

## 2.7 第七次习题课

### 2.7.1 正交投影与最小二乘法

#### 正交投影

在三维几何空间中, 设  $U$  是过原点的一个平面 (则其为一个二维子空间),  $l$  为过原点且垂直于  $U$  的直线 (其为一个一维子空间), 则  $l$  中的每个向量与  $U$  中的每个向量都正交, 称  $l$  为  $U$  的正交补, 类似地, 定义一般欧几里得空间的一个子空间的正交补:

##### 定义 2.31 (正交补)

设  $V$  是一个欧式空间,  $S$  是  $V$  的一个非空子集, 则  $V$  中与  $S$  里每个向量都正交的所有向量组成的集合称为  $S$  的正交补, 记作  $S^\perp$ . 即

$$S^\perp \stackrel{\text{def}}{=} \{\alpha \in V | (\alpha, \beta) = 0, \forall \beta \in S\}.$$

容易验证  $S^\perp$  是  $V$  的一个子空间.

##### 定理 2.6

设  $U$  是欧式空间  $V$  的一个有限维子空间, 则有  $V = U \oplus U^\perp$ .

**证明** 首先说明  $U \cap U^\perp = 0$ , 这是因为:  $\forall a \in U \cap U^\perp, (a, a) = 0 \Rightarrow a = 0$ .

其次说明  $\forall \alpha \in V$ , 有  $\alpha = \alpha_1 + \alpha_2$ , 其中  $\alpha_1 \in U, \alpha_2 \in U^\perp$ :

取  $U$  的一组标准正交基  $e_1, e_2, \dots, e_m$ , 令  $\alpha_1 = \sum_{i=1}^m (\alpha, e_i) e_i \in U$ , 则有  $\alpha_2 = \alpha - \alpha_1$ ,

$(\alpha_2, e_i) = (\alpha, e_i) - (\alpha_1, e_i) = (\alpha, e_i) - (\alpha, e_i) = 0$ , 即  $\alpha_2 \in U^\perp$ , 得证.

从上述证明过程我们可以看到,  $V$  中的任意向量  $\alpha$  可以唯一地表示为  $\alpha = \alpha_1 + \alpha_2, \alpha_1 \in U, \alpha_2 \in U^\perp$ , 从而:

##### 定义 2.32 (正交投影)

$V$  上的线性变换  $P_U: \alpha \rightarrow \alpha_1$  称为  $V$  在  $U$  上的正交投影,  $\alpha_1$  称为  $\alpha$  在  $U$  上的正交投影.

在三维几何空间中, 求一个向量  $\alpha$  到一个过原点平面的正交投影即从这个向量端点向平面做垂线得到垂足, 连接原点与垂足的向量即为正交投影. 同时可以看到, 平面内其他向量到  $\alpha$  的距离都大于正交投影到  $\alpha$  的距离. 类似地, 对于一般欧式空间, 有:

##### 定理 2.7

对于  $\alpha \in V, \alpha_1 \in U, \alpha_1$  是  $\alpha$  在  $U$  上的正交投影当且仅当

$$d(\alpha, \alpha_1) \leq d(\alpha, \gamma), \forall \gamma \in U.$$

**证明** 必要性: 设  $\alpha_1$  为  $\alpha$  在  $U$  上的正交投影, 则  $\alpha - \alpha_1 \in U^\perp$ . 对  $\forall \gamma \in U$ , 有  $\alpha_1 - \gamma \in U, (\alpha - \alpha_1, \alpha_1 - \gamma) = 0$ , 则有  $|\alpha - \alpha_1|^2 + |\alpha_1 - \gamma|^2 = |\alpha - \gamma|^2 \Rightarrow d(\alpha, \alpha_1) \leq d(\alpha, \gamma)$ . 等号成立当且仅当  $\gamma = \alpha_1$ .

充分性: 设  $\beta$  为  $\alpha$  在  $U$  上的正交投影, 则由必要性知  $d(\alpha, \beta) \leq d(\alpha, \alpha_1)$ , 又在式 (2.67) 中取  $\gamma = \beta$  得  $d(\alpha, \alpha_1) \leq d(\alpha, \beta)$ , 结合两式得  $d(\alpha, \alpha_1) = d(\alpha, \beta)$ , 这等价于  $\alpha_1 = \beta$ .

