矩阵分解简介

统计 91 董晟渤

2021 年 8 月 29 日

目录

1 前言

所谓的**矩阵分解**, 指的是将给定的矩阵写为一系列矩阵的乘积.

在进行矩阵的分解的时候, 首先需要清楚**分解的对象**是什么. 例如, QR 分解的对象是非奇异的实矩阵, 而 UR 分解的对象是非奇异的复矩阵. 除此之外, 我们也可以尝试分解一些较为一般的矩阵.

除此之外, 我们还需要明白**分解后的形式**是什么. 通常, 我们需要将矩阵分解得足够简洁, 或者性质足够好, 才能够简化问题. 那么哪些矩阵是符合要求的呢? 在此, 给出几种相对简洁, 或者性质较好的矩阵的定义.

最为简单的矩阵就是对角矩阵, 其的定义如下.

定义 1.1 (对角矩阵). *形如*

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}_{n\times n}$$

的矩阵称为对角矩阵, 记作 $\mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$.

当然, 在前面的课程中, 我们知道, 并不是所有的矩阵都可以被化成对角阵. 但是, Jordan 型矩阵的理论指出, 所有的矩阵都可以被化作 Jordan 型矩阵.

定义 1.2 (Jordan 块和 Jordan 形矩阵). *形如*

$$\boldsymbol{J_i} = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{pmatrix}_{n_i\times n_i}$$

的矩阵称为 Jordan 块, 而由若干个 Jordan 块形成的准对角矩阵

$$\boldsymbol{J} = \mathrm{diag}\{\boldsymbol{J}_1, \boldsymbol{J}_2, \cdots, \boldsymbol{J}_n\}$$

称为 Jordan 形矩阵.

再退一步, 上面的这两种矩阵, 它们的元素其实都集中在右上角. 从而, 我们可以定义更一般的, 像三角形一样的矩阵.

定义 1.3 (上三角矩阵与下三角矩阵). 形如

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ 0 & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{nn} \end{pmatrix}_{n\times n}$$

的矩阵称为上三角矩阵; 对应地, 形如

$$\begin{pmatrix} x_{11} & 0 & \cdots & 0 \\ x_{21} & x_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}_{n\times n}$$

的矩阵称为下三角矩阵. 若上 (下) 三角矩阵对角线上的元素 $a_{ii} > 0(1 \le i \le n)$, 则称其为正线上 (下) 三角矩阵. 若上 (下) 三角矩阵对角线上的元素 $a_{ii} = 1(1 \le i \le n)$, 则称其为单位上 (下) 三角矩阵.

上三角矩阵和下三角矩阵是比较容易研究的, 并且可以通过转置来相互转化. 接下来, 对于 $\mathbb{R}^n$ 中的实矩阵, 我们有正交矩阵 的定义. 同时, 对于 $\mathbb{C}^n$ 中的矩阵, 我们推广了正交矩阵的概念, 定义了所谓的酉矩阵.

定义 1.4 (酉矩阵与正交矩阵). 设 $\boldsymbol{U} \in \mathbb{C}^{n\times n}$, 且满足

$$\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^H = \boldsymbol{I},$$

则称其为酉矩阵. 对应地, 设 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, 且满足

$$\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I},$$

则称其为正交矩阵.

更一般地, 我们可以给出如下的定义, 相当于是酉矩阵的一般情况.

定义 1.5 (正规矩阵)**.** 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 且满足

$$\boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{A}^H\boldsymbol{A},$$

则称其为正规矩阵.

如果我们可以把矩阵分解为对角矩阵, 三角矩阵或者酉矩阵, 那么就能够方便地研究矩阵. 除此之外, 矩阵的分解也有许多实际中的应用.

以下, 我们用 $\mathbb{C}_r^{m\times n}$ 来表示复数域 $\mathbb{C}$ 上的 m 行 n 列且秩为 r 的矩阵. 对应地, $\mathbb{R}_r^{m\times n}$ 表示实数域 $\mathbb{R}$ 上的 m 行 n 列且秩为 r 的矩阵.

2　矩阵的初步分解

之所以将这些分解方式叫作初步分解, 是因为这些方法几乎只需要用到高代中的基本方法. 所谓的基本方法, 包括初等行变换和正交化.

