线性代数

Linear Algebra

涂轩丞

浙江大学

目录

目录

1. Orthogonal Complements

Theorem(Theorem for null space and range of T^*). Suppose $T \in \mathcal{L}(V, W)$, then

(i) $NT^* = (RT)^\perp$.

(ii) $RT^* = (NT)^\perp$.

Proof:

■

2. Singular value decomposition

For a linear map $T \in \mathcal{L}(V, W)$, we could decompose it as we have for self-adjoint operator or normal operator.

Recall the important Riesz representation theorem in inner product space.

Theorem(Riesz representation theorem). Assume V is finite-dimensional and φ is a linear functional on V, then there exists a unique vector $v \in V$ such that

$$\varphi(u) = \langle v, u \rangle, \quad \forall u \in V. \tag{1}$$

Proof:

■

In functional analysis, we have actually a similar result for infinite-dimensional spaces.

The following lemma of T^*T is necessary.

Theorem(Lemma: Properties of T^*T). Suppose $T \in \mathcal{L}(V, W)$.

(i) T^*T is a self-adjoint operator on V. We could also check $\mathbb{T}^*$ is a self-adjoint operator on W.

(ii) $NT^*T = NT$.

(iii) $RT^*T = RT^*$.

(iv) dimension. Dim $RT =$ Dim $RT^* =$ Dim RT^*T.

Proof: (i) by definition.

$$\langle T^*Tv, w\rangle = \langle Tv, Tw\rangle = \langle v, T^*Tw\rangle \Rightarrow T^*T = (T^*T)^*. \tag{2}$$

(ii) $NT \subset NT^*T$ is apparent. Assume $v \in NT^*T$, $T^*Tv = 0$, so $\langle v, T^*Tv\rangle = 0$, so $\langle Tv, Tv\rangle = |Tv|^2 = 0$, which means $Tv = 0$.

(iii) $RT^*T \subset RT^*$ is apparent. For another direction, we use (ii) $RT^*T = (NT^*T)^\perp = (NT)^\perp = RT^*$.

(iv) Use fundamental theorem of linear maps.

■

Definition(Definition of singular value). Assume a linear operator $T \in \mathcal{L}(V, W)$, the singular values of T are defined as the nonnegative square roots of the eigenvalues of T^*T, listed in decreasing order.

Theorem((SVD) Singular value decomposition). Assume a linear operator $T \in \mathcal{L}(V, W)$, with its positive singular values $s_1, ..., s_r$. Then there exists orthonormal lists $e_1, ..., e_r \subset V$, $f_1, ..., f_r \subset W$, such that

$$Tv = \sum_{k=1}^{r} s_k \langle v, e_k \rangle f_k. \tag{3}$$

Proof: Here we denote that V and W is finite-dimensional. And the proof is constructive. This method also gives info about the eigenvectors construction.

Let $s_1, ..., s_n$ to be the singular value of $T(\text{Dim } V = n)$, where $s_{\{r+1\}}, ..., s_n$ are zero singular values.

- Apply spectral theorem to T^*T and there exists orthonormal basis $e_1, ..., e_n \subset V$, such that

$$T^*Te_k = s_k^2 e_k, \quad k = 1, ..., n. \tag{4}$$

- Define $f_k = \frac{Te_k}{s_k}$ for $k = 1, ..., r$.

this is actually orthonormal basis in W. This is also inspired by $T = W\Sigma V^T$, so $TV = W\Sigma$, so $TV\Sigma^{-1} = W$, which shows a relationship of basis from V to W space.

- Prove the proposition by expressing v in the constructed orthonormal basis

$$\begin{aligned} Tv &= T\left(\sum_{k=1}^{n} \langle v, e_k\rangle e_k\right) \\ &= \sum_{k=1}^{n} \langle v, e_k\rangle Te_k \\ &= \sum_{k=1}^{r} \langle v, e_k\rangle s_k f_k \end{aligned} \tag{5}$$

for $k \geq r$, $Te_k = 0$ because $T^*Te_k = 0 \cdot e_k$ and Property of self-adjoint T^*T (ii).

We could also check that the matrix with respect to basis $\{e_k\}_{1\leq k\leq r}$ and $\{f_k\}_{1\leq k\leq r}$ which should be extended.