#### 最小二乘法

##### ¶ 离散数据的多项式拟合

对于观察/测量得到的一组离散数据  $(x_i, y_i), i = 1, 2, \dots, m$ , 我们通常采用插值/拟合的方式来构造它的逼近函数  $\phi(x)$ , 插值的原则是  $\phi(x_i) = y_i$ , 在此不做详述; 而为给出拟合遵循的原则, 首先给出向量范数的定义:

**定义 2.33 (范数)**

对  $\forall X \in \mathbb{R}^n$ , 它对应一个实数  $\|X\|$ , 满足下面三条性质:

- (1) (非负性)  $\|X\| \geq 0$ , 等号成立当且仅当  $X = 0$ ;
  - (2) (齐次性)  $\forall \alpha \in \mathbb{R}$ , 有  $\|\alpha X\| = |\alpha| \|X\|$ ;
  - (3) (三角不等式)  $\|X + Y\| \leq \|X\| + \|Y\|$ ,
- 则称  $\|X\|$  为  $X$  的范数.



易见, 欧式空间中定义的向量模长即为向量的一种范数, 称为欧几里得范数, 记作  $\|X\|_2$ .

利用不同的范数可以定义出向量  $Z = (\phi(x_1), \phi(x_2), \dots, \phi(x_m))$  与  $Y = (y_1, y_2, \dots, y_m)$  之间的误差, 进而给出拟合方法, 特别地, 对欧几里得范数  $\|X\|_2$ , 有:

**定义 2.34**

拟合的误差平方和记为  $Q = \|Z - Y\|_2^2 = \sum_{i=1}^m (\phi(x_i) - y_i)^2$ . 按误差平方和达到极小来构造拟合曲线的方法称为最小二乘法.

**定义 2.35**

多项式拟合: 当拟合函数  $\phi(x)$  形如  $a_0 + a_1x + \dots + a_nx^n$  时, 拟合问题称为  $n$  次多项式拟合, 即求  $\alpha = (a_0, a_1, \dots, a_n)^T$  使得  $Q(\alpha) = \sum_{i=1}^m (a_0 + a_1x_i + \dots + a_nx_i^n - y_i)^2$  取到最小. 我们将会在下一节讨论具体做法.

**矛盾方程组**

对于多项式拟合问题, 取矩阵  $A = \begin{pmatrix} 1 & x_1 & \cdots & x_1^n \\ 1 & x_2 & \cdots & x_2^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \cdots & x_m^n \end{pmatrix}$ , 有  $Q(\alpha) = \|A\alpha - Y\|_2^2$ . 当方程组  $A\alpha = Y$  有解

时, 拟合函数过所有数据点, 误差平方和为 0.

然而, 实际的大多数情况数据点或多或少会存在误差 (“噪声”), 这导致方程组  $A\alpha = Y$  无解, 这时我们称它为矛盾方程组, 使  $\|A\alpha - Y\|_2$  最小的  $\alpha$  称为矛盾方程组的最小二乘解.

我们下面讨论对一般的线性方程组, 求其最小二乘解:

设  $A \in \mathbb{R}^{m \times n}$ , 线性方程组  $Ax = \beta$  的最小二乘解设为  $\alpha$ ,  $A$  的列向量组记为  $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ , 它们张成的子空间记为  $U$ , 则有:

$\alpha$  是  $Ax = \beta$  的最小二乘解

$$\iff |A\alpha - \beta| \leq |A\alpha' - \beta| \forall \alpha' \in \mathbb{R}^n$$

$$\iff |A\alpha - \beta| \leq |\gamma - \beta|, \forall \gamma \in U$$

$$\iff A\alpha \text{ 是 } \beta \text{ 在 } U \text{ 上的正交投影}$$

$$\iff \alpha_j^T (\beta - A\alpha) = 0, j = 1, 2, \dots, n$$

$$\iff A^T (\beta - A\alpha) = 0$$

$$\iff A^T A\alpha = A^T \beta$$

$$\iff \alpha \text{ 是 } A^T A x = A^T \beta \text{ (称为方程组 } Ax = \beta \text{ 的法方程) 的解.}$$

进一步地, 由于  $\text{rank}(A^T A) \leq \text{rank}(A^T A, A^T \beta) \leq \text{rank}(A^T) = \text{rank}(A^T A)$ , 则  $\text{rank}(A^T A) = \text{rank}(A^T A, A^T \beta)$ , 即法方程一定有解.