2.1　满秩分解

我们首先利用初等行变换, 寻求**一般矩阵**的分解方式. 并且在这里, 我们希望将矩阵分解为满秩的矩阵, 从而方便研究问题.

定理 2.1 (满秩分解)**.** *设* $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$, *则存在* $\boldsymbol{B} \in \mathbb{C}_r^{m\times r}$ *及* $\boldsymbol{C} \in \mathbb{C}_r^{r\times n}$, *使得*

$$\boldsymbol{A} = \boldsymbol{BC}.$$

证明. 用初等行变换, 可以将矩阵 $\boldsymbol{A}$ 化成阶梯矩阵, 也即存在 $\boldsymbol{P} \in \mathbb{C}_m^{m\times m}$, 使得

$$\boldsymbol{PA} = \begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{O} \end{pmatrix},$$

其中 $\boldsymbol{C} \in \mathbb{C}_r^{r\times n}$. 再记

$$\boldsymbol{P}^{-1} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{D} \end{pmatrix},$$

其中 $\boldsymbol{B} = \mathbb{C}_r^{m\times r}$, 则有

$$\boldsymbol{A} = \boldsymbol{P}^{-1}\begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{O} \end{pmatrix} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{D} \end{pmatrix}\begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{O} \end{pmatrix} = \boldsymbol{BC}.$$

□

需要注意的是, 上面的证明过程, 同样给出了求矩阵的满秩分解的方式. 其中, $\boldsymbol{B}$ 为列满秩矩阵, $\boldsymbol{C}$ 为行满秩矩阵. 上面的分解只用到了初等行变换, 并不复杂. 为了求出 $\boldsymbol{P}^{-1}$, 可以考虑对分块矩阵 $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{I} \end{pmatrix}$ 作初等行变换, 使得 $\boldsymbol{A}$ 被化为阶梯矩阵.

2.2 UR 分解

接下来, 我们可以对一个 n 阶的**非奇异复矩阵**进行分解, 将其写成一个酉矩阵和上三角矩阵的乘积. 这个分解只需要应用 Gram-Schmidt 正交化方法.

定理 2.2 (UR 分解)**.** *设 $\boldsymbol{A} \in \mathbb{C}_n^{n\times n}$, 则存在酉矩阵 $\boldsymbol{U}$ 及正线上三角矩阵 $\boldsymbol{R}$, 使得*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{R}.$$

证明. 记

$$\boldsymbol{A} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n).$$

应用 Gram-Schmidt 正交化方法, 我们可以将 $\boldsymbol{A}$ 的列向量化为标准正交向量组 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$, 且有

$$\begin{cases} \boldsymbol{\alpha}_1 = r_{11}\boldsymbol{u}_1, \\ \boldsymbol{\alpha}_2 = r_{12}\boldsymbol{u}_1 + r_{22}\boldsymbol{u}_2, \\ \cdots, \\ \boldsymbol{\alpha}_n = r_{1n}\boldsymbol{u}_1 + r_{2n}\boldsymbol{u}_2 + \cdots + r_{nn}\boldsymbol{u}_n. \end{cases}$$

其中 $r_{ii} > 0$. 记

$$\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n), \boldsymbol{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{pmatrix},$$

则有 $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{R}$, 其中 $\boldsymbol{U}$ 为酉矩阵, $\boldsymbol{R}$ 为正线上三角矩阵. □

将范围限制在实数上, 立即得到如下推论.

推论 2.3 (QR 分解)**.** *设 $\boldsymbol{A} \in \mathbb{R}_n^{n\times n}$, 则存在正交矩阵 $\boldsymbol{Q}$ 及正线上三角矩阵 $\boldsymbol{R}$, 使得*

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}.$$

上面的证明过程, 其实也给出了求矩阵的 UR 分解的方式.

2.3 LU 分解

再考虑更加特殊的 n 阶的实矩阵. 我们不但要求它是非奇异的, 还要求它的所有顺序主子式都非零.

定理 2.4. 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, 且所有顺序主子式都非零, 则存在单位下三角矩阵 $\boldsymbol{L}$ 和非奇异上三角矩阵 $\boldsymbol{U}$, 使得

$$\boldsymbol{A} = \boldsymbol{LU}.$$

证明. 当 $n = 1$ 时, 结论一定成立.

