Conversely, if $\langle v, v\rangle_{\mathcal{F}} \in \mathbb{R}$, then given $(V, \mathbf{w})$, the matrix of the form $\mathcal{M}_{\mathcal{F}}$, and $v = (\mathbf{x}, \mathbf{w})$,

$$\langle v, v\rangle_{\mathcal{F}} = \mathbf{x}^* \mathcal{M}_{\mathcal{F}} \mathbf{x} = (\mathbf{x}^* \mathcal{M}_{\mathcal{F}} \mathbf{x})^* = \mathbf{x}^* \mathcal{M}_{\mathcal{F}}^* \mathbf{x}. \tag{17}$$

Thus we see $\mathcal{M}_{\mathcal{F}} = \mathcal{M}_{\mathcal{F}}^*$, i.e., the matrix of the form is Hermitian. by Proposition 2.12, the form is Hermitian symmetric. $\square$

Orthogonality

Definition 2.15. Let $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ be a bilinear form on a real vector space V or a sesquilinear form on a complex vector space V. We say v_1 *is orthogonal to* v_2 if

$$\langle v_2, v_1 \rangle_{\mathcal{F}} = 0. \tag{18}$$

Discussion 2.16. Here our definition of orthogonality doesn't require symmetry or Hermitian symmetry of the form. As a consequence, we have to take care of the order of v_1 and v_2.

Definition 2.17. The *orthogonal complement* of a subspace W of V, denoted by $W^{\perp}$, is the subspace of all vectors $v \in V$ that are orthogonal to every vector in W:

$$W^{\perp} := \{v \in V \mid \langle w, v \rangle_{\mathcal{F}} = 0 \text{ for all } w \in W\}. \tag{19}$$

Definition 2.18. The form is said to be *nondegenerate* if for every nonzero vector v, there is a vector v' such that $\langle v', v \rangle_{\mathcal{F}} \neq 0$. Otherwise it is said to be *degenerate*. It is said to be *nondegenerate on a subspace* W if its restriction to W is a nondegenerate form.

Proposition 2.19. The form is nondegenerate on W if and only if $W \cap W^{\perp} = \{0\}$.

Proof. If $v \in W \cap W^{\perp}$ is a nonzero vector, then since $v \in W^{\perp}$, $\langle w, v \rangle_{\mathcal{F}} = 0$ for all $w \in W$. This means that for this $v \in W$, we cannot find a vector w in W such that $\langle w, v \rangle_{\mathcal{F}} \neq 0$, thus the form is degenerate on W. On the other hand, if the form is degenerate, then for some vector $v \in W$, $\langle w, v \rangle_{\mathcal{F}} = 0$ for all $w \in W$. Thus v is also in $W^{\perp}$, and $W \cap W^{\perp}$ contains the nonzero vector v.

□

Corollary 2.20. The form is nondegenerate on V if and only if $V^{\perp} = \{0\}$.

Corollary 2.21. Let $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ be a nondegenerate bilinear form or sesquilinear form on V, and let v_1 and v_2 be vectors in V. If $\langle w, v_1 \rangle_{\mathcal{F}} = \langle w, v_2 \rangle_{\mathcal{F}}$ for all vectors w in V, then $v_1 = v_2$.

Proof. Move the right side to the left side, and using linearity in the second slot, we see

$$\langle w, v_1 - v_2 \rangle_{\mathcal{F}} = 0 \tag{20}$$

for all $w \in V$. Thus $v_1 - v_2 \in V^{\perp}$. Since the form is nondegenerate, $V^{\perp} = \{0\}$. Hence $v_1 - v_2 = 0 \Leftrightarrow v_1 = v_2$. □

Proposition 2.22. Let $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ be a bilinear form or a sesquilinear form on V, and let $\mathcal{M}_{\mathcal{F}}$ be its matrix with respect to $(V, \mathbf{w})$ for some basis $\mathbf{w}$. Let $v = (\mathbf{y}, \mathbf{w})$.

1. $\langle w, v \rangle_{\mathcal{F}} = 0$ for all $w \in V$ if and only if $\mathcal{M}_{\mathcal{F}}\mathbf{y} = 0$.

2. The form is nondegenerate if and only if $\mathcal{M}_{\mathcal{F}}$ is invertible.

Proof. 1.

$$\begin{aligned}\langle w, v\rangle_{\mathcal{F}} \text{ for all } w \in V &\Leftrightarrow \mathbf{x}^t \mathcal{M}_{\mathcal{F}} \mathbf{y} = 0 \text{ for all } \mathbf{x} \in F^n \\ &\Leftrightarrow \mathcal{M}_{\mathcal{F}} \mathbf{y} = 0.\end{aligned} \tag{21}$$

2. $\mathcal{M}_{\mathcal{F}}$ is not invertible if and only if there is $\mathbf{y} \neq \mathbf{0}$ such that $\mathcal{M}_{\mathcal{F}}\mathbf{y} = 0$ if and only if the form is degenerate by Eq. (21).

□

Theorem 2.23. Let $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ be a bilinear form or a sesquilinear form on V.

1. The form is nondegenerate on W if and only if $V = W \oplus W^{\perp}$.
2. If the form is nondegenerate on V and on W, then it is nondegenerate on $W^{\perp}$.

Proof. 1. "$\Leftarrow$" : If $V = W \oplus W^{\perp}$, then $W \cap W^{\perp} = \{0\}$, hence the form is nondegenerate by Proposition 2.19.

"$\Rightarrow$" : If the form is nondegenerate, then $W \cap W^{\perp} = \{0\}$ again by Proposition 2.19. Thus we are left to show every vector $v \in V$ can be written as $v = w + u$, where $w \in W$ and $u \in W^{\perp}$.

Let $(w_1, w_2, \ldots, w_k)$ be a basis of W. Our objective is to extend this basis of W to a basis $\mathbf{w} = (w_1, w_2, \ldots, w_k; v_1, \ldots, v_{n-k})$ of V such that $\langle w_j, v_i\rangle_{\mathcal{F}} = 0$ for $j = 1, \ldots, k$ and $i = 1, \ldots, n-k$.