综上, 我们将求一个线性方程组的最小二乘解问题等价为了求解法方程的问题.

回到多项式拟合问题, 取相应的矩阵  $A$ , 可得其对应的法方程为:

$$\begin{pmatrix} m & \sum_{i=1}^m x_i & \cdots & \sum_{i=1}^m x_i^n \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 & \cdots & \sum_{i=1}^m x_i^{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m x_i^n & \sum_{i=1}^m x_i^{n+1} & \cdots & \sum_{i=1}^m x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \\ \vdots \\ \sum_{i=1}^m x_i^n y_i \end{pmatrix} \quad (2.67)$$

实际在进行离散数据拟合的时候, 数据点的个数  $m$  一般远大于  $n$ , 需要借助计算机来得到矩阵元并进一步通过算法求解法方程组.

**例题 2.36** 对于给定的一组数据点  $(x_i, y_i), i = 1, 2, \dots, m$  进行线性拟合.

**解** 列出相应的法方程为

$$\begin{pmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{pmatrix} \quad (2.68)$$

$$\text{解之得 } a_0 = \frac{(\sum_{i=1}^m x_i^2)(\sum_{i=1}^m y_i) - (\sum_{i=1}^m x_i)(\sum_{i=1}^m x_i y_i)}{m \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2}, a_1 = \frac{m \sum_{i=1}^m x_i y_i - (\sum_{i=1}^m x_i)(\sum_{i=1}^m y_i)}{m \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2}.$$

## 2.7.2 酉空间

### 酉空间与酉变换

考虑定义在复数域上的线性空间  $V$ , 为引进度量, 不能单纯地定义内积为:  $(\alpha, \beta) = \alpha^T \beta$ , 因为这样无法保证内积的正定性, 从而无法给出长度的概念. 因此, 我们舍弃实内积空间关于内积的对称性, 改为如下定义:

#### 定义 2.36 (酉空间)

<sup>a</sup> 在复数域上线性空间  $V$  中定义一个二元函数  $(\alpha, \beta)$ , 若它满足下列四条性质:  $\forall \alpha, \beta, \gamma \in V, k \in \mathbb{C}$ , 有

- (1) (Hermite 性)  $(\alpha, \beta) = \overline{(\beta, \alpha)}$ ;
- (2) (线性性 1)  $(\alpha + \gamma, \beta) = (\alpha, \beta) + (\gamma, \beta)$ ;
- (3) (线性性 2)  $(k\alpha, \beta) = k(\alpha, \beta)$ ;
- (4) (正定性)  $(\alpha, \alpha) \geq 0$ , 取等当且仅当  $\alpha = 0$ ,

则称这个二元函数为  $V$  上的一个内积, 定义了内积的复线性空间  $V$  称为酉空间.

<sup>a</sup> 此处的定义与课本稍有区别

由内积的 Hermite 性及对第一个变量的线性, 可得

$$(\alpha, k_1\beta + k_2\gamma) = \overline{(k_1\beta + k_2\gamma, \alpha)} = \overline{k_1(\beta, \alpha)} + \overline{k_2(\gamma, \alpha)} = \bar{k}_1(\alpha, \beta) + \bar{k}_2(\alpha, \gamma).$$

类似于欧几里得空间, 可以在酉空间中给出长度、角度 ( $\angle \alpha, \beta \stackrel{\text{def}}{=} \arccos \frac{|(\alpha, \beta)|}{\|\alpha\| \|\beta\|}$ )、正交、距离等度量概念, 酉空间也同样有 Cauchy 不等式、三角不等式、勾股定理等, 也可以仿照 Schmidt 正交化过程构造标准正交基.

#### 定义 2.37 (酉矩阵)

复数域上的  $n$  级矩阵  $P$  若满足  $P^*P = PP^* = I$ , 则称  $P$  为酉矩阵.