假设所有的 $n-1$ 阶的所有顺序主子式都非零的矩阵都存在 LU 分解, 考虑满足条件的 n 阶的矩阵 $\boldsymbol{A}$, 将其写成分块形式

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix},$$

其中 $\boldsymbol{A}_{11} \in \mathbb{R}^{n\times n}$, 并且其顺序主子式都非零，从而存在单位下三角矩阵 $\boldsymbol{L}_{11}$ 和非奇异上三角矩阵 $\boldsymbol{U}_{11}$, 使得

$$\boldsymbol{A}_{11} = \boldsymbol{L}_{11}\boldsymbol{U}_{11}.$$

令 $\boldsymbol{L}_{21} = \boldsymbol{A}_{21}\boldsymbol{U}_{11}^{-1}$, $\boldsymbol{U}_{12} = \boldsymbol{L}_{11}^{-1}\boldsymbol{A}_{12}$, $\boldsymbol{U}_{22} = \boldsymbol{A}_{22} - \boldsymbol{L}_{21}\boldsymbol{U}_{12}$, 则

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{L}_{11} & \boldsymbol{O} \\ \boldsymbol{L}_{21} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{O} & \boldsymbol{U}_{22} \end{pmatrix}.$$

记 $\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}_{11} & \boldsymbol{O} \\ \boldsymbol{L}_{21} & 1 \end{pmatrix}$, $\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{O} & \boldsymbol{U}_{22} \end{pmatrix}$, 则 $\boldsymbol{A} = \boldsymbol{LU}$, 其中 $\boldsymbol{L}$ 为单位下三角矩阵, $\boldsymbol{U}$ 为非奇异上三角矩阵. □

要实现矩阵的 LU 分解, 有许多的算法, 最基本的是 **Gauss 消去法**. 限于篇幅有限, 在此没有展开介绍.

3 矩阵的奇异值分解

3.1 奇异值的定义

为了给出奇异值的定义, 我们需要先证明一些结论.

引理 3.1. 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$, 则

$$\operatorname{rank}(\boldsymbol{A}^H\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^H) = r.$$

证明. 一方面, 设 $\boldsymbol{x} \in \mathbb{C}^n$ 是齐次线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ 的解, 则有

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^H\boldsymbol{0} = \boldsymbol{0},$$

从而 $\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$; 另外一方面, 设 $\boldsymbol{x} \in \mathbb{C}^n$ 是方程组 $\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ 的解, 则有

$$(\boldsymbol{A}\boldsymbol{x})^H\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^H\boldsymbol{0} = \boldsymbol{0},$$

从而 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. 由上知 $\operatorname{rank}(\boldsymbol{A}^H\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}) = r$. 同理 $\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^H) = r$. □

引理 3.2. 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$, 则 $\boldsymbol{A}^H\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^H$ 的特征值都是非负实数.

证明. 设 λ 是 $\boldsymbol{A}^H\boldsymbol{A}$ 的任一特征值, 其对应的特征向量为 $\boldsymbol{\alpha}$, 则

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}.$$

从而

$$\boldsymbol{\alpha}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{\alpha} = (\boldsymbol{A}\boldsymbol{\alpha})^H\boldsymbol{A}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}^H\boldsymbol{\alpha},$$

解得 $\lambda = \dfrac{(\boldsymbol{A}\boldsymbol{\alpha})^H\boldsymbol{A}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^H\boldsymbol{\alpha}} \geq 0$. 同理 $\boldsymbol{A}\boldsymbol{A}^H$ 的特征值都是非负实数. □

引理 3.3. 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$, 则 $\boldsymbol{A}^H\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^H$ 非零特征值相同, 且个数都为 r.

证明. 设 λ 是 $\boldsymbol{A}^H\boldsymbol{A}$ 的任一特征值, 其对应的特征向量为 $\boldsymbol{\alpha}$, 则

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}.$$

从而

$$\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{\alpha}=\lambda\boldsymbol{A}\boldsymbol{\alpha},$$

也即 λ 是 $\boldsymbol{A}\boldsymbol{A}^H$ 的特征值, 其对应的特征向量为 $\boldsymbol{A}\boldsymbol{\alpha}$. 另外, 根据 $\operatorname{rank}(\boldsymbol{A}^H\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^H)=r$ 知特征值的个数为 r. □

在上面的基础下, 我们可以开始考虑矩阵 $\boldsymbol{A}^H\boldsymbol{A}$ 和 $\boldsymbol{A}\boldsymbol{A}^H$ 的特征值了. 我们知道, 它们的特征值都是相同的, 且必定是非负的. 因此以下定义是合理的.