Note we have $\{e_k\}_{1\leq k\leq n}$, and from the above proof we have $Te_k = s_k f_k$ for $k \leq r$ and 0 for $k > r$. We shall extend $\{f_k\}_{1\leq k\leq r}$ to $\{f_k\}_{1\leq k\leq m}$(Dim $W = m$) by utilizing NT^*. This is because we want to solve $R(T)^\perp$, which equals NT^* by Theorem for null space and range of T^*. (Readers should double check the dimension of NT^*, which is $m - r$, for Dim $RT = r$.)

■

Theorem(Matrix version of SVD, a compact SVD form). Assume A is an m-by-n matrix of rank $r \geq 1$. Then there exists an m-by-r matrix W with orthogonal columns, an r-by-r diagonal matrix Σ with positive numbers on the diagonal, and an n-by-r matrix V with orthonormal columns such that

$$A = W \Sigma V^*. \tag{6}$$

Proof: Let $T : \mathbb{F}^n \to \mathbb{F}^m$ whose matrix with respect to the standard basis equals A. From the above proof of the SVD theorem, we have Dim $RT = r$ and

$$Tv = \sum_{k=1}^{r} s_k \langle v, e_k \rangle f_k. \tag{7}$$

we make use of the above structure. Let

W to be the m-by-r matrix whose columns are $f_1, \ldots, f_r$,

Σ to be the r-by-r diagonal matrix Σ with entries $s_1, \ldots, s_r$,

V to be the n-by-r matrix whose columns are $e_1, \ldots, e_r$.

Choose u_k, a standard base of $\mathbb{F}^m$, then apply this matrix

$$(AV - W\Sigma)u_k = Ae_k - W s_k u_k = s_k f_k - s_k f_k = 0. \tag{8}$$

so $AV = W\Sigma$, multiply both sides by V^* and we have $A = W\Sigma V^*$. But we have to be careful.

Here actually $VV^* = I$ does not hold absolutely. We have to argue as follows. If $k \leq r$, $V^* e_k = u_k$, so $VV^* e_k = e_k$. Thus $AVV^* v = Av$ for all $v \in \text{span}(e_1, ..., e_m)$. For $v \in \text{span}(e_1, ..., e_m)^{\perp}$, we have $Av = 0$ and $V^* v = 0$, so we also have $AVV^* v = Av = 0$.

■

Proof: Another version.

Denote $S = \text{diag}(s_1, ..., s_r)$, $\Sigma = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}_{n \times n}$, $V_1 = (e_1, ..., e_r)$, $V_2 = \left(e_{\{r+1\}}, ..., e_n\right)$ where the orthonormal basis in V_2 is with respect to eigenvalue 0. Notice

$$\begin{aligned} A^* A V_1 &= S^2 V_1 = V_1 S^2 \\ V_1^* A^* A V_1 &= S^2 \\ \Rightarrow S^{-1} V_1^* A^* A V_1 S^{-1} &= I_r. \end{aligned} \tag{9}$$

define $W_1 = AV_1 S^{-1}$, we have $W_1^* W_1 = I_r$. As for V_2, we have $A^* A V_2 = V_2 0^2 = 0$, So $V_2^* A^* A V_2 = 0$, $AV_2 = 0$.

Choose W_2 to be an orthogonal complement of W_1, which is actually calculated from NA^*, $A^* W_2 = 0$. So let $W = (W_1, W_2)$, we have

$$W^T AV = \begin{pmatrix} W_1^T AV_1 & W_1^T AV_2 \\ W_2^T AV_1 & W_2^T AV_2 \end{pmatrix} = \begin{pmatrix} W_1^T AV_1 & 0 \\ W_2^T AV_1 & 0 \end{pmatrix} \quad \text{by} \quad AV_2 = 0 = \begin{pmatrix} W_1^T W_1 S & 0 \\ W_2^T W_1 S & 0 \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \tag{10}$$

■

3. Principle Component Analysis

We first talk about total PCA.

Definition(Principle Component Analysis). Assume $X, Y \in \mathbb{R}^n$ are random vectors. A linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$Y = TX, \quad y_i = \alpha_i^T X, \quad i = 1, ..., n. \tag{11}$$

where T has an orthonormal matrix $A = (\alpha_i)^T$ with respect to standard basis, $\alpha_i \in \mathbb{R}^n$ and $\alpha_i^T \alpha_j = \delta_{ij}$. We could show that there exists α_1 such that after transformation, y_1 has the maximum variance, which is called a principle component.