The matrix of the form with respect to $(V, \mathbf{w})$ is

$$\mathcal{M}_{\mathcal{F}} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{22}$$

where A is the matrix of the form restricted to W. Since the form is nondegenerate on W, A is invertible. We are to find $v_1, \ldots, v_{n-k}$ that make $B = 0$. We do this by first using an arbitrary basis $\mathbf{w} = (w_1, w_2, \ldots, w_k; v_1, \ldots, v_{n-k})$ of V and achieve $B = 0$ by changing basis.

Let the base change matrix be

$$P = \begin{pmatrix} I & Q \\ 0 & S \end{pmatrix} \tag{23}$$

for some new basis $\mathbf{w}' = (w_1, w_2, \ldots, w_k; v'_1, \ldots, v'_{n-k})$. Then the matrix of the form $\mathcal{M}'_{\mathcal{F}}$ with respect to $(V, \mathbf{w}')$ is

$$\begin{aligned}\mathcal{M}'_{\mathcal{F}} = P^* \mathcal{M}_{\mathcal{F}} P &= \begin{pmatrix} I & 0 \\ Q^* & S^* \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & Q \\ 0 & S \end{pmatrix} \\ &= \begin{pmatrix} A & AQ + BS \\ Q^*A + S^*C & \cdot \end{pmatrix}.\end{aligned} \tag{24}$$

Set $S = I$ and $Q = -A^{-1}B$, the upper right block of $\mathcal{M}'_{\mathcal{F}}$ will become 0, as desired.

2. Exercise.

$\square$

Exercise 2.24. Prove that if both $\mathcal{M} = \begin{pmatrix} A & 0 \\ C & D \end{pmatrix}$ and its upper left block A are invertible, then the block D is invertible. Then prove (2) of Theorem 2.23.

Example 2.25. Starting with an arbitrary basis $\mathbf{w} = (w_1, w_2, \ldots, w_n)$, the matrix of the form is

$$\mathcal{M}_{\mathcal{F}} = \begin{pmatrix} \langle w_1, w_1\rangle_{\mathcal{F}} & \langle w_1, w_2\rangle_{\mathcal{F}} & \cdots & \langle w_1, w_n\rangle_{\mathcal{F}} \\ \langle w_2, w_1\rangle_{\mathcal{F}} & \langle w_2, w_2\rangle_{\mathcal{F}} & \cdots & \langle w_2, w_n\rangle_{\mathcal{F}} \\ \vdots & \vdots & \ddots & \vdots \\ \langle w_n, w_1\rangle_{\mathcal{F}} & \langle w_n, w_2\rangle_{\mathcal{F}} & \cdots & \langle w_n, w_n\rangle_{\mathcal{F}} \end{pmatrix}. \tag{25}$$

Suppose $\lambda_1 := \langle w_1, w_1\rangle_{\mathcal{F}} \neq 0$, so the form is nondegenerate on $W = \operatorname{span} w_1$. What we want to do in the above theorem is to acheive

$$B = \begin{pmatrix} \langle w_1, w_2\rangle_{\mathcal{F}} & \cdots & \langle w_1, w_n\rangle_{\mathcal{F}} \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 \end{pmatrix}.$$

We do this by setting $Q = -\frac{1}{\lambda_1}\begin{pmatrix} \langle w_1, w_2\rangle_{\mathcal{F}} & \cdots & \langle w_1, w_n\rangle_{\mathcal{F}} \end{pmatrix}$ and $S = I$. Then the matrix of the form with respect to the new basis becomes

$$\mathcal{M}'_{\mathcal{F}} = \begin{pmatrix} \langle w_1, w_1\rangle_{\mathcal{F}} & 0 & \cdots & 0 \\ \langle w'_2, w_1\rangle_{\mathcal{F}} & \langle w'_2, w'_2\rangle_{\mathcal{F}} & \cdots & \langle w'_2, w'_n\rangle_{\mathcal{F}} \\ \vdots & \vdots & \ddots & \vdots \\ \langle w'_n, w_1\rangle_{\mathcal{F}} & \langle w'_n, w'_2\rangle_{\mathcal{F}} & \cdots & \langle w'_n, w'_n\rangle_{\mathcal{F}} \end{pmatrix}. \tag{26}$$

Note that since we don't require (Hermitian) symmetry in our definition of orthogonality, there is no guarantee that the first column of $\mathcal{M}'_{\mathcal{F}}$ (except $\langle w_1, w_1\rangle_{\mathcal{F}}$) is also zero.

Definition 2.26. Let V be a vector space together with a bilinear form or a sesquilinear form $\langle\cdot,\cdot\rangle_{\mathcal{F}}$. An *orthogonal basis* $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ of V is a basis such that $\langle w_i, w_j\rangle_{\mathcal{F}} = 0$ for *all* $i \neq j, i, j \in \{1, \ldots, n\}$.

Discussion 2.27. Orthogonality is symmetric by definition.

Theorem 2.28. Let $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ be a bilinear form or a sesquilinear form on V. In addition, assume the form is symmetric or Hermitian symmetric. Then there exists an orthogonal basis of V.

Proof. If the form is identically zero, then every basis is orthogonal. So assume it is not identically zero. Then

Lemma 2.29. there is a vector v such that $\langle v, v\rangle_{\mathcal{F}} \neq 0$.

Exercise 2.30. Prove the above lemma.

Assume by induction that there exists an orthogonal basis of any proper subspace of V. Apply the above lemma and choose a vector $v_1 \in V$ with $\langle v_1, v_1\rangle_{\mathcal{F}} \neq 0$ as the first vector in our basis. Let $W = \operatorname{span} v_1$. The matrix of the form restricted to W is the 1×1 matrix whose entry is $\langle v_1, v_1\rangle_{\mathcal{F}}$. It is invertible, so the form is nondegenerate on W. By Theorem 2.23, $V = W \oplus W^{\perp}$. By our induction hypothesis, $W^{\perp}$ has an orthogonal basis $(v_2, \ldots, v_n)$.

