

MIT 18.06 线性代数

第三十一讲：基变换和图像压缩

整理自 MIT 18.06 课程及相关公开笔记

引言

本讲将探讨基变换的一个重要应用——**图像压缩** (Image Compression)。通过选择合适的基向量，可以极大地提高数据存储和传输的效率。这一过程不仅展示了基变换的实用价值，也揭示了线性变换在不同基下所对应的矩阵之间的内在联系。

1 图像压缩的基本思想

1.1 图像的向量表示

首先，考虑一幅 512×512 像素的黑白图像。每个像素点的亮度值在 $0 \leq x_i \leq 255$ 范围内变化，占用 8 个比特 (bits)。因此，这整幅图像可以被抽象为一个包含 $n = 512 \times 512$ 个分量的向量 $\mathbf{x} \in \mathbb{R}^n$ 。对于彩色图像，每个像素需要三个数值（如 RGB 三个通道），向量长度是黑白图像的 3 倍。

1.2 选择合适的基

在图像压缩中，标准基并不总是最优的选择。标准基的每个向量只负责描述单个像素的亮度，对于整幅图像中普遍存在的“颜色相近或相同”的区域（例如黑板、纯色背景等），标准基无法利用这一特点。

示例：一张全黑的图像，若使用标准基表示，需要将每一个分量的值都设为 0 或 255，需要存储 n 个数据。但如果选择一个新的基向量 $\mathbf{f}_1 = [1 \ 1 \ \dots \ 1]^T$ （代表全黑或全白的低频信号），则这张图像可以仅用 $255\mathbf{f}_1$ 一个系数来表示，存储量从 n 大幅降至 1。

因此，图像压缩算法的核心思想就是：**寻找一组基，使得图像在该基下的系数只有少数几个较大，其余系数接近于零，从而可以通过丢弃小系数来达到压缩的目的。**

2 两类重要的基

2.1 傅里叶基 (Fourier Basis)

JPEG (Joint Photographic Experts Group, 联合图像专家组) 压缩标准采用的是与傅里叶矩阵密切相关的**傅里叶基**。该算法将图像划分为 8×8 的区块，每个区块包含 64 个像素，然后在这 64 维空间中进行基变换。

$\mathbb{R}^8$ 中的傅里叶基向量为 (其中 $\omega = e^{2\pi i/8}$):

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega \\ \omega^2 \\ \omega^3 \\ \omega^4 \\ \omega^5 \\ \omega^6 \\ \omega^7 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega^2 \\ \omega^4 \\ \omega^6 \\ \omega^8 \\ \omega^{10} \\ \omega^{12} \\ \omega^{14} \end{bmatrix}, \dots, \begin{bmatrix} 1 \\ \omega^7 \\ \omega^{14} \\ \omega^{21} \\ \omega^{28} \\ \omega^{35} \\ \omega^{42} \\ \omega^{49} \end{bmatrix}$$

JPEG 实际使用的是傅里叶基中 ω^{jk} 的实数部分, 即余弦函数, 因此该方法被称为**离散余弦变换** (Discrete Cosine Transform, DCT)。整个压缩过程为:

$$\text{信号 } \mathbf{x} \xrightarrow{\text{无损}} 64 \text{ 个系数 } \mathbf{c} \xrightarrow{\text{有损压缩}} \hat{\mathbf{c}} (\text{许多为零}) \rightarrow \hat{\mathbf{x}} = \sum \hat{c}_i \mathbf{v}_i$$

第一步是基变换, 不丢失信息; 第二步是根据阈值 (人眼可分辨的差异) 丢弃小系数, 得到压缩后的近似图像。

2.2 小波基 (Wavelet Basis)

小波基是傅里叶基的有力竞争者, 在 JPEG2000 标准中使用的是离散小波变换。在 $\mathbb{R}^8$ 中, 小波基向量如下:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

这些向量除了全 1 向量外, 其非零元素均为 1 与 -1 的交替, 部分向量中甚至包含大量的零元素。小波基向量之间是相互正交的, 可以进一步规范化为标准正交基。

3 基变换与矩阵表示

3.1 从标准基到新基

设矩阵 W 的列向量是新基向量 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ 。如果 $\mathbf{x}$ 是旧基 (标准基) 下的向量, $\mathbf{c}$ 是新基下的坐标向量, 则它们之间的关系为:

$$\mathbf{x} = W\mathbf{c}$$

因此, 可以通过求解线性方程组得到新基下的坐标:

$$\mathbf{c} = W^{-1}\mathbf{x}$$

3.2 好基的标准

在图像压缩中，一个“好”的基需要满足以下两个条件：

1. **快速运算** (FFT、FWT)：基矩阵及其逆矩阵与向量的乘法可以快速完成；
2. **良好的压缩性** (Few is Enough)：图像可以用极少数的基向量来近似，即大部分系数 c_i 很小可以安全地置为零。

3.3 变换矩阵的相似性

考虑一个线性变换 $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ 。当使用一组基 $\mathbf{v}_1, \dots, \mathbf{v}_n$ 时，变换矩阵为 A ；当改用另一组基 $\mathbf{w}_1, \dots, \mathbf{w}_n$ 时，变换矩阵为 B 。

对于同一个线性变换，不同基下的变换矩阵之间必然存在**相似**关系：

$$B = M^{-1}AM$$

其中 M 是基变换矩阵（通常可以取 $M = W$ ）。

特别情形：如果选择的基恰好是变换的特征向量，即 $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ ，那么变换矩阵 A 就是一个对角矩阵 Λ ，对角线元素为特征值。使用特征向量作为基是压缩的理想情况，但寻找图像的特征向量计算量过大，因此退而求其次，选择傅里叶基或小波基进行近似。

总结

- **图像压缩**的本质是选择更合适的基，使得图像在该基下的系数尽可能稀疏（大部分为零），从而减少存储和传输量。
- **傅里叶基**和**小波基**因其正交性和快速运算特性，成为了图像压缩中的主流选择。
- **基变换**对向量的影响体现在坐标的变化： $\mathbf{x} = W\mathbf{c}$ 。
- **基变换**对线性变换矩阵的影响体现在矩阵的相似性： $B = M^{-1}AM$ 。