奇异值分解

1.奇异值

我们将在之后的论述中用到T^*T的如下性质.

1.1 T^*T的性质

设$T \in \mathcal{L}(V,W)$,那么我们有

(a) T^*T是V上的正算子.

(b) null $T^*T=$ null T.

(c) range $T^*T=$ range T^*.

(d) $\dim \operatorname{range} T=\dim \operatorname{range} T^*=\dim \operatorname{range} T^*T$.

Proof.

(a) 我们有

$$(T^*T)^*=T^*(T^*)^*=T^*T$$

于是T^*T是自伴算子.又对于任意$v \in V$有

$$\langle T^*Tv,v\rangle=\langle Tv,Tv\rangle=||Tv||^2 \geqslant 0$$

于是T^*T是正算子.

(b) 对于任意$v \in$ null T^*T,有

$$||Tv||^2=\langle Tv,Tv\rangle=\langle T^*Tv,v\rangle=\langle \mathbf{0},v\rangle=0$$

于是$Tv=\mathbf{0}$,即null $T^*T \subseteq$ null T.另一方面,对于任意$v \in$ null T,都有

$$T^*Tv=T^*\mathbf{0}=\mathbf{0}$$

于是null $T \subseteq$ null T^*T.综上可知null $T^*T=$ null T.

(c) 由于T^*T是自伴算子,于是

$$\text{range } T^*T=(\text{null } T^*T)^{\perp}=(\text{null } T)^{\perp}=\text{range } T$$

(d) 我们有

$$\dim \operatorname{range} T=\dim(\text{null } T^*)^{\perp}=\dim W-\dim \operatorname{null} T^*=\dim \operatorname{range} T^*$$

算子的特征值可以反应算子的一些性质.还有一种被称为"奇异值"的数也是很有用的.

1.2 定义:奇异值

设$T \in \mathcal{L}(V,W)$.T的**奇异值**是T^*T的特征值的非负平方根,按降序排列,且每个奇异值的出现次数等于T^*T对应特征空间的维数.

1.3 正奇异值的作用

设$T \in \mathcal{L}(V,W)$,那么我们有

(a) T是单射,当且仅当0不是T的奇异值.

(b) T的正奇异值个数等于$\dim \operatorname{range} T$.

(c) T是满射,当且仅当T的正奇异值个数等于$\dim W$.

Proof.

(a) 我们有

$$T\text{是单射} \Leftrightarrow \operatorname{null} T = \{\mathbf{0}\} \Leftrightarrow \operatorname{null} T^*T = \{\mathbf{0}\} \Leftrightarrow 0\text{不是}T^*T\text{的特征值} \Leftrightarrow 0\text{不是}T\text{的奇异值}$$

(b) 对T^*T应用谱定理可知$\dim \operatorname{range} T^*T$等于$T^*T$的正特征值个数.
又因为$\operatorname{range} T^*T = \operatorname{range} T$,于是$\dim \operatorname{range} T$等于$T$的正奇异值的数目.

(c) 由**(b)**和满射的性质可以得到.

接下来的这条结果用奇异值很好地刻画了等距映射.

1.4 等距映射的奇异值

设$S \in \mathcal{L}(V,W)$,那么S是等距映射当且仅当S的所有奇异值均等于1。

Proof.

我们有

$$\begin{aligned} S\text{是等距映射} &\Leftrightarrow S^*S = I \\ &\Leftrightarrow S^*S\text{的所有特征值均为}1 \\ &\Leftrightarrow S\text{的所有奇异值均等于}1 \end{aligned}$$

2.奇异值分解

接下来的结果表明,对于V到W的每个线性映射,都可以用它的奇异值和V和W中的规范正交组给出简洁的描述.在下一节,我们也将看到奇异值分解(SVD)的重要应用.

2.1 奇异值分解

设$T \in \mathcal{L}(V,W)$,且T的正奇异值为$s_1,\cdots,s_m$.

那么存在V中的规范正交组$e_1,\cdots,e_m$和W中的规范正交组$f_1,\cdots,f_m$使得

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m$$

对任意$v \in V$成立.

Proof.

令$s_1,\cdots,s_n$表示T的奇异值,则$n = \dim V$.

由于T^*T是正算子,于是根据谱定理可知存在V的规范正交基$e_1,\cdots,e_n$使得

$$T^*Te_k = s_k^2 e_k$$

对于任意$k \in \{1,\cdots,n\}$成立.

对于任意$k \in \{1,\cdots,m\}$,令$f_k = \dfrac{Te_k}{s_k}$.对于任意$j,k \in \{1,\cdots,m\}$有

$$\langle f_j, f_k \rangle = \frac{\langle Te_j, Te_k \rangle}{s_j s_k} = \frac{\langle e_j, T^*Te_k \rangle}{s_j s_k} = \frac{s_k}{s_j} \langle e_j, e_k \rangle = \begin{cases} 0, j \neq k \\ 1, j = k \end{cases}$$

于是$f_1,\cdots,f_m$是W中的规范正交组.

对于任意$k \in \{m+1,\cdots,n\}$有$s_k = 0$,从而$T^*Te_k = \mathbf{0}$,也即$Te_k = \mathbf{0}$.

对于任意$v \in V$有

$$\begin{aligned} Tv &= T\left(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n\right) \\ &= \langle v, e_1 \rangle Te_1 + \cdots + \langle v, e_m \rangle Te_m \\ &= s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m \end{aligned}$$

于是欲证等式成立.

根据奇异值分解的结果,对于任意$T \in \mathcal{L}(V,W)$,将SVD中的$e_1,\cdots,e_m$和$f_1,\cdots,f_m$分别扩充为V和W的基.