There is no guarantee that $(v_1, v_2, \ldots, v_n)$ is an orthogonal basis of V in the sense of Definition 2.26 unless the form $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ is (Hermitian) symmetric. Without this assumption, the matrix of the form $\mathcal{M}_{\mathcal{F}}$ is at best lower triangular, but not diagonal, since we only know $\langle v_1, v_2\rangle_{\mathcal{F}} = 0, \langle v_1, v_3\rangle_{\mathcal{F}} = 0, \ldots, \langle v_1, v_n\rangle_{\mathcal{F}} = 0$, but there is no guarantee that $\langle v_2, v_1\rangle_{\mathcal{F}} = 0, \langle v_3, v_1\rangle_{\mathcal{F}} = 0, \ldots, \langle v_n, v_1\rangle_{\mathcal{F}} = 0$. □

Definition 2.31. Let $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ be a bilinear form or a Hermitian symmetric sesquilinear form (so that $\langle v, v\rangle_{\mathcal{F}} \in \mathbb{R}$) on V. The form is said to be

- *positive definite* if $\langle v, v\rangle_{\mathcal{F}} > 0$ for any nonzero $v \in V$;
- *positive semidefinite* if $\langle v, v\rangle_{\mathcal{F}} \geq 0$ for any nonzero $v \in V$;
- *negative definite* if $\langle v, v\rangle_{\mathcal{F}} < 0$ for any nonzero $v \in V$;
- *negative semidefinite* if $\langle v, v\rangle_{\mathcal{F}} \leq 0$ for any nonzero $v \in V$;
- *indefinite* if it is neither positive nor negative semidefinite.

Remark. Let $\mathcal{M}_{\mathcal{F}}$ be the matrix of the form with respect to some basis $\mathbf{w}$. The form is positive definite if and only if $\mathbf{x}^*\mathcal{M}_{\mathcal{F}}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$. Similar inequalities hold for other definitions above.

Theorem 2.32. Let V be a vector space over F, where $F = \mathbb{R}$ or $\mathbb{C}$. There is a basis $\mathbf{w}$ of V such that $\langle w_i, w_i\rangle_{\mathcal{F}} = 1$ for all $i = 1, \ldots, n$ and $\langle w_i, w_j\rangle_{\mathcal{F}} = 0$ for all $i \neq j$, if and only if the form $\langle\cdot,\cdot\rangle_{\mathcal{F}} : V \times V \to F$ is (Hermitian) symmetric and positive definite. Such a basis $\mathbf{w}$ is called an *orthonormal basis* of V.

Discussion 2.33. With respect to $(V, \mathbf{w})$, the matrix of the form becomes the identity matrix I, and thus for $v_1 = (\mathbf{x}, \mathbf{w})$ and $v_2 = (\mathbf{y}, \mathbf{w})$,

$$\langle v_1, v_2\rangle_{\mathcal{F}} = \mathbf{x}^* I \mathbf{y} = \mathbf{x}^*\mathbf{y}. \tag{27}$$

That is, the form becomes dot product on $(V, \mathbf{w})$. A matrix $\mathcal{M}$ is (Hermitian) symmetric and positive definite if and only if there is invertible matrix P such that $P^*\mathcal{M}P = I$. But P is necessarily not unitary unless $\mathcal{M}$ itself is the identity matrix. Also, if $\mathcal{M}$ is the matrix of some linear operator, then since $I = P^*\mathcal{M}P$ is not the same as $P^{-1}\mathcal{M}P$, the identity matrix I is not the same as the operator.

Proof of Theorem 2.32. Given an orthogonal basis $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ of V, use positive definiteness to adjust each positive real number $\langle w_i, w_i\rangle_{\mathcal{F}}$ to 1. □

Example 2.34. Let a two dimensional real vector space $(V, \mathbf{w})$ be given. Suppose the matrix of the form with respect to this basis is

$$\mathcal{M}_{\mathcal{F}} = \begin{pmatrix} \langle w_1, w_1\rangle_{\mathcal{F}} & \langle w_1, w_2\rangle_{\mathcal{F}} \\ \\ \langle w_2, w_1\rangle_{\mathcal{F}} & \langle w_2, w_2\rangle_{\mathcal{F}} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \tag{28}$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$, $\lambda_1, \lambda_2 > 0$. The matrix is symmetric and positive definite, so we know that it can be changed to the identity matrix. We set

$$P = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} \end{pmatrix}. \tag{29}$$

Then

$$P^t \mathcal{M}_{\mathcal{F}} P = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{30}$$

Exercise 2.35. Show that if we don't assume the form to be positive definite in Theorem 2.32, but merely (Hermitian) symmetric, then there is a basis $\mathbf{w}$ of V such that the matrix of the form is at best

$$\mathcal{M}_{\mathcal{F}} = \begin{pmatrix} I_p & & \\ & -I_q & \\ & & \mathbf{0} \end{pmatrix} \tag{31}$$

for some p, q, where I_p and I_q are respectively $p \times p$ and $q \times q$ identity matrices.

3 Application: Operators on Inner Product Space

Definition 3.1. A (Hermitian) symmetric and positive definite bilinear form or sesquilinear form is called an *inner product*. We denote such form as $\langle \cdot, \cdot \rangle_I$.

Definition 3.2. Let $(V, \mathbf{w})$ be a vector space together with an inner product $\langle \cdot, \cdot \rangle_I$, where $\mathbf{w}$ is an orthonormal basis. For $T \in \mathscr{L}(V)$, its *adjoint* $T^* \in \mathscr{L}(V)$ is the operator whose matrix is the adjoint of the matrix of T. Thus $\mathcal{M}_{T^*} := \mathcal{M}_T^*$.

Let's develop two useful properties of adjoint.

Proposition 3.3. With the definitions and notations above, $\langle Tv, w\rangle_I = \langle v, T^* w\rangle_I$ and $\langle v, Tw\rangle_I = \langle T^* v, w\rangle_I$ for all $v, w \in V$.

Proof. Let $v = (\mathbf{x}, \mathbf{w})$ and $w = (\mathbf{y}, \mathbf{w})$. Then

$$\langle Tv, w\rangle_I = (\mathcal{M}_T \mathbf{x})^* \mathbf{y} = \mathbf{x}^* \mathcal{M}_T^* \mathbf{y} = \langle v, T^* w\rangle_I, \tag{32}$$

and

$$\langle v, Tw\rangle_I = \mathbf{x}^* \mathcal{M}_T \mathbf{y} = \mathbf{x}^* (\mathcal{M}_T^*)^* \mathbf{y} = (\mathcal{M}_T^* \mathbf{x})^* \mathbf{y} = \langle T^* v, w\rangle_I. \tag{33}$$

□

Proposition 3.4. Let W be a subspace of V.