定义 3.4 (奇异值)**.** 设 $\boldsymbol{A}\in\mathbb{C}_r^{m\times n}$, 记 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_r$, 则

$$\delta_i=\sqrt{\lambda_i},1\le i\le r$$

称为矩阵 $\boldsymbol{A}$ 的奇异值.

事实上, 上面的定义还有另外一种等价的写法.

定义 3.5 (奇异值)**.** 设 $\boldsymbol{A}\in\mathbb{C}_r^{m\times n}$, 若存在非负实数 δ 及向量 $\boldsymbol{u}\in\mathbb{C}^n$, $\boldsymbol{v}\in\mathbb{C}^m$, 使得

$$\boldsymbol{A}\boldsymbol{u}=\delta\boldsymbol{v},\boldsymbol{A}^H\boldsymbol{v}=\delta\boldsymbol{u},$$

则称 δ 为 $\boldsymbol{A}$ 的奇异值, $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 为 $\boldsymbol{A}$ 对应于奇异值 δ 的右奇异向量 和左奇异向量.

一方面, 设 λ 是 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值, 对应的特征向量为 $\boldsymbol{\alpha}$, 则

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{\alpha}=\delta^2\boldsymbol{\alpha}.$$

由上知 $\boldsymbol{\beta}=\dfrac{1}{\delta}\boldsymbol{A}\boldsymbol{\alpha}$ 是 $\boldsymbol{A}\boldsymbol{A}^H$ 的特征向量, 计算得

$$\boldsymbol{A}^H\boldsymbol{\beta}=\frac{1}{\delta}\cdot\delta^2\alpha=\delta\alpha.$$

从而 α 和 β 分别是 $\boldsymbol{A}$ 的右奇异向量和左奇异向量.

另一方面, 设存在非负实数 δ 及非零向量 $\boldsymbol{u}\in\mathbb{C}^n$, $\boldsymbol{v}\in\mathbb{C}^m$, 使得

$$\boldsymbol{A}\boldsymbol{u}=\delta\boldsymbol{v},\boldsymbol{A}^H\boldsymbol{v}=\delta\boldsymbol{u},$$

计算得

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{u} = \delta\boldsymbol{A}^H\boldsymbol{v} = \delta^2\boldsymbol{u}.$$

从而 δ^2 是 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值. 这便说明了上述两种定义的等价性.

3.2 奇异值分解定理

接下来, 对于**一般的矩阵**, 我们给出如下的分解方式.

定理 3.6 (奇异值分解)**.** 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$, 则存在酉矩阵 $\boldsymbol{V} \in \mathbb{C}^{m\times m}$ 和 $\boldsymbol{U} \in \mathbb{C}^{n\times n}$, 使得

$$\boldsymbol{V}^H\boldsymbol{A}\boldsymbol{U} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix},$$

其中 $\boldsymbol{S} = \operatorname{diag}\{\delta_1, \delta_2, \cdots, \delta_r\}$, $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_r \geq 0$ 为 $\boldsymbol{A}$ 的奇异值.

证明. 设 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值为 $\delta_1^2, \delta_2^2, \cdots, \delta_r^2$, 注意到此时 $\boldsymbol{A}^H\boldsymbol{A}$ 为 Hermite 矩阵, 从而存在酉矩阵 $\boldsymbol{U} \in \mathbb{C}^{n\times n}$, 使得

$$\boldsymbol{U}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{U} = \operatorname{diag}\{\delta_1^2, \delta_2^2, \cdots, \delta_r^2, 0, \cdots, 0\} = \begin{pmatrix} \boldsymbol{S}^2 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}.$$

记 $\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{pmatrix}$, 其中 $\boldsymbol{U}_1 \in \mathbb{C}^{n\times r}$, 则

$$\begin{cases} \boldsymbol{A}^H\boldsymbol{A}\boldsymbol{U}_1 = \boldsymbol{U}_1\boldsymbol{S}^2, \\ \boldsymbol{A}^H\boldsymbol{A}\boldsymbol{U}_2 = \boldsymbol{O}. \end{cases}$$