Firstly, let us recall that $\boldsymbol{\mu} = (\mathbb{E}x_1, \ldots, \mathbb{E}x_n)^T$ is the mean vector, and corresponding covariance matrix $\Sigma = \left(\operatorname{cov}(x_i, x_j)\right)_{ij} = \mathbb{E}(X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^T = \mathbb{E}XX^T - \boldsymbol{\mu}\boldsymbol{\mu}^T$.

After transformation, we have the following property by linearity of ME.

Theorem(Property of ME after Transformation). (i) $\boldsymbol{\mu}_y = A\boldsymbol{\mu}$, that is, $\mathbb{E}y_i = \alpha_i^T \boldsymbol{\mu}$.

(ii) $\Sigma_y = A^T \Sigma A$, that is, $\sigma_{ij} = \operatorname{cov}(x_i, x_j) = \alpha_i^T \Sigma \alpha_j$.

Proof: We prove for (ii). By definition

1 内积空间上的算子

$$
\begin{aligned}
\sigma_{ij} &= \mathbb{E}(y_i - \alpha_i^T \boldsymbol{\mu})(y_j - \alpha_j^T \boldsymbol{\mu})^T \\
&= \mathbb{E}(\alpha_i^T X - \alpha_i^T \boldsymbol{\mu})(\alpha_j^T X - \alpha_j^T \boldsymbol{\mu}) \\
&= \mathbb{E}\alpha_i^T (X - \boldsymbol{\mu}) \alpha_j^T (X - \boldsymbol{\mu}) \\
&= \mathbb{E}\alpha_i^T (X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^T \alpha_j \quad \text{by symmetry of inner product} \\
&= \alpha_i^T \mathbb{E}[(X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^T] \alpha_j \\
&= \alpha_i^T \Sigma \alpha_j.
\end{aligned} \tag{12}
$$

∎

From the above property, we could explain: we ask the matrix to be orthonormal because we want the covariance matrix of Y to be diagonal, i.e. y_i and y_j are mutually irrelevant unless $i = j$.

Theorem(Theorem for principle component analysis). The maximum of variance of y_1 is reached when α_1 is the eigenvector of the maximum eigenvalue λ_1 of matrix Σ, and satisfies $\mathrm{Var}(y_1) = \lambda_1$.

Proof: To maximize $\mathrm{Var}(y_1)$, is equivalently to maximize $\alpha_1^T \Sigma \alpha_1$ for all possible $\alpha_1 \in \mathbb{R}^n$. Take derivative (gradient for $\alpha_1 \in \mathbb{R}^n$) of the corresponding Lagrangian function with condition $\alpha_1^T \alpha_1 = 1$ and we have $2\Sigma\alpha_1 - 2l(\alpha_1) = \theta$. So to reach the maximum, l is an eigenvalue of Σ, and at this time the goal function equals

$$\alpha_1^T \Sigma \alpha_1 = \alpha_1^T l \alpha_1 = l \alpha_1^T \alpha_1 = l. \tag{13}$$

To reach the maximum, let l to be the maximum eigenvalue of Σ, denoted by λ_1, and choose a eigenvector α_1 correspondingly.

■

If we want to get k principle components, which are mutually irrelevant, i.e. $\operatorname{cov}(y_i, y_j) = 0$ unless $i = j$, we could have the following conclusion.

Theorem(Theorem for k principle components analysis). The k principle components of X is determined by a transformation T defined by

$$y_i = \alpha_i^T X, \quad i = 1, \ldots, k, \tag{14}$$

where $\alpha_i (i = 1, \ldots, k)$ is the eigenvector with respect to the maximum k eigenvalues of Σ.

Proof: We only prove for $k = 2$, the other situation could be deduced by induction.

We aim to find a vector α_2, such that we maximize $\alpha_2 \Sigma \alpha_2$, with a condition $\alpha_2^T \alpha_2 = 1, \langle \alpha_2, \alpha_1 \rangle = 0$. Take a gradient we have

$$2\Sigma\alpha_2 - 2l_1\alpha_2 - l_2\alpha_1 = \theta. \tag{15}$$

apply inner product with α_1, we have

$$2\lambda_1 \alpha_2^T \alpha_1 - l_2 = 0, \quad \Sigma \text{ is self-adjoint} \tag{16}$$

so $l_2 = 0$. Then by the same logic of Theorem for principle component analysis, we also have λ_2 to be the second largest eigenvalue of Σ. Apparently $l_2 \neq \lambda_1$ otherwise $\langle \alpha_2, \alpha_1 \rangle \neq 0$.