- If W is T-invariant, then $W^\perp$ is T^*-invariant;
- If W is T^*-invariant, then $W^\perp$ is T-invariant.

Proof.

$$
\begin{aligned}
\langle w, u\rangle_I &= 0 && \forall w \in W, u \in W^\perp; \\
\langle Tw, u\rangle_I &= 0 && \forall w \in W, u \in W^\perp \text{ since } W \text{ is } T\text{-invariant}; \\
\langle w, T^*u\rangle_I &= 0 && \forall w \in W, u \in W^\perp \text{ by Proposition 3.3.}
\end{aligned}
$$

Thus $T^*u \in W^\perp\ \ \forall u \in W^\perp \implies W^\perp$ is T^*-invariant. Since $(T^*)^* = T$, substitute T^* into the first statement we get the second. □

Definition 3.5. Let $(V, \mathbf{w})$ be a complex vector space with inner product $\langle\cdot,\cdot\rangle_I$, where $\mathbf{w}$ is an orthonormal basis. Let $T \in \mathscr{L}(V)$.

- T is *Hermitian* (or *self-adjoint*) if $T = T^*$;
- T is *normal* if $TT^* = T^*T$;
- T is *unitary* if $T^*T = I$.

Proposition 3.6. Let $v, w \in V$ be arbitrary. With the definitions as above,

- T is Hermitian if and only if $\langle Tv, w\rangle_I = \langle v, Tw\rangle_I$;
- T is normal if and only if $\langle Tv, Tw\rangle_I = \langle T^*v, T^*w\rangle_I$;
- T is unitary if and only if $\langle Tv, Tw\rangle_I = \langle v, w\rangle_I$.

Proof.

1. If $T = T^*$, then $\langle Tv, w\rangle_I = \langle v, Tw\rangle_I$ by Proposition 3.3. Conversely, if $\langle Tv, w\rangle_I = \langle v, Tw\rangle_I$, then $\langle v, Tw\rangle_I = \langle v, T^*w\rangle_I$ for all $v, w \in V$, again by Proposition 3.3. Since the form is nondegenerate (its matrix is the identity matrix, which is invertible; see Proposition 2.22), $Tw = T^*w$ for all $w \in V$ by Corollary 2.21. Thus $T = T^*$.

2. - Substitute T^*v for v into the first equation of Proposition 3.3, we have

   $$\langle TT^*v, w\rangle_I = \langle T^*v, T^*w\rangle_I;$$

   - Substitute Tv for v into the second equation of Proposition 3.3, we have

   $$\langle Tv, Tw\rangle_I = \langle T^*Tv, w\rangle_I.$$

   Then

- $TT^* = T^*T \implies \langle Tv, Tw\rangle_I = \langle T^*v, T^*w\rangle_I$;
- $\langle Tv, Tw\rangle_I = \langle T^*v, T^*w\rangle_I \implies \langle T^*Tv, w\rangle_I = \langle TT^*v, w\rangle_I \implies T^*Tv = TT^*v$ by Corollary 2.21 $\implies TT^* = T^*T$.

3. Substitute Tw for w into the first equation of Proposition 3.3 and proceed as above.

$\square$

Discussion 3.7 (On definition of Adjoint). Our definition of adjoint of an operator is independent up to orthonormal basis:

- Given $(V, \mathbf{w})$, $T \in \mathscr{L}(V)$, its matrix $\mathcal{M}_T$, T^* is defined by $\mathcal{M}_{T^*} := (\mathcal{M}_T)^*$;
- Given $(V, \mathbf{w}')$, $T \in \mathscr{L}(V)$, its matrix $\mathcal{M}'_T$, T^* is defined by $\mathcal{M}'_{T^*} := (\mathcal{M}'_T)^*$.

Since $\mathcal{M}'_{T^*} = P^{-1}\mathcal{M}_{T^*}P$, where P is the base change matrix, the right sides of the two equations above should also satisfy this relation, i.e.,

$$(\mathcal{M}'_T)^* = (P^{-1}\mathcal{M}_T P)^* = P^*\mathcal{M}_T^*(P^{-1})^* = P^{-1}\mathcal{M}_T^* P = (\mathcal{M}_T^*)'. \tag{34}$$

This hold if and only if $P^* = P^{-1}$, i.e., P is unitary and thus $\mathbf{w}'$ is again orthonormal (see Proposition 3.6). Since we always use orthonormal basis when we are dealing with inner product, our definition of adjoint is suffice.

Proposition 3.8. Let T be a normal operator. If $Tv = \lambda v$, then $T^*v = \bar{\lambda}v$.

Proof. Since T is normal, $T - \lambda I$ is also normal (Exercise 3.9). Then since $(T - \lambda I)v = 0$, we have

$$\begin{aligned} 0 &= \langle (T - \lambda I)v, (T - \lambda I)v\rangle_I \\ &= \langle (T - \lambda I)^*v, (T - \lambda I)^*v\rangle_I \\ &= \langle (T^* - \bar{\lambda} I)v, (T^* - \bar{\lambda} I)v\rangle_I. \end{aligned} \tag{35}$$

Since the form is positive definite, $\langle v, v\rangle_I = 0$ implies that $v = 0$. This shows that $T^*v = \bar{\lambda}v$. $\square$

Exercise 3.9. Verify that if T is normal, then $T - \lambda I$ is normal.

Theorem 3.10 (Spectral Theorem for Normal Operators). Let $(V, \mathbf{w})$ be a vector space together with an inner product $\langle \cdot, \cdot\rangle_I$, where $\mathbf{w}$ is an orthonormal basis. If $T \in \mathscr{L}(V)$ is a normal operator with matrix $\mathcal{M}_T$, then

- there is a unitary matrix P such that $P^*\mathcal{M}_T P$ is diagonal; or equivalently,
- there is an orthonormal basis $\mathbf{w}'$ such that all of $\mathbf{w}'$ are at the same time eigenvectors of T.