进而可得 $\boldsymbol{A}\boldsymbol{U}_2 = 0$. 令 $\boldsymbol{V}_1 = \boldsymbol{A}\boldsymbol{U}_1 \in \mathbb{C}^{m\times r}\boldsymbol{S}^{-1}$, 可以验证 $\boldsymbol{V}_1^H\boldsymbol{V}_1 = \boldsymbol{I}$. 将 $\boldsymbol{V}_1$ 扩充为 $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{pmatrix} \in \mathbb{C}^{m\times m}$, 使得 $\boldsymbol{V}$ 为酉矩阵, 此时 $\boldsymbol{V}_2^H\boldsymbol{A}\boldsymbol{U}_1 = \boldsymbol{V}_2^H\boldsymbol{V}_1\boldsymbol{S} = \boldsymbol{O}$, 则有

$$\begin{cases} \boldsymbol{V}_1^H\boldsymbol{A}\boldsymbol{U}_1 = \boldsymbol{S}, \\ \boldsymbol{V}_2^H\boldsymbol{A}\boldsymbol{U}_1 = \boldsymbol{O}. \end{cases}$$

代入计算得

$$\boldsymbol{V}^H\boldsymbol{A}\boldsymbol{U}=\begin{pmatrix}\boldsymbol{V}_1^H\\ \boldsymbol{V}_2^H\end{pmatrix}\boldsymbol{A}\begin{pmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{V}_1^H\boldsymbol{A}\boldsymbol{U}_1 & \boldsymbol{V}_1^H\boldsymbol{A}\boldsymbol{U}_2\\ \boldsymbol{V}_2^H\boldsymbol{A}\boldsymbol{U}_1 & \boldsymbol{V}_2^H\boldsymbol{A}\boldsymbol{U}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{S} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}.$$

□

事实上, 上面的证明过程也给出了求矩阵的奇异值分解的方式. 对于矩阵 $\boldsymbol{A}$,

- 首先, 我们需要求出矩阵 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值与特征向量, 开根号后即可得到奇异值;
- 其次, 根据特征向量, 进行单位正交化后可以得到矩阵 $\boldsymbol{U}=\begin{pmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_2\end{pmatrix}$;
- 接下来, 利用等式 $\boldsymbol{V}_1=\boldsymbol{A}\boldsymbol{U}_1\boldsymbol{S}^{-1}$, 再单位正交扩充得到 $\boldsymbol{V}=\begin{pmatrix}\boldsymbol{V}_1 & \boldsymbol{V}_2\end{pmatrix}$;
- 最后, 得到了矩阵 $\boldsymbol{A}$ 的奇异值分解 $\boldsymbol{V}^H\boldsymbol{A}\boldsymbol{U}=\begin{pmatrix}\boldsymbol{S} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}$.

另外一方面, 如果 $\boldsymbol{A}\boldsymbol{A}^H$ 的特征向量更容易求出的话, 也可以从 $\boldsymbol{V}$ 出发, 再算出 $\boldsymbol{U}$.

在这里, 奇异值分解还有另外一种写法. 记

$$\boldsymbol{U}=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r,\boldsymbol{u}_{r+1},\cdots,\boldsymbol{u}_n),\boldsymbol{V}=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r,\boldsymbol{v}_{r+1},\cdots,\boldsymbol{v}_m).$$

代入计算得此时

$$\begin{aligned}\boldsymbol{A}=\boldsymbol{V}\begin{pmatrix}\boldsymbol{S} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\boldsymbol{U}^H&=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)\begin{pmatrix}\boldsymbol{S} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\begin{pmatrix}\boldsymbol{u}_1\\ \boldsymbol{u}_2\\ \vdots\\ \boldsymbol{u}_n\end{pmatrix}\\ &=\sum_{i=1}^{r}\delta_i\boldsymbol{v}_i\boldsymbol{u}_i^H.\end{aligned}$$

据最后一式, 我们成功地将矩阵 $\boldsymbol{A}$ 写成了一系列矩阵的线性组合. 其中, $\boldsymbol{v}_i\boldsymbol{u}_i^H$ 都是酉矩阵, 其前面的系数是从大到小排列的. 这在奇异值分解的后续应用中非常重要.