■

After transformation to $Y \in \mathbb{R}^n$, we have an amazing result of total variance of Y

$$\sum_{i=1}^n \mathrm{Var}(y_i) = \sum_{i=1}^n \lambda_i = \sum_{i=1}^n \sigma_{ii}. \tag{17}$$

which is given by taking the trace of

$$\mathrm{tr}(\Sigma_y) = \mathrm{tr}(A^T \Sigma A) = \sum_{i=1}^n (\alpha_i^T \Sigma \alpha_i) = (\lambda_i) \tag{18}$$

and making use of $\Sigma = A^T \Sigma_y A = A \Sigma_y A^T$.

After choosing n principle components, we also want to find some relationship between y_i and x_j.

Definition(Factor of Loading). The factor of loading for y_i with respect to x_j is defined by

$$\rho(y_i, x_j) = \frac{\sqrt{\lambda_i}\alpha_{ji}}{\sqrt{\sigma_{jj}}}. \tag{19}$$

where α_{ji} is the j-th component of vector α_i. We have to compute this element-wisely.

Proof: Just by definition.

$$
\begin{aligned}
\rho(y_i, x_j) &= \frac{\operatorname{cov}(y_i, x_j)}{\sqrt{\lambda_i \sigma_{jj}}} \\
&= \frac{\operatorname{cov}(\alpha_i^T X, e_j^T X)}{\sqrt{\lambda_i \sigma_{jj}}} \\
&= \frac{\alpha_i^T \ \operatorname{cov}(X, X) e_j}{\sqrt{\lambda_i \sigma_{jj}}} \\
&= \frac{e_j^T \Sigma \alpha_i}{\sqrt{\lambda_i \sigma_{jj}}} \\
&= \frac{\lambda_i e_j^T \alpha_i}{\sqrt{\lambda_i \sigma_{jj}}} \\
&= \frac{\sqrt{\lambda_i} \alpha_{ji}}{\sqrt{\sigma_{jj}}}.
\end{aligned} \tag{20}
$$

■

Theorem(Properties of factor of loading). (i) Sum over original variable.

$$\sum_{j=1}^{n} \sigma_{jj}\rho^2(y_i, x_j) = \lambda_i. \tag{21}$$

(ii) Sum over all principle components

$$\sum_{i=1}^{n} \rho^2(y_i, x_j) = 1. \tag{22}$$

Proof: We give proof for (ii) using outer product formula.

Since $\Sigma = A\Sigma_y A^T = \sum_{i=1}^{n} \lambda_i \alpha_i \alpha_i^T$, so

$$\sigma_{jj} = \sum_{i=1}^{n} \lambda_i \alpha_{ji}^2. \tag{23}$$

■

3.1. Normalized version

Usually different random variables have distinct values. We have to normalize them if we want to analyse them together.

$$x_i^* = \frac{x_i - \mathbb{E}x_i}{\sqrt{\mathrm{Var}(x_i)}}, \quad i = 1, ..., n. \tag{24}$$

So all the content above would be the same except the following changes.

Theorem(Changes applied to normalized random vectors). (i) $\boldsymbol{\mu}^* = \theta$ and $\Sigma^* = R$, where R is the correlation coefficient matrix with $r_{jj} = \sigma_{ii} = 1$.

(ii) sum over variance after transformation. $\sum_{i=1}^{n} \lambda_i^* = n$.

(iii) load of factors. $\rho(y_i, x_j) = \sqrt{\lambda_i^*}\alpha_{ji}$.

3.2. Truncated principle components

In practice, we usually do not choose n principle components but rather $k \ll n$ to achieve compression of data.

How to choose these k components? we based on the following criterion.

Definition(Contribution to variance). the contribution to total variance of principle component y_i is defined by

$$\eta_i = \frac{\lambda_i}{\sum_{k=1}^{n} \lambda_k} \tag{25}$$

usually we need to let $\sum_{i=1}^{k} \eta_i$ to be larger than 70%.

3.3. Sampled PCA

In actual experiments, we have to observe independently m times. We have to replace mean and covariance matrix with their empirical versions. Assume $X_1, \ldots, X_m$ are m mutually independent random vectors (samples in $\mathbb{R}^n$), then the unbiased estimates of mean and variance are

$$\boldsymbol{\mu} \approx \overline{X} = \sum_{k=1}^{m} X_k, \quad \sigma_{jj} = \frac{1}{m-1} \sum_{k=1}^{m} \left(X_k - \overline{X}\right)^2. \tag{26}$$

So we have its empirical covariance matrix $S = \frac{1}{m-1} \sum_{k=1}^{m} \left(X_k - \overline{X}\right)\left(X_k - \overline{X}\right)^T$. Tackle this matrix with the same method we used in PCA, then we are done.