Proof. We prove the second statement. Pick an eigenvector v of T with eigenvalue λ, and normalize its length to 1 (that is, make $\langle v, v\rangle_I = 1$. Note that $Tv = \lambda v$ still hold). Then $Tv = \lambda v$ and $T^*v = \bar{\lambda} v$ by Proposition 3.8. Let $W := \operatorname{span} v$. Then $V = W \oplus W^\perp$ by Theorem 2.23 since the form is nondegenerate on W ($\langle v, v\rangle_I = 1 \neq 0$). Since W is T^*-invariant, $W^\perp$ is T-invariant by Proposition 3.4. Thus T is normal on $W^\perp$. Then we may assume by induction that $W^\perp$ has an orthonormal basis of eigenvectors, say $(v_2, \dots, v_n)$. Then $(v, v_2, \dots, v_n)$ is an orthonormal basis of eigenvectors of T. □

Discussion 3.11. The above spectral theorem for normal operators applies in particular to Hermitian operators and unitary operators, since they are both normal. Thus we see that for a matrix $\mathcal{M}_T$ of a Hermitian operator, there is a unitary matrix P such that $P^*\mathcal{M}_T P$ is diagonal. See Discussion 2.33 for the case of forms.

4 Application: Square Root, Polar Decomposition and SVD

Proposition 4.1. Let $(V, \mathbf{w})$ be a complex vector space together with an inner product $\langle\cdot,\cdot\rangle_I$, where $\mathbf{w}$ is an orthonormal basis as usual. Given $T \in \mathscr{L}(V)$,

$$\langle v, w\rangle_T := \langle v, Tw\rangle_I \tag{36}$$

defines a new sesquilinear form on V. $T \in \mathscr{L}(V)$ is Hermitian if and only if the form $\langle\cdot,\cdot\rangle_T$ is Hermitian symmetric.

Exercise 4.2. Prove the above Proposition.

Exercise 4.3. Prove that with respect to the new form $\langle\cdot,\cdot\rangle_T$, $V^\perp = \operatorname{null} T$. Thus the form is nondegenerate if and only if T is invertible.

Corollary 4.4. Every eigenvalue of a Hermitian operator is real.

Proof. By Proposition 4.1 and Corollary 2.14, $\langle v, v\rangle_T \in \mathbb{R}$ for any $v \in V$. Thus if $Tv = \lambda v$, then

$$\langle v, v\rangle_T = \langle v, Tv\rangle_I = \langle v, \lambda v\rangle_I = \lambda\langle v, v\rangle_I \in \mathbb{R} \Rightarrow \lambda \in \mathbb{R}. \tag{37}$$

□

Definition 4.5. An operator $R \in \mathscr{L}(V)$ is called a *square root* of an operator $T \in \mathscr{L}(V)$ if $R^2 = T$.

Definition 4.6. An operator $T \in \mathscr{L}(V)$ is called *positive* if the form $\langle\cdot,\cdot\rangle_T$ is Hermitian symmetric and positive semidefinite.

Theorem 4.7. Every positive operator $T \in \mathscr{L}(V)$ has a unique positive square root.

Proof. First observe that if $\langle v, v\rangle_T \geq 0$ for all $v \in V$ (i.e., the form is positive semidefinite), then all eigenvalues of T are nonnegative (see the proof of Corollary 4.4). Also, with respect to some orthonormal basis $\mathbf{w}$, the matrix of T is a diagonal matrix $\mathcal{M}_T$ consisting of these nonnegative eigenvalues:

$$\mathcal{M}_T = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}. \tag{38}$$

Thus we can take square roots along the diagonal and obtain a matrix $\mathcal{M}_R$:

$$\mathcal{M}_R = \begin{pmatrix} \sqrt{\lambda}_1 & & & \\ & \sqrt{\lambda}_2 & & \\ & & \ddots & \\ & & & \sqrt{\lambda}_n \end{pmatrix}. \tag{39}$$

The operator $R \in \mathscr{L}(V)$ is defined by $Rw_j = \sqrt{\lambda}_j w_j$ for $j = 1, \ldots, n$. Thus one can see $\mathcal{M}_R^2 = \mathcal{M}_T$ and $R^2 = T$. Hermitian symmetry and positive semidefiniteness are obvious.

Uniqueness is just the fact that every nonnegative number has a unique (nonnegative) square root. To prove it formally, suppose there is another positive operator $R' \in \mathscr{L}(V)$ such that $R'^2 = T$. It has an orthonormal basis $\mathbf{v}$, and suppose $R'v_j = \alpha_j v_j$ for some v_j in $\mathbf{v}$ and $\alpha_j \geq 0$. Then $Tv_j = R'^2 v_j = \alpha_j^2 v_j$ so that v_j is also an eigenvector of T. By our construction it is also an eigenvector of R, so $Rv_j = \beta_j v_j$ for some $\beta_j \geq 0$. Then $Tv_j = R^2 v_j = \beta_j^2 v_j$, and therefore $\alpha_j^2 = \beta_j^2$. Since both α_j and β_j are nonnegative, this implies that $\alpha_j = \beta_j$. The same situation is true for all basis elements in $\mathbf{v}$, so $R = R'$ as desired. $\square$

Theorem 4.8 (Polar Decomposition). Every operator $T \in \mathscr{L}(V)$ can be written as $T = UR$ where $U \in \mathscr{L}(V)$ is unitary and $R^2 = T^*T$.

Remark. Since T^*T is a positive operator, as can be easily checked, R is uniquely determined by Theorem 4.7. In the proof below, we shall define U separately on each component of $V = R(V) \oplus [R(V)]^\perp$.