SVD的结果表明

$$Te_k = \begin{cases} s_k f_k, 1 \leqslant k \leqslant m \\ 0, m < k \leqslant \dim V \end{cases}$$

那么T关于$e_1,\cdots,e_{\dim V}$和$f_1,\cdots,f_{\dim W}$的矩阵具有如下简单形式.

$$\mathcal{M}\left(T, (e_1,\cdots,e_{\dim V}), (f_1,\cdots,f_{\dim W})\right)_{j,k} = \begin{cases} s_k, 1 \leqslant j = k \leqslant m \\ 0, \text{其它情形} \end{cases}$$

当$\dim V = \dim W$时,上述矩阵描述的就是对角矩阵.而其余情形下,我们可以看到一种类似于对角矩阵的形式.我们对对角矩阵的定义做一些扩展.

2.2 定义:对角矩阵

$m\times n$矩阵A被称为**对角矩阵**,如果A中除了$A_{k,k}(k\in\{1,\cdots,\min\{m,n\}\})$可能不为0以外,所有元素均为0.

奇异值分解给出了一种新的理解伴随和伪逆的方式.

2.3 伴随和伪逆的奇异值分解

设$T\in\mathcal{L}(V,W)$且T的正奇异值为$s_1,\cdots,s_m$.设$e_1,\cdots,e_m$和$f_1,\cdots,f_m$是V和W中的规范正交组,满足

$$Tv=s_1\langle v,e_1\rangle f_1+\cdots+s_m\langle v,e_m\rangle f_m$$

对任意$v\in V$成立.那么对于任意$w\in W$有

$$T^*w=s_1\langle w,f_1\rangle e_1+\cdots+s_m\langle w,f_m\rangle e_m$$

以及

$$T^\dagger w=\frac{\langle w,f_1\rangle}{s_1}e_1+\cdots+\frac{\langle w,f_m\rangle}{s_m}e_m$$

Proof.

对于$v\in V$和$w\in W$,我们有

$$\begin{aligned}\langle Tv,w\rangle&=\langle s_1\langle v,e_1\rangle f_1+\cdots+s_m\langle v,e_m\rangle f_m,w\rangle\\&=s_1\langle v,e_1\rangle\langle f_1,w\rangle+\cdots+s_m\langle v,e_m\rangle\langle f_m,w\rangle\\&=\langle v,s_1\langle v,e_1\rangle f_1+\cdots+s_m\langle v,e_m\rangle f_m\rangle\end{aligned}$$

于是

$$T^*w=s_1\langle v,e_1\rangle f_1+\cdots+s_m\langle v,e_m\rangle f_m$$

对于$w\in W$,令

$$v=\frac{\langle w,f_1\rangle}{s_1}e_1+\cdots+\frac{\langle w,f_m\rangle}{s_m}e_m$$

将T作用于上式两侧可得

$$\begin{aligned}Tv&=\frac{\langle w,f_1\rangle}{s_1}Te_1+\cdots+\frac{\langle w,f_m\rangle}{s_m}Te_m\\&=\langle w,f_1\rangle f_1+\cdots+\langle w,f_m\rangle f_m\\&=P_{\text{range }T}w\end{aligned}$$

于是根据伪逆的定义可知

$$T^\dagger w=\frac{\langle w,f_1\rangle}{s_1}e_1+\cdots+\frac{\langle w,f_m\rangle}{s_m}e_m$$

我们还可以给出SVD的矩阵版本.

2.4 奇异值分解的矩阵版本

设A是$p\times n$矩阵,秩$m\geqslant 1$.那么存在列规范正交的$p\times m$矩阵B,正定的$m\times m$对角矩阵D和列规范正交的$n\times m$矩阵C使得$A=BDC^*$.

Proof.

令线性映射$T:\mathbb{F}^n\to\mathbb{F}^p$关于标准基的矩阵为$A$,那么$\dim\operatorname{range}T=m$.令

$$Tv=s_1\langle v,e_1\rangle f_1+\cdots+s_m\langle v,e_m\rangle f_m$$

是T的奇异值分解.令B为各列为$f_1,\cdots,f_m$的$p\times m$矩阵,D为对角线元素为$s_1,\cdots,s_m$的对角矩阵,C为各列为$e_1,\cdots,e_m$的$n\times m$矩阵.令$u_1,\cdots,u_m$表示$\mathbb{F}^m$的标准基,对于任意$k\in\{1,\cdots,m\}$有

$$(AC-BD)u_k=Ae_k-B(s_ku_k)=s_kf_k-s_kf_k=\mathbf{0}$$

于是$AC=BD$,两边右乘C^*有$ACC^*=BDC^*$.

注意到C^*的行为$e_1,\cdots,e_m$的复共轭,因此对于任意$k\in\{1,\cdots,m\}$由矩阵乘法的定义可得$C^*e_k=u_k$,于是$CC^*e_k=e_k$.因此$ACC^*v=Av$对所有$v\in\operatorname{span}(e_1,\cdots,e_m)$成立.

对于任意$v\in(\operatorname{span}(e_1,\cdots,e_m))^{\perp}$,都有$Av=\mathbf{0}$.根据矩阵乘法的定义可知$C^*v=\mathbf{0}$.因此$ACC^*v=Av=\mathbf{0}$.

综合上述推理可知$ACC^*=A$.于是$A=BDC^*$,命题即得证.

注意到以上结果中的A有pn个元素,而B,D,C一共有$m(p+m+n)$个元素.因此,如果秩m相对于p和n很小,就可以用相对小的空间开销在计算机中存储矩阵A.