其中  $P^* = \bar{P}^T$  称  $P$  的共轭转置 (厄米共轭), 有时为了区别于伴随矩阵, 也将其记为  $P^H, P^\dagger$ .

**例题 2.37** 我们曾在课本习题三.17. 求得所有的二阶酉矩阵.

**定理 2.8**

$n$  级复方阵  $P$  为酉矩阵  $\iff P$  的行 (列) 向量组构成酉空间的一组标准正交基.

**命题 2.18**

酉矩阵的性质:

- (1) 单位阵是酉方阵;
- (2)  $A, B$  为同阶酉方阵, 则  $AB$  也是酉方阵;
- (3) 设  $A$  是酉方阵, 则  $\bar{A}, A^T, A^{-1}$  也是酉方阵;
- (4) 酉方阵的行列式、特征值模长均为 1.

**定义 2.38 (酉变换)**

酉空间  $V$  中保持内积的线性变换  $\mathcal{A}$  称为酉空间的一个酉变换, 即:

$$(\mathcal{A}\alpha, \mathcal{A}\beta) = (\alpha, \beta), \forall \alpha, \beta \in V.$$

**定理 2.9**

下列命题等价:

- (1)  $\mathcal{A}$  是酉变换;
- (2)  $\mathcal{A}$  保内积;
- (3)  $\mathcal{A}$  保模长;
- (4)  $\mathcal{A}$  把标准正交基变为标准正交基;
- (5)  $\mathcal{A}$  在标准正交基下的矩阵是酉矩阵.



**证明** 完全类似实内积空间的正交变换.

**定理 2.10**

设  $\mathcal{A}$  是酉空间  $V$  上的酉变换, 则  $V$  中存在一个标准正交基, 使得  $\mathcal{A}$  在此基下的矩阵为对角阵, 且主对角元都是模长为 1 的复数.



**证明** 对酉空间的维数  $n$  进行归纳:

$n = 1$  时酉空间的基为  $\mathcal{A}$  的特征向量, 命题成立.

假设对  $n-1$  维酉空间命题成立, 对于  $n$  维酉空间: 取  $\mathcal{A}$  的一个特征值  $\lambda_1$ , 其模长为 1, 设  $\eta_1$  是  $\mathcal{A}$  属于特征值  $\lambda_1$  的一个单位特征向量, 则  $\langle \eta_1 \rangle$  是  $\mathcal{A}$  的一个不变子空间.  $V = \langle \eta_1 \rangle \oplus \langle \eta_1 \rangle^\perp$ .  $\langle \eta_1 \rangle^\perp$  也是  $\mathcal{A}$  的一个不变子空间 (证明类似于正交变换), 则  $\mathcal{A}|_{\langle \eta_1 \rangle^\perp}$  是  $\langle \eta_1 \rangle^\perp$  上的酉变换, 根据归纳假设, 存在  $\langle \eta_1 \rangle^\perp$  中的一个标准正交基  $\eta_2, \dots, \eta_n$  使得  $\mathcal{A}|_{\langle \eta_1 \rangle^\perp}$  在这组基下的矩阵为对角矩阵  $\text{diag}\{\lambda_2, \dots, \lambda_n\}$ , 且  $\lambda_2, \dots, \lambda_n$  模为 1. 于是  $\mathcal{A}$  在标准正交基  $\lambda_1, \lambda_2, \dots, \lambda_n$  下的矩阵为对角阵  $\text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ , 且对角元模长都为 1. 由归纳法原理命题得证.

**推论 2.9**

$n$  级酉矩阵  $A$  一定酉相似于一个对角矩阵, 其对角元模长均为 1.



**证明** 只需取  $A$  对应的酉变换在定理 2.3 中对应的对角阵即可.

**厄米 (Hermite) 变换与厄米矩阵**

类似于欧式空间的实对称变换, 我们定义厄米变换:

**定义 2.39**

酉空间  $V$  上的变换  $\mathcal{A}$  满足:

$$(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}\beta), \forall \alpha, \beta \in V,$$

则称  $\mathcal{A}$  为  $V$  上的一个 **Hermite** 变换或自伴随变换.