3.3 奇异值分解的编程实现

在 MATLAB 中, 函数svd可以用来求矩阵的奇异值分解, 例如对矩阵 $\boldsymbol{A}$, 用法如下.

```
[V S U] = svd(A);
```

根据该命令, 可以将矩阵 $\boldsymbol{A}$ 分解为

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{U},$$

其中 $\boldsymbol{V}$ 和 $\boldsymbol{U}$ 都是酉矩阵, $\boldsymbol{S}$ 的对角线上的元素都是 $\boldsymbol{A}$ 的奇异值.

例 3.7. *利用 MATLAB 求矩阵*

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 2 \\ 1 & 0 \end{pmatrix}$$

的奇异值分解.

解答 在 MATLAB 中输入如下代码.

```
A = [0  1;  -1  0;  0  2;  1  0];
[V S U] = svd(A);
```

最终计算得到的结果是

$$\boldsymbol{V} = \begin{pmatrix} -0.447 & 0 & -0.784 & -0.430 \\ 0 & 0.707 & -0.340 & 0.620 \\ -0.894 & 0 & 0.392 & 0.215 \\ 0 & -0.707 & -0.340 & 0.620 \end{pmatrix}, \boldsymbol{S} = \begin{pmatrix} 2.236 & 0 \\ 0 & 1.414 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \boldsymbol{U} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

从而我们得到了矩阵 $\boldsymbol{A}$ 的奇异值分解 $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{U}$.

3.4 奇异值分解的应用

在这里, 个人尝试自己编写了程序, 成功实现了**图片压缩**. 设此时目录中有一个彩色图片pic.png, 我们首先将其转化为黑白图片. 黑白图片上的每一点都有所谓的灰度, 从

而其对应着一个矩阵 $\boldsymbol{C}$. 我们尝试对这个矩阵 $\boldsymbol{C}$ 进行奇异值分解

$$\boldsymbol{C} = \sum_{i=1}^{r} \delta_i \boldsymbol{v}_i \boldsymbol{u}_i^H,$$

其中 $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_r$, 再对 $n < r$, 取上式的前 n 项 $\sum_{i=1}^{n} \delta_i \boldsymbol{v}_i \boldsymbol{u}_i^H$, 其可以作为矩阵 A 的近似替代. 这便是利用奇异值分解完成图像压缩的基本原理, 对应的代码如下.

```
% 输入图像
A = imread('.\pic.png');

% 将图像转化为黑白
B = rgb2gray(A);
imwrite(B, '.\pic-pre.png','png');
figure();
imshow(B);

% 计算SVD分解，并取前n项
C = im2double(B);
[U, S, V] = svd(C);
n = 50;
D = S(1, 1) * U(:, 1) * V(:, 1)';
for k = 2 : n
    D = D + S(k, k) * U(:, k) * V(:, k)';
end

% 输出压缩结果
P = im2uint8(D);
imwrite(P, '.\pic-print.png','png');
figure();
imshow(P);
```

在上面, 取 $n = 50$, 可以得到较为理想的分解式. 在实际应用中, 改变 n 的值, 可以得到不同的结果. 为了体现这一点, 在此取 $n = 1, 5, 10, 30, 50$, 并与原图作比较.

从而, 我们利用矩阵的奇异值分解, 成功实现了简易的图像压缩. 但事实上, 上面的

方法效率较低. 通过查阅资料得知, 通常图像压缩需要结合图片本身的特征, 例如图像中往往有一部分是连续的, 同时有一些棱角. 从而, 使用小波分解的方法, 效率会更高.

4 矩阵的谱分解

4.1 Jordan 标准形理论回顾

接下来, 我们处理的对象是**方阵**. 在前序课程中, 我们已经引入了将矩阵对角化, 或者化为 Jordan 标准形的理论和方式. 在这里不加证明地给出相关的定理, 并且指出, 这也是一种矩阵分解的方式.

定理 4.1 (Jordan 标准形)**.** *设 $\boldsymbol{A} \in \mathbb{C}_r^{n\times n}$, 则存在 $\boldsymbol{P} \in \mathbb{C}_n^{n\times n}$ 及 n 阶 Jordan 形矩阵 $\boldsymbol{J}$, 使得*

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J}.$$

其中 $\boldsymbol{J}$ 称为矩阵 $\boldsymbol{A}$ 的<u>Jordan 标准形</u>.

