



Math Notes

My Math Note

Xia Wenxuan



Written by Xia Wenxuan, 2021

<https://github.com/gegeji>

All the material in this document is taken from my textbook or video material, and some material uses optical character recognition (OCR) to aid input. It may contains typographical or content inaccuracies. This document is for my personal study only. I am not responsible for the accuracy of the content of the text.

Edited and Revised on November 1, 2021

Contents

I	Vectors	
1	对向量的介绍	8
1.1	Vector	8
1.2	Vector Space	9
1.3	向量运算	9
1.4	内积	10
1.4.1	常用的内积等式	11
1.5	Cauchy-Schwartz Inequality	12
1.6	浮点运算	13
2	Linear Function	14
2.1	Linear Function	14
2.2	泰勒展开	15
2.3	Regression Model	17
3	Norm and Distance	19
3.1	Vector Norm	19
3.2	Root Mean Square Value (RMS)	21
3.3	Chebyshev's Inequality	21
3.4	Distance	21
3.4.1	Feature Distance and Nearest Neighbor	22

3.5	Standard Derivation	22
3.6	Angle	22
3.6.1	相关系数	23
4	优化问题初步	24
4.1	优化问题引入	24
4.2	Convex Set	25
4.3	向量偏导	26
4.4	投影问题: 标量优化问题	27
4.5	Clustering	28
5	Linear Independence	30
5.1	线性相关、线性无关	30
5.2	Basis	31
5.3	标准正交向量	31
5.4	Gram-Schmidt Algorithm	33
5.4.1	The Analysis of Gram-Schmidt Algorithm	34

II

Matrices

6	Matrices	37
6.1	Matrices	37
6.2	矩阵运算	38
6.2.1	Matrix Power	41
6.2.2	矩阵乘法的算法复杂度	41
6.2.3	矩阵向量乘积复杂度	42
6.3	Special Matrices and Matrices in Different Applications	43
6.3.1	$f(x) = Ax$ 中的 A	43
6.3.2	Selectors	44
6.3.3	图论: 节点弧关联矩阵	45
6.3.4	Convolution	45
6.3.5	多项式	46
6.3.6	Fourier Transform	46
6.3.7	Semi-Definite Matrices	47
6.4	Gram 矩阵	48
7	Matrices Norms	49
7.1	矩阵范数	49
8	适定问题	53
8.1	The Definition of Well-posed Problem	53
8.2	绝对误差的界限	54
8.3	相对误差的界限	54

9	Inverse of Matrices	55
9.1	Left Inverse, Right Inverse, Inverse	55
9.2	Linear Equation Systems	56
9.2.1	线性方程组求解	56
9.3	Fundamental Theorem of Linear Algebra	57
9.4	Invertible Matrices	57
9.5	转置和共轭转置的逆	60
9.6	Gram Matrix 非奇异的性质	60
9.7	伪逆	60
10	Orthogonal Matrices	64
10.1	预备知识	64
10.1.1	标准正交向量	64
10.1.2	Gram 矩阵与标准正交的关系	64
10.1.3	矩阵-向量乘积与标准正交的关系	64
10.1.4	左可逆性与正交的关系	65
10.2	正交矩阵	66
10.3	Permutation Matrices	66
10.4	平面旋转	67
10.5	Householder Matrix	67
10.5.1	The Geometry of Householder Transformation	68
10.6	正交矩阵乘积	69
10.7	具有正交矩阵的线性方程	69
10.7.1	The Complexity of the Multiplication Ax of Orthogonal Matrix A	69
10.8	列标准正交的高矩阵	69
10.9	值域范围、列空间	70
10.9.1	投影到列标准正交的矩阵 A 的列空间	70
11	QR 分解与 Householder 变换	72
11.1	Triangular Matrices	72
11.1.1	高斯消元法	72
11.1.2	The Inverses of Triangular Matrices	73
11.2	QR Factorization	73
11.2.1	复矩阵的 QR 分解	76
11.3	QR 分解的应用	76
11.3.1	QR 分解和求解线性方程组 $Ax = b$	76
11.3.2	QR 分解和求解伪逆 A^\dagger 、逆 A^{-1}	77
11.3.3	A 的列空间和 Q 的列空间相同	77
11.3.4	往 A 列空间上的投影也是往 Q 列空间上的投影	78
11.4	QR Algorithm Using Gram-Schmidt Algorithm	78
11.4.1	Gram-Schmidt Algorithm	81
11.4.2	基于 Gram-Schmidt 方法进行 QR 分解的时间复杂度	82

11.5	The Numerical Instability of QR Decomposition based on Gram-Schmidt Algorithm	82
11.6	QR Decomposition Using Householder Transformation	84
11.6.1	Householder Matrix	84
11.6.2	构造反射算子	85
11.6.3	Householder 三角化	87
11.6.4	An Example for Householder Algorithm	90
11.6.5	Complexity of Householder Algorithm	91
11.7	Householder 变换进行 QR 分解的 Q 因子	91
11.7.1	矩阵-向量积 $H_k x$ 算法复杂度	92
11.8	Fast Orthogonalization	92

III

Extensive Reading

12	Fourier Series, Fourier Transform	96
12.1	基本概念	96
12.2	Fourier Series	98
12.3	Fourier transform	101
12.4	Discrete Fourier Transform	102
13	Factorization of Matrices	104
13.1	主要的矩阵分解	104
13.2	$A = LU$	104
14	List Of Definitions	106

Vectors