**定义 2.40**

$n$  阶方阵  $A$  若满足  $A^* = A$ , 则称其为 **Hermite** 矩阵或自伴随矩阵.

**定理 2.11**

$n$  维酉空间  $V$  上的线性变换  $\mathcal{A}$  是 Hermite 变换  $\iff \mathcal{A}$  在  $V$  的一个标准正交基下的矩阵  $A$  为 Hermite 矩阵

**定理 2.12**

Hermite 变换 (矩阵) 的特征值为实数.



**证明** 设  $\lambda_1$  为 Hermite 变换的一个特征值, 其对应的特征向量为  $\eta$ , 则有

$$\lambda_1(\eta, \eta) = (\lambda_1\eta, \eta) = (\mathcal{A}\eta, \eta) = (\eta, \mathcal{A}\eta) = (\eta, \lambda_1\eta) = \bar{\lambda}_1(\eta, \eta). \quad (2.69)$$

故  $\lambda_1 = \bar{\lambda}_1$ , 即  $\lambda_1$  为实数.

**定理 2.13**

Hermite 变换在  $V$  中的某个标准正交基下的矩阵为对角阵, 且对角元均为实数.



**证明** 证明过程与定理 2.10 类似.

**推论 2.10**

Hermite 矩阵可以酉相似到一个实对角阵.



在许多物理问题中, 我们需要在复数空间讨论问题, 但得到的物理量又要求是实数, 这时, Hermite 矩阵特征值为实数这一特点便能得到很好的应用.

**例题 2.38** 量子力学第三条基本假设: 量子力学体系的力学量由态矢量空间  $\mathbb{H}$  中的线性厄米算符表示; 其本征矢量的全体构成  $\mathbb{H}$  的一组完备基; 在任一量子态下观测此力学量, 总是以一定概率得到对应厄米算符的某一个本征值.

用数学语言描述为:  $\hat{\mathcal{A}} = \hat{\mathcal{A}}^\dagger$ ,  $\hat{\mathcal{A}}|a_i\rangle = a_i|a_i\rangle$ ,  $|\Psi\rangle = \sum_i c_i|a_i\rangle$ .

上述式 1 即  $\hat{\mathcal{A}}$  为厄米变换, 式 2 即变换的特征方程 (由厄米变换知  $a_i$  为实数), 式 3 为任一态矢量在本征矢量构成的基下展开, 其中  $|c_i|^2$  代表了  $|\Psi\rangle$  态坍缩到本征态  $|a_i\rangle$  的概率, 亦即观测到力学量  $\hat{\mathcal{A}}$  的值为  $a_i$  的概率.

**2.7.3 矩阵的分解**

将一个矩阵分解为数个矩阵相乘是我们处理实际问题时常用的简化手段, 良好的分解方法可以使解决问题事半功倍, 本节列举几种分解方法的实际应用.

**QR 分解****定理 2.14**

QR 分解:  $n$  阶可逆实矩阵  $A$  可分解为一个正交阵  $Q$  和一个上三角阵  $R$  的乘积,  $A = QR$ .



**证明** 证明过程即 Schmidt 正交化过程.

QR 分解在计算数学中常用于求矩阵的特征值和特征向量, 因其具有数值稳定性高、可通过优化算法加速收敛等特点被广泛使用, 下面不加证明地给出求特征值的原理:

**例题 2.39**  $A$  为给定的  $n$  阶实矩阵, 记  $A_1 = A$ , 对  $A_1$  进行 QR 分解得  $A_1 = Q_1 R_1$ , 令  $A_2 = R_1 Q_1$ , 对  $A_2$  进行 QR 分解得到  $A_2 = Q_2 R_2$ , 令  $A_3 = R_2 Q_2$ , 依此类推, 若  $A_k = Q_k R_k$ , 则令  $A_{k+1} = R_k Q_k$ , 这样迭代得到  $n$  阶矩阵序列  $\{A_k\}$ , 满足:

(1)  $\{A_k\}$  为正交相似序列. 事实上, 有  $Q_k A_{k+1} Q_k^T = Q_k R_k Q_k Q_k^T = Q_k R_k = A_k$ .