根据这个定理, 我们可以将矩阵 $\boldsymbol{A}$ 分解为

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}.$$

同时, 如果矩阵 $\boldsymbol{A}$ 所对应的 Jordan 标准形是由一系列一阶的 Jordan 块组成的, 则此时 $\boldsymbol{J}$ 为对角矩阵, 称此时矩阵 $\boldsymbol{A}$<u>可对角化</u>.

4.2 谱分解定理

谱分解的对象是**可对角化的矩阵**, 这种分解方式也被叫作特征分解.

首先, 我们需要明白分解后的结果应该是什么样的. 在谱分解中, 可以将一个可对角化的矩阵, 分解成一系列的幂等矩阵的线性组合. 而所谓的幂等矩阵, 与投影矩阵是等价的. 在此, 不再介绍较为复杂的幂等矩阵和投影矩阵的理论, 直接给出幂等矩阵的定义.

定义 4.2 (幂等矩阵)**.** *设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 且满足*

$$\boldsymbol{A}^2 = \boldsymbol{A},$$

则称其为<u>幂等矩阵</u>.

定理 4.3 (谱分解)**.** 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 且可对角化, 则存在一系列正交的幂等矩阵 $P_1, P_2, \cdots, P_k$, 使得

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{P}_i,$$

其中 λ_i 为 $\boldsymbol{P}$ 的特征值, 且

$$\sum_{i=1}^{n} \boldsymbol{P}_i = \boldsymbol{I}.$$

证明. 首先, 设 $\boldsymbol{A}$ 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 线性无关的特征向量为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$, 记 $\boldsymbol{P} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$, 则有

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_1^H \\ \boldsymbol{y}_2^H \\ \vdots \\ \boldsymbol{y}_n^H \end{pmatrix} = \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i \boldsymbol{y}_i^H,$$

其中 $\boldsymbol{y}_i^H$ 是矩阵 $\boldsymbol{P}^{-1}$ 的行向量. 记 $\boldsymbol{x}_i\boldsymbol{y}_i^H = \boldsymbol{P}_i$, 则有 $\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{P}_i$.

其次, 我们证明 $P_1, P_2, \cdots, P_k$ 是正交的幂等矩阵, 根据

$$\boldsymbol{I} = \boldsymbol{P}^{-1}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{y}_1^H \\ \boldsymbol{y}_2^H \\ \vdots \\ \boldsymbol{y}_n^H \end{pmatrix} (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = \begin{pmatrix} \boldsymbol{y}_1^H\boldsymbol{x}_1 & \boldsymbol{y}_1^H\boldsymbol{x}_2 & \cdots & \boldsymbol{y}_1^H\boldsymbol{x}_n \\ \boldsymbol{y}_2^H\boldsymbol{x}_1 & \boldsymbol{y}_2^H\boldsymbol{x}_2 & \cdots & \boldsymbol{y}_2^H\boldsymbol{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{y}_n^H\boldsymbol{x}_1 & \boldsymbol{y}_n^H\boldsymbol{x}_2 & \cdots & \boldsymbol{y}_n^H\boldsymbol{x}_n \end{pmatrix},$$

可以得知 $\boldsymbol{y}_i^H\boldsymbol{x}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j \end{cases}$, 从而

$$\boldsymbol{P}_i\boldsymbol{P}_j = (\boldsymbol{x}_i\boldsymbol{y}_i^H = \boldsymbol{P}_i)(\boldsymbol{x}_j\boldsymbol{y}_i^H = \boldsymbol{P}_j) = \boldsymbol{x}_i(\boldsymbol{y}_i^H\boldsymbol{x}_j)\boldsymbol{y}_j^H = \begin{cases} \boldsymbol{P}_i, & i = j, \\ \boldsymbol{O}, & i \neq j. \end{cases}$$

接下来, 我们证明 $\sum_{i=1}^{n} \boldsymbol{P}_i = \boldsymbol{I}$, 这是因为

$$\boldsymbol{I} = \boldsymbol{P}\boldsymbol{P}^{-1} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) \begin{pmatrix} \boldsymbol{y}_1^H \\ \boldsymbol{y}_2^H \\ \vdots \\ \boldsymbol{y}_n^H \end{pmatrix} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i^H = \sum_{i=1}^{n} \boldsymbol{P}_i.$$

□

事实上, 谱分解本质上还是矩阵的对角化, 只是将其换了一种写法. 我们还可以深入地证明, 每个矩阵 $\boldsymbol{P}_i$ 都可以由特征值直接算出, 在此就不再阐述.