Proof. For $v \in V$,

$$\langle Rv, Rv\rangle_I = \langle R^2 v, v\rangle_I = \langle T^*Tv, v\rangle_I = \langle Tv, Tv\rangle_I. \tag{40}$$

Thus if we define $U' : R(V) \to T(V)$ by

$$U'(Rv) = Tv, \tag{41}$$

then by Eq. (40), $\langle U'(Rv), U'(Rw)\rangle_I = \langle Tv, Tw\rangle_I = \langle Rv, Rw\rangle_I$, so U' is indeed unitary on $R(V)$. We use definiteness of $\langle\cdot,\cdot\rangle_I$ to show U' is well defined. If $Rv_1 = Rv_2$, then we want their images Tv_1 and Tv_2 to be equal. This is

$$Tv_1 - Tv_2 = 0.$$

Using definiteness of $\langle\cdot,\cdot\rangle_I$, it suffices to show

$$\langle Tv_1 - Tv_2, Tv_1 - Tv_2\rangle_I = 0,$$

which is

$$\langle T(v_1 - v_2), T(v_1 - v_2)\rangle_I = 0.$$

According to Eq. (40), the above expression is equal to

$$\begin{aligned}&\langle R(v_1 - v_2), R(v_1 - v_2)\rangle_I \\ &= \langle Rv_1 - Rv_2, Rv_1 - Rv_2\rangle_I \\ &= 0.\end{aligned}$$

Thus U' is indeed well defined.

Exercise 4.9. Check that U' is indeed linear.

Reverse the above verifications that used to show U' was well defined, we see that U' is injective. Thus $\dim \operatorname{null}(U) = 0$ and therefore $\dim R(V) = \dim T(V)$. Then $\dim[R(V)]^\perp = \dim[T(V)]^\perp$ and we can choose orthonormal basis $\mathbf{u} = (u_1, \ldots, u_m)$ of $[R(V)]^\perp$ and $\mathbf{v} = (v_1, \ldots, v_m)$ of $[T(V)]^\perp$. Define $U'' : [R(V)]^\perp \to [T(V)]^\perp$ by $(\mathbf{x}, \mathbf{u}) \mapsto (\mathbf{x}, \mathbf{v})$. If $z_1 = (\mathbf{x}_1, \mathbf{u})$ and $z_2 = (\mathbf{x}_2, \mathbf{u})$, then

$$\langle U''z_1, U''z_2\rangle_I = \langle(\mathbf{x}_1, \mathbf{v}), (\mathbf{x}_2, \mathbf{v})\rangle_I = \mathbf{x}_1^*\mathbf{x}_2 = \langle(\mathbf{x}_1, \mathbf{u}), (\mathbf{x}_2, \mathbf{u})\rangle_I = \langle z_1, z_2\rangle_I. \tag{42}$$

Write each element $v \in V = R(V) \oplus [R(V)]^\perp$ as $v = w + u$ where $w \in R(V)$ and $u \in [R(V)]^\perp$. Define $U \in \mathscr{L}(V)$ by

$$Uv = U'w + U''u. \tag{43}$$

Then for all $v \in V$, $URv = U'(Rv) = Tv$, so $T = UR$ as desired. We are left to check that

U is indeed unitary. Given $v_1, v_2 \in V$, where $v_1 = w_1 + u_1$ and $v_2 = w_2 + u_2$, we have

$$\begin{aligned}
&\langle U v_1, U v_2\rangle_I \\
&= \langle U(w_1 + u_1), U(w_2 + u_2)\rangle_I \\
&= \langle U' w_1 + U'' u_1, U' w_2 + U'' u_2\rangle_I \\
&= \langle U' w_1, U' w_2\rangle_I + \langle U'' u_1, U'' u_2\rangle_I \\
&= \langle w_1, w_2\rangle_I + \langle u_1, u_2\rangle_I \\
&= \langle w_1 + u_1, w_2 + u_2\rangle_I \\
&= \langle v_1, v_2\rangle_I .
\end{aligned} \tag{44}$$

□

Definition 4.10. For $T \in \mathscr{L}(V), T = UR$, the eigenvalues of R are called the *singular values* of T.

Theorem 4.11 (Singular Value Decomposition, SVD). For every $T \in \mathscr{L}(V)$, there are two orthonormal basis $\mathbf{w} = (w_1, \dots, w_n)$ and $\mathbf{v} = (v_1, \dots, v_n)$ such that

$$T w_j = s_j v_j \tag{45}$$

for each $j = 1, \dots, n$, where $s_1, \dots, s_n$ are singular values of T.

Proof. There is an orthonormal basis $\mathbf{w} = (w_1, \dots, w_n)$ of V such that $R w_j = s_j w_j$ for $j = 1, \dots, n$. Then $T w_j = U R w_j = s_j U w_j$. Since U is unitary, it preserves orthonormal basis, thus $(v_1, \dots, v_n) := (U w_1, \dots, U w_n)$ is again an orthonormal basis. □

5 Application: Quadratic Forms and Optimization

Definition 5.1. A *quadratic form* $Q_A : \mathbb{R}^n \to \mathbb{R}$ is a homogeneous polynomial of degree 2, i.e.,

$$Q_A(x_1, x_2, \dots, x_n) = \sum_{i \le j} a_{ij} x_i x_j \tag{46}$$

for some $a_{ij} \in \mathbb{R}, i, j = 1, \dots, n$.

Proposition 5.2. Every quadratic form $Q_A : \mathbb{R}^n \to \mathbb{R}$ can be represented by $Q_A(\mathbf{x}) = \mathbf{x}^t A \mathbf{x}$ for some real symmetric $n \times n$ matrix. Conversely, for every real symmetric matrix A we can associate a quadratic form $Q_A(\mathbf{x}) = \mathbf{x}^t A \mathbf{x}$.

Proof. The last statement is obvious. For the first statement we look at two dimensional case. Given a quadratic form $Q_A(\mathbf{x}) = Q_A(x_1, x_2) = a x_1^2 + b x_1 x_2 + c x_2^2$,

$$Q_A(x_1, x_2) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}^t A \mathbf{x}. \tag{47}$$

□

Corollary 5.3. Let $\langle\cdot,\cdot\rangle_I$ be dot product on $\mathbb{R}^n$. Then

$$Q_A(\mathbf{x}) = \mathbf{x}^t A\mathbf{x} = \langle\mathbf{x}, A\mathbf{x}\rangle_I = \langle\mathbf{x},\mathbf{x}\rangle_A. \tag{48}$$

Definition 5.4. A quadratic form $Q_A : \mathbb{R}^n \to \mathbb{R}$ is said to be

- *positive definite* if $Q_A(\mathbf{x}) > 0$ for any $\mathbf{x} \neq \mathbf{0}, \mathbf{x} \in \mathbb{R}^n$;
- *positive semidefinite* if $Q_A(\mathbf{x}) \geq 0$ for any $\mathbf{x} \neq \mathbf{0}, \mathbf{x} \in \mathbb{R}^n$;
- *negative definite* if $Q_A(\mathbf{x}) < 0$ for any $\mathbf{x} \neq \mathbf{0}, \mathbf{x} \in \mathbb{R}^n$;
- *negative semidefinite* if $Q_A(\mathbf{x}) \leq 0$ for any $\mathbf{x} \neq \mathbf{0}, \mathbf{x} \in \mathbb{R}^n$;
- *indefinite* if it is neither positive nor negative semidefinite.