For actual calculation, we usually let $X_k = \frac{X_k - \overline{X}}{\sqrt{s_{kk}}}$ for each $k = 1, ..., m$, and solve singular values of $X' = (X_1, ..., X_m)_{n\times m}$ as $s_1 > ... > s_n$, then $\lambda_i = s_i^2$ for $i = 1, ..., n$. And $V = A$ and $Y = V^T X$. If we choose k principle components, then choose k columns of V as eigenvectors.

目录

1. Bilinear forms

Definition(Definition of bilinear form). A **bilinear form** on V is a function $\beta : V \times V \to \mathbb{F}$, such that

$$v \mapsto \beta(v, u), \quad v \mapsto \beta(u, v) \tag{27}$$

are both linear functionals on V for each $u \in V$.

Example. (i) $\mathbb{F} = \mathbb{R}$, inner product on V, i.e. $(u, v) \mapsto \langle u, v \rangle$ is a bilinear form.

Note that for $\mathbb{F} = \mathbb{R}$, a bilinear form differs from inner product in that inner product requires symmetry ($\beta(u, v) = \beta(v, u)$) and positive definiteness ($\beta(v, v) > 0$ for all $v \in V - \{\theta\}$), whereas these properties are not required for a bilinear form.

Example. Show that a bilinear form β on V, is also a linear map on $V \times V$, then $\beta = \theta$.

For simplicity, we denote the set of all the bilinear forms on V by $V^{(2)}$.

Definition(Matrix form for a bilinear form).

Theorem(composition of a bilinear form and an operator). Suppose β is a bilinear form on V and T is a linear operator on V. Define two supplementary bilinear forms

$$\alpha(u,v) = \langle u, Tv\rangle, \quad \rho(u,v) = \langle Tu, v\rangle. \tag{28}$$

Let $e_1, ..., e_n$ be a basis of V, then

$$\mathcal{M}(\alpha) = \mathcal{M}(\beta)\mathcal{M}(T), \quad \mathcal{M}(\rho) = \mathcal{M}(T)^t\mathcal{M}(\beta). \tag{29}$$

Theorem(change-of-basis formula).

2. Symmetric bilinear form

Definition(Definition of Symmetric bilinear form). A bilinear form $\rho \in V^{(2)}$ is called symmetric if

$$\rho(u,w) = \rho(w,u) \tag{30}$$

for all $u, w \in V$. The set of symmetric bilinear form on V is denoted by $V^{(2)}_{\text{sym}}$.

Example. (i) Suppose V is a real inner product space, then

$$\rho(u,w) = \langle u, Tw\rangle \tag{31}$$

is symmetric bilinear form iff T is self-adjoint.

Definition(Alternating bilinear form(交错双线性型)). A bilinear form α on V is call alternating, if

$$\alpha(v, v) = 0, \quad \forall v \in V. \tag{32}$$

The set of all alternating bilinear form is denoted by $V_{\text{alt}}^{(2)}$.

Example. (i) Suppose $n \geq 3$ and then $\alpha : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ defined by

$$\alpha((x_1, ..., x_n), (y_1, ..., y_n)) = x_1y_2 - x_2y_1 + x_1y_3 - y_1x_3 \tag{33}$$

is alternating.

Theorem(Characterization of alternating bilinear form). A bilinear form α on V is alternating, iff

$$\alpha(u, w) = -\alpha(w, u), \quad \forall u, w \in V. \tag{34}$$

Proof:

■

Now the following theorem describes the composition of $V^{(2)}$.