(2) 记  $A_k$  的元素为  $(a_{ij}^{(k)})_{n \times n}$ , 若矩阵序列  $\{A_k\}$  基本收敛, 则有

$$\begin{aligned} a_{ij}^{(k)} &\rightarrow 0, \quad k \rightarrow \infty, 1 \leq i < j \leq n \\ a_{ii}^{(k)} &\rightarrow \lambda_i, \quad k \rightarrow \infty, i = 1, 2, \dots, n \end{aligned} \quad (2.70)$$

这样便得到了  $A$  的全部特征值  $\lambda_i$ .

**例题 2.40** 除 Schmidt 正交化外, 还可利用 Givens 旋转变换、Householder 反射变换等方法实现 QR 分解.

## 奇异值分解 (SVD)

本小节参考 (<https://zhuanlan.zhihu.com/p/480389473>)

### 定理 2.15

奇异值分解 (SVD): 对  $\forall A \in \mathbb{C}^{m \times n}$ , 存在酉矩阵  $U \in \mathbb{C}^{m \times m}, V \in \mathbb{C}^{n \times n}$ , 使得  $A = U \Sigma V^H$ , 其中  $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_r, 0, \dots, 0\}$ , 其中  $\sigma_i = \sqrt{\lambda_i}$ ,  $\lambda_i$  为  $A^H A$  的非零特征值且  $\lambda_1 \geq \dots \geq \lambda_r > 0$ .

奇异值分解有很多应用, 本节以图像压缩为例给出 SVD 在图像处理中的一个应用:

**例题 2.41** 利用 SVD 对下图进行压缩:



图 2.6: Kobe(612 × 408)

图像一般被存储为一个  $\mathbb{R}^{m \times n \times 3}$  的矩阵, 前两个指标代表像素点的位置, 最后一个指标包含了这个像素点的 RGB 信息, 决定了像素的颜色, 我们可以分别处理三种颜色, 每种颜色对应一个  $\mathbb{R}^{m \times n}$  的矩阵, 可分别进行奇异值分解.

对于一个矩阵  $A$ , 较大的奇异值 (位于  $\Sigma$  矩阵的左上) 及其对应的行、列向量对  $A$  的影响比重大, 所以我们可以扔掉 0 及靠后的奇异值, 仅保留较大的一些奇异值及其对应的行、列向量来实现对图片的压缩. 对于本图,  $A$  有 408 个奇异值, 下面给出仅保留前 1, 4, 20, 40, 200 个奇异值得到的图像压缩结果:

可以看到, 仅取 1 个奇异值时, 得到的图片有明显的色块分布, 几乎看不出图片内容; 取 4 个奇异值时初具人形; 取 20 个奇异值时已经能看清面部表情, 但图片还是很糊; 取 40 个奇异值时效果已经很好了, 但还存在一些噪点; 取 200 个 (一半) 奇异值时已经和原图十分接近.



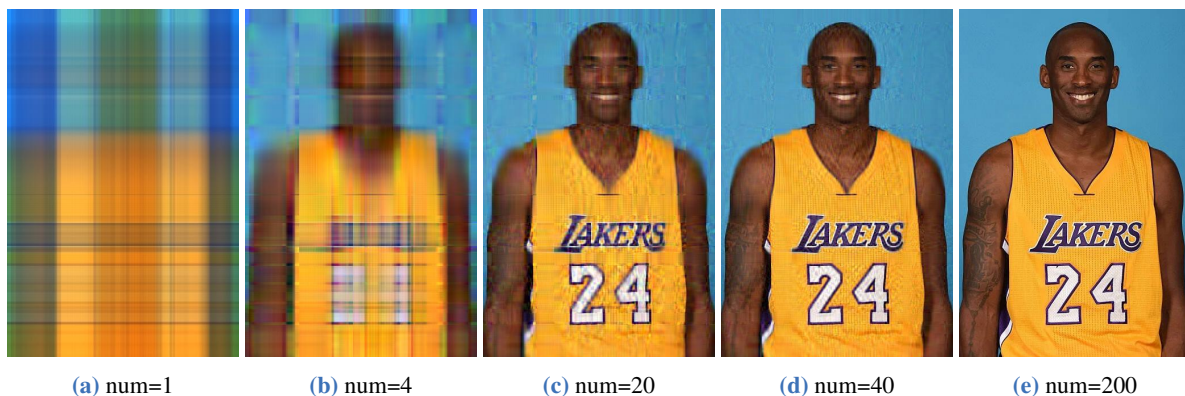


图 2.7: 图像压缩结果

我们选出取 40 个特征值的情况进行详细计算:

原图所需存储的矩阵大小为  $612 \times 408 \times 3 = 749088$ , 对三个  $612 \times 408$  的矩阵分别进行奇异值分解, 每个矩阵得到  $U \in \mathbb{R}^{612 \times 612}$ ,  $V^T \in \mathbb{R}^{408 \times 408}$ , 奇异值有 408 个, 选  $U$  的前 40 列得到  $U' \in \mathbb{R}^{612 \times 40}$ , 选  $V^T$  的前 40 行得到  $V'^T \in \mathbb{R}^{40 \times 408}$ , 这样我们构造出新的矩阵  $A' = U' \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_{40}\} V'^T$ , 这样压缩后我们仅需要存储  $U'$ ,  $V'^T$  及 40 个奇异值, 相应的存储空间大小为  $40 \times (612 + 408 + 1) \times 3 = 122520$ , 为原本的 1/6. 可以看到, 我们通过这种方法将图片压缩了六倍, 且得到的图片仍较为清晰 (虽然丢掉较小的奇异值不可避免地丢失了信息).

## LU 分解

### 定理 2.16

$LU$  分解:  $n$  级矩阵  $A$  可以分解为一个下三角矩阵  $L$  和一个上三角矩阵  $U$  的乘积:  $A = LU$ , 这实际上与高斯消元法对应.

不同形式的  $LU$  分解常用于求解线性方程组, 相比于高斯消元法可以节省部分运算量. 其大致的计算思路为: 求解线性方程组  $Ax = \beta \iff$  求解  $Ly = \beta, Ux = y$  (这两个方程都是很容易求解的).

**例题 2.42** 几种  $LU$  分解的求解步骤.

$$(1) \text{ Doolittle 分解: } \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix}$$

用以下顺序可以逐层求解出  $L, U$  矩阵的矩阵元: 计算  $U$  的第一行元素, 计算  $L$  的第一列元素, 计算  $U$  的第二行元素, 计算  $L$  的第二列元素,  $\dots$ , 计算  $U$  的第  $n$  行元素, 计算  $L$  的第  $n$  列元素.

$$(2) \text{ Crout 分解: } \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} 1 & u_{12} & \cdots & u_{1n} \\ & 1 & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$$

求解方法与 Doolittle 分解完全类似, 只是先计算  $L$  的列, 再计算  $U$  的行.

(3)  $LDL^T$  分解: 对正定阵  $A$  进行 Doolittle 分解, 再提取出  $U$  的对角元  $D = \text{diag}\{u_{11}, u_{22}, \dots, u_{nn}\}$ , 可证  $U = DL^T$ , 从而  $A$  被分解为  $LDL^T$ , 且  $D$  为正定阵的标准型, 分解步骤与 Doolittle 分解完全相同.

**例题 2.43** 利用 Doolittle 分解求解方程组

$$\begin{cases} 2x_1 + x_2 + x_3 = 4 \\ x_1 + 3x_2 + 2x_3 = 6 \\ x_1 + 2x_2 + 2x_3 = 5 \end{cases} \quad (2.71)$$

**解** 系数矩阵经 Doolittle 分解得到:  $\begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.5 & 0.6 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 0 & 2.5 & 1.5 \\ 0 & 0 & 0.6 \end{pmatrix}$

解  $Ly = \beta$  得  $y = (4, 4, 0.6)^T$ , 解  $Ux = y$  得  $x = (1, 1, 1)^T$ .

## 2.7.4 利用初等变换将二次型化为标准型

### 定理 2.17

每一个实对称阵  $A$  都相合于对角阵, 即存在初等阵  $P_1, P_2, \dots, P_r$  使得

$$P_r^T P_{r-1}^T \cdots P_1^T A P_1 P_2 \cdots P_r = \text{diag}(\mu_1, \mu_2, \dots, \mu_n). \quad (2.72)$$

这里  $\mu_1, \mu_2, \dots, \mu_n$  为实数.