We say a real symmetric matrix A has the above property if its associated quadratic form (Proposition 5.2) is so.

In view of Corollary 5.3, definiteness for quadratic forms, matrices and bilinear forms are really the same thing.

We see from Definition 5.4 that if Q_A is positive definite, then $\mathbf{x} = \mathbf{0}$ minimizes $Q_A(\mathbf{x})$ on $\mathbb{R}^n$; it is called a *global minimizer* of Q_A. Similarly, if Q_A is negative definite, then $\mathbf{x} = \mathbf{0}$ maximizes $Q_A(\mathbf{x})$ on $\mathbb{R}^n$; it is called a *global maximizer* of Q_A. Quadratic forms give us an opportunity to visualize definiteness (Fig. 1).

Exercise 5.5. Work out the 5 symmetric matrices of the conrresponding quadratic forms in Fig. 1.

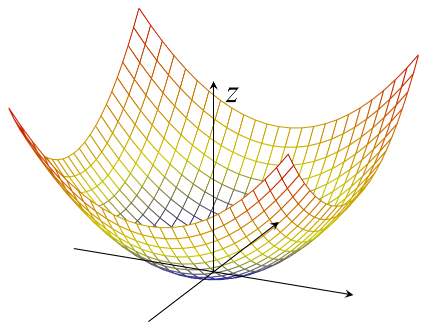
Definition 5.6. Let $f : A \to \mathbb{R}$ be twice continuously differentiable on $A \subset \mathbb{R}^n$. It's *Hessian matrix* at $\mathbf{a} \in A$ is

$$\mathbf{H}_f(\mathbf{a}) = \begin{pmatrix} D_{11}f(\mathbf{a}) & D_{12}f(\mathbf{a}) & \cdots & D_{1n}f(\mathbf{a}) \\ D_{21}f(\mathbf{a}) & D_{22}f(\mathbf{a}) & \cdots & D_{2n}f(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1}f(\mathbf{a}) & D_{n2}f(\mathbf{a}) & \cdots & D_{nn}f(\mathbf{a}) \end{pmatrix}. \tag{49}$$

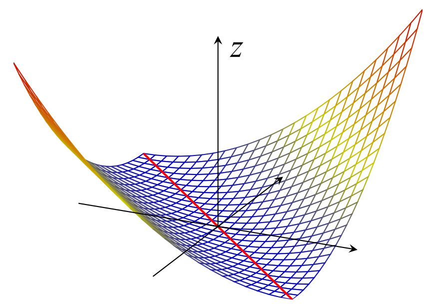
Theorem 5.7 (Second Partial Derivative Test). Let $f : \mathbb{R}^n \to \mathbb{R}$ be smooth. Suppose $\mathbf{a} \in \mathbb{R}^n$ is a critical point of f.

- If $\mathbf{H}_f(\mathbf{a})$ is positive definite, then $\mathbf{a}$ is a local minimum of f;
- If $\mathbf{H}_f(\mathbf{a})$ is negative definite, then $\mathbf{a}$ is a local maximum of f.

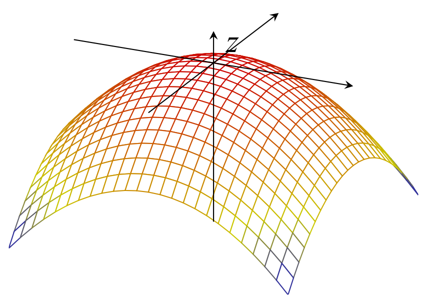
$Q(x_1, x_2) = x_1^2 + x_2^2$ is positive definite



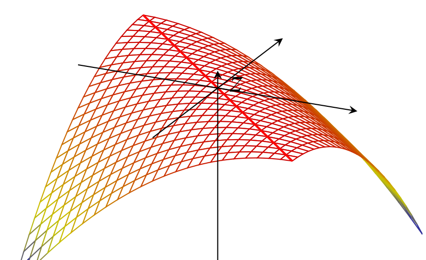
$Q(x_1, x_2) = (x_1 + x_2)^2$ is positive semidefinite



$Q(x_1, x_2) = -(x_1^2 + x_2^2)$ is negative definite



$Q(x_1, x_2) = -(x_1 + x_2)^2$ is negative semidefinite



$Q(x_1, x_2) = x_1^2 - x_2^2$ is indefinite

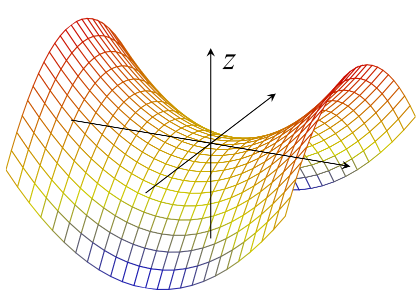


Figure 1: Quadratic forms

Sketch of Proof. Taylor's Theorem says that:

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{a}) + \sum_{i=1}^{n}(x_i - a_i)\left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{a}} + \frac{1}{2}\sum_{i,j=1}^{n}(x_i - a_i)(x_j - a_j)\left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{a}} + \cdots \\ &= f(\mathbf{a}) + \langle \mathbf{x}-\mathbf{a}, \nabla f(\mathbf{a})\rangle_I + \frac{1}{2}\langle \mathbf{H}_f(\mathbf{a})(\mathbf{x}-\mathbf{a}), \mathbf{x}-\mathbf{a}\rangle_I + \cdots \end{aligned} \tag{50}$$

Since $\mathbf{a}$ is a critical point of f, $\nabla f(\mathbf{a}) = \mathbf{0}$, so that

$$f(\mathbf{x}) - f(\mathbf{a}) = \frac{1}{2}\langle \mathbf{H}_f(\mathbf{a})(\mathbf{x}-\mathbf{a}), \mathbf{x}-\mathbf{a}\rangle_I + \cdots \tag{51}$$