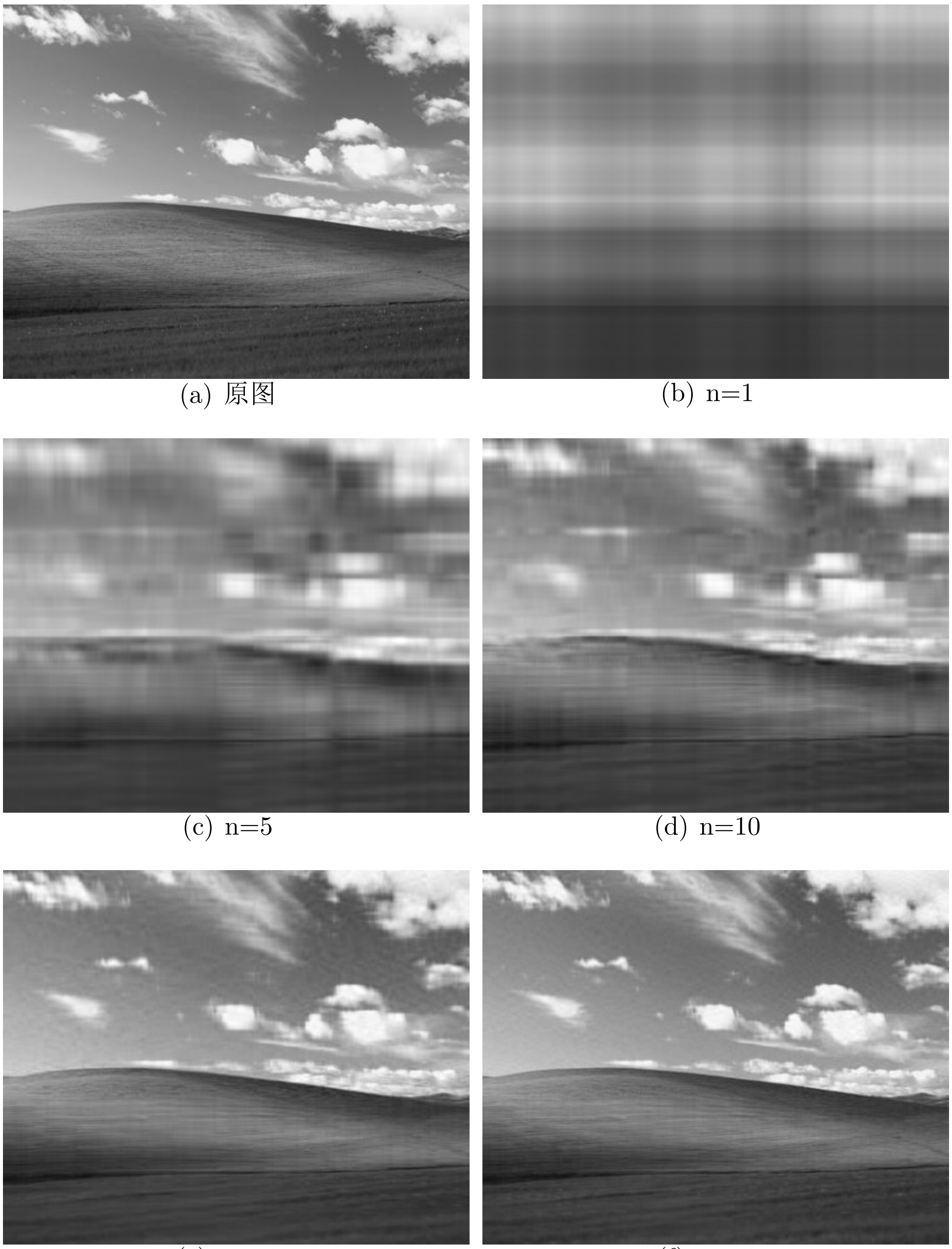
(a) 原图 (b) n=1
(c) n=5 (d) n=10
(e) n=30 (f) n=50

图 1: 图像压缩示例