**证明** 定理的证明即初等变换法的具体操作过程:

(1) 若  $a_{11} \neq 0$  则利用  $a_{11}$  将第一行、第一列的其他位置消为零, 得  $\begin{pmatrix} a_{11} & 0 \\ 0 & A_{n-1} \end{pmatrix}$ , 继续对  $A_{n-1}$  做类似操作即可.

(2) 当  $a_{11} = 0$ , 第一行其他位置一定有一个非零元素, 设为  $a_{1i}$ , 则将  $A$  的第  $i$  列加到第一行, 再将第  $i$  行加到第一行, 即令  $A' = T_{1i}(1)AT_{1i}(1)^T$ , 则转化为第一种情况.

**例题 2.44** 利用初等变换将二次型  $x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 - 8x_1x_3 - 4x_2x_3$  化为标准型, 并求变换矩阵.

**解** 写出二次型对应的矩阵  $A = \begin{pmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{pmatrix}$ , 将其与单位阵  $I$  上下写在一起, 对  $A$  做一对对行、列变换, 同时只对  $I$  做列变换, 有:

$$\begin{pmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & -4 \\ 0 & 0 & -10 \\ -4 & -10 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -10 \\ 0 & -10 & -15 \\ 1 & 2 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -35 & -25 \\ 0 & -25 & -15 \\ 1 & 6 & 4 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -35 & 0 \\ 0 & 0 & 20/7 \\ 1 & 6 & -2/7 \\ 0 & 1 & -5/7 \\ 0 & 1 & 2/7 \end{pmatrix}$$

即变换矩阵  $P$  写为  $\begin{pmatrix} 1 & 6 & -2/7 \\ 0 & 1 & -5/7 \\ 0 & 1 & 2/7 \end{pmatrix}$ , 令  $(y_1, y_2, y_3)^T = P^{-1}(x_1, x_2, x_3)^T$ , 二次型化为标准型  $y_1^2 - 35y_2^2 + \frac{20}{7}y_3^2$ .

**注** 上面变换的原理是: 对  $A$  做成对的行、列变换,  $A$  变为  $P^TAP$ ; 仅对  $I$  做列变换,  $I$  变为  $P$ , 这样, 我们同时求得了二次型的标准型及其变换矩阵. 同样的, 也可以将  $A$  和  $I$  左右写在一起, 在对  $A$  做成对变换的同时只对  $I$  做行变换, 区别只在于这样得到的  $I$  变换后的矩阵为  $P^T$ .

**注** 从上例可以看出, 由于同一个二次型对应的标准型不唯一, 且经过不同初等变换步骤得到的标准型也不相同, 这是我们不希望看到的. 因此, 在大部分考题中, 我们要求使用正交方阵将其对角化, 这样得到的结果在对角元相差顺序的情况下唯一, 且变换过程不会对二次曲面有拉伸, 良好保持了原有形状.



另解：利用正交变换将上述二次型化为标准型<sup>4</sup>：

**解** 矩阵  $A$  对应特征值为  $\lambda_1 = \lambda_2 = 5, \lambda_3 = -4$ ，对应的特征向量为  $\alpha_1 = (-1, 0, 1)^T, \alpha_2 = (-1, 2, 0)^T, \alpha_3 = (2, 1, 2)^T$ ，经 Schmidt 正交化得到一组标准正交向量  $e_1 = (-1/\sqrt{2}, 0, 1/\sqrt{2})^T, e_2 = (-\sqrt{2}/6, 2\sqrt{2}/3, -\sqrt{2}/6)^T, e_3 = (2/3, 1/3, 2/3)^T$ ，记  $P = (e_1, e_2, e_3)$ ，则有  $P$  为正交阵，且  $P^T A P = \text{diag}\{5, 5, -4\}$ ，对应的正交变换为  $y = P^{-1}x$ ，将二次型化为了标准型  $5y_1^2 + 5y_2^2 - 4y_3^2$ 。特别地， $5y_1^2 + 5y_2^2 - 4y_3^2 = 1$  对应单叶双曲面。

---

<sup>4</sup>2021-2022 第二学期期末 3.