If say, $\mathbf{H}_f(\mathbf{a})$ is positive definite, then $\langle \mathbf{H}_f(\mathbf{a})(\mathbf{x}-\mathbf{a}), \mathbf{x}-\mathbf{a}\rangle_I > 0$ for $\mathbf{x} \neq \mathbf{a}$, so that $f(\mathbf{x}) \geq f(\mathbf{a})$ for $\mathbf{x}$ sufficient close to $\mathbf{a}$. $\square$

We now give a criterion for a real symmetric matrix (hence for a symmetric bilinear form and in particular quadratic form) to be positive or negative definite. Let A be a real symmetric $n \times n$ matrix.We let A_k denote the upper left $k \times k$ block of A. Thus if we let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \tag{52}$$

Then $A_1 = \begin{pmatrix} a_{11} \end{pmatrix}$, $A_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, etc.

Theorem 5.8. Let A be a real $n \times n$ symmetric matrix. Then

- A is positive definite if and only if $\det A_k > 0$ for all $k = 1, \ldots, n$;
- A is negative definite if and only if $\det A_k$ alternating signs as $\det A_1 < 0, \det A_2 > 0, \det A_3 < 0, \ldots$

We first prove two lemmas.

Lemma 5.9. If A is positive or negative definite, then A is nonsingular.

Proof. If A is singular, then there is $\mathbf{x} \neq 0$ such that $A\mathbf{x} = \mathbf{0}$. But then $\mathbf{x}^t A\mathbf{x} = 0$, contradicting to the definiteness of A. $\square$

Lemma 5.10. If A is real symmetric and Q is nonsingular, then Q^tAQ is symmetric, and A is positive (negative) definite if and only if Q^tAQ is positive (negative) definite.

Proof. 1. $A^t = A \Longrightarrow (Q^tAQ)^t = Q^tA^tQ = Q^tAQ$.

2. "$\Rightarrow$": Since $Q\mathbf{z} \neq 0$ for any $\mathbf{z} \neq 0$, we have

$$0 < (Q\mathbf{z})^t A(Q\mathbf{z}) = \mathbf{z}^t (Q^t AQ)\mathbf{z}. \tag{53}$$

Thus $Q^t AQ$ is positive definite.

"$\Leftarrow$": Since Q is nonsingular, for every $\mathbf{x} \neq 0$ there is $\mathbf{y} \neq 0$ such that $\mathbf{x} = Q\mathbf{y}$. Then

$$\mathbf{x}^t A\mathbf{x} = (Q\mathbf{y})^t A(Q\mathbf{y}) = \mathbf{y}^t (Q^t AQ)\mathbf{y} > 0. \tag{54}$$

This shows that A is positive definite.

□

Proof of Theorem 5.8. We prove the first statement. The proof of the second statement is similar and is left as exercise.

Partition A as

$$A = \left(\begin{array}{c|c} A_{n-1} & \mathbf{a} \\ \hline \mathbf{a}^t & a_n \end{array}\right), \tag{55}$$

where

$$\mathbf{a} = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{(n-1)n} \end{pmatrix}.$$

Decompose A as

$$\begin{aligned} A &= \left(\begin{array}{c|c} I_{n-1} & \mathbf{0} \\ \hline (A_{n-1}^{-1}\mathbf{a})^t & 1 \end{array}\right) \left(\begin{array}{c|c} A_{n-1} & \mathbf{0} \\ \hline \mathbf{0} & d \end{array}\right) \left(\begin{array}{c|c} I_{n-1} & A_{n-1}^{-1}\mathbf{a} \\ \hline \mathbf{0} & 1 \end{array}\right) \\ &= Q^t BQ. \end{aligned} \tag{56}$$

Exercise 5.11. Work out the expression for d.

We see that $\det Q = \det Q^t = 1$, and $\det B = d \cdot \det A_{n-1}$. Therefore

$$\det A = \det(Q^t BQ) = (\det Q^t)(\det B)(\det Q) = d \cdot \det A_{n-1}. \tag{57}$$

"$\Leftarrow$": This is obviously true for $n = 1$. Suppose by induction that it is also true for any $(n-1) \times (n-1)$ symmetric matrix. Since $\det A_{n-1} > 0$ and $\det A > 0$, we have $d > 0$ in Eq. (57).

Let $\mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$. Single out the last element of $\mathbf{x}$ as

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_{n-1} \\ x_n \end{pmatrix},$$

where $\mathbf{x}_{n-1}$ is an $(n-1)$-vector. Then

$$\begin{aligned} \mathbf{x}^t B\mathbf{x} &= \begin{pmatrix} \mathbf{x}_{n-1} & x_n \end{pmatrix} \left(\begin{array}{c|c} A_{n-1} & \mathbf{0} \\ \hline \mathbf{0} & d \end{array}\right) \begin{pmatrix} \mathbf{x}_{n-1} \\ x_n \end{pmatrix} \\ &= \mathbf{x}_{n-1}^t A_{n-1} \mathbf{x}_{n-1} + d x_n^2. \end{aligned} \tag{58}$$

Since A_{n-1} is positive definite by induction hypothesis and $d > 0$, the last expression is strictly positive. Therefore B is positive definite. By Lemma 5.10, A is positive definite, as desired.

"$\Rightarrow$": Suppose A is positive definite. Then all $A_k, k = 1, \ldots, n$ are positive definite: just take $\mathbf{x} = \begin{pmatrix} \mathbf{x}_k \\ \mathbf{0} \end{pmatrix}$. Thus we can suppose by induction that all of $A_k, k = 1, \ldots, n-1$ have positive determinant. We need only to show A also has positive determinant, which according to Eq. (57) reduced to show $d > 0$. Let $\mathbf{x} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}$. Then

$$\mathbf{x}^t B\mathbf{x} = d > 0 \tag{59}$$

since B is positive definite by Lemma 5.10.

□

Exercise 5.12. Prove the second statement of Theorem 5.8. (Hint: The proof is almost the same as in the case of positive definiteness. Go through the proof again and see where should be modified.)