Theorem(Theorem). The set $V_{\text{sym}}^{(2)}$ and $V_{\text{alt}}^{(2)}$ are subsets of $V^{(2)}$. Furthermore,

$$V^{(2)} = V_{\text{alt}}^{(2)} \oplus V_{\text{sym}}^{(2)}. \tag{35}$$

Proof:

- Show that $V_{\text{sym}}^{(2)}$ and $V_{\text{alt}}^{(2)}$ are subsets of $V^{(2)}$ by definition.
- Show that $V^{(2)} = V_{\text{sym}}^{(2)} + V_{\text{alt}}^{(2)}$. Suppose $\beta \in V^{(2)}$, then define $\rho, \alpha \in V^{(2)}$ by

$$\rho(u,w) = \frac{1}{2}(\beta(u,w) + \beta(w,u)), \quad \alpha(u,w) = \frac{1}{2}(\beta(u,w) - \beta(w,u)) \tag{36}$$

  so $\rho \in V_{\text{sym}}^{(2)}$ and $\alpha \in V_{\text{alt}}^{(2)}$, and $\beta = \rho + \alpha$.
- Show that $V_{\text{sym}}^{(2)} \cap V_{\text{alt}}^{(2)} = \{0\}$. That is, let $\beta \in V_{\text{sym}}^{(2)} \cap V_{\text{alt}}^{(2)}$, then

$$\beta(u,w) = \beta(w,u) = -\beta(u,w) \Rightarrow \beta(u,w) = 0, \quad \forall u, w \in V. \tag{37}$$

  So $\beta = 0$.

■

3. Quadratic form

Definition(Quadratic form induced by bilinear form). Suppose β is a bilinear form on V, define a function $q_\beta : V \to \mathbb{F}$ by $q_\beta(v) = \beta(v,v)$.

A function $q : V \to \mathbb{F}$ is called a **quadratic form** on V if there exists a bilinear form β such that $q = q_\beta$.

Example. Quadratic form.

(i) For $\beta((x_1, x_2, x_3), (x_1, x_2, x_3)) = x_1 y_1 - 4x_1 y_2 + 8x_1 y_3 - 3x_3 y_3$, q_β is given by

$$q_\beta((x_1, x_2, x_3)) = x_1^2 - 4x_1 x_2 + 8x_1 x_3 - 3x_3^2. \tag{38}$$

Theorem(Quadratic form on $\mathbb{F}^n$). Suppose n is an positive integer and q is a function from $\mathbb{F}^n$ to $\mathbb{F}$. Then q is a quadratic form on V iff there exist numbers $A_{j,k}$ for $j, k = 1, ..., n$ such that

$$q(x_1, ..., x_n) = \sum_{k=1}^{n} \sum_{j=1}^{n} A_{j,k} x_j x_k, \quad \forall (x_1, ..., x_n) \in \mathbb{F}^n. \tag{39}$$

Proof:

- Necessary. By definition.
- Sufficient. Given a quadratic form, define a corresponding bilinear form.

■

Theorem(characterization of quadratic forms). Suppose $q : V \to \mathbb{F}$ is a function. TFAE.

(i) q is a quadratic form.

(ii) There exists a **unique** symmetric bilinear form ρ on V such that $q = q_\rho$.

(iii) $q(\lambda v) = \lambda^2 q(v)$ for all $\lambda \in \mathbb{F}$ and all $v \in V$. Furthermore, the function

$$(u, w) \mapsto q(u + w) - q(u) - q(w) \tag{40}$$

is a symmetric bilinear form on V.

(iv) $q(2v) = 4q(v)$ for all $v \in V$. Furthermore, the function

$$(u, w) \mapsto q(u + w) - q(u) - q(w) \tag{41}$$

is a symmetric bilinear form on V.

Proof: (i) $\Rightarrow$ (ii). By decomposition of $V^{(2)}$.

(ii) $\Rightarrow$ (iii). By utilizing the bilinear form.

(iii) $\Rightarrow$ (iv) is apparent.

(iv) $\Rightarrow$ (i). Just define

$$\rho(u, w) = \frac{q(u + w) - q(u) - q(w)}{2} \tag{42}$$

which is a symmetric bilinear form. Then the corresponding q_ρ satisfies

$$q_\rho(v) = \rho(v, v) = \frac{q(v+v) - q(v) - q(v)}{2} = q(v). \tag{43}$$

which means q is a quadratic form.

■

Theorem(diagonalization of quadratic form). Suppose q is a quadratic form on V.

(i) There exists a basis $e_1, ..., e_n$ of V and $\lambda_1, ..., \lambda_n \in \mathbb{F}$ such that

$$q(x_1 e_1 + ... + x_n e_n) = \lambda_1 x_1^2 + ... + \lambda_n x_n^2, \quad \forall x_1, ..., x_n \in \mathbb{F}^n. \tag{44}$$

(ii) If $\mathbb{F} = \mathbb{R}$ and V is an inner product space, then the basis in (a) can be chosen to be an orthogonal basis of V.

浙江大學
ZHEJIANG UNIVERSITY

Thank You For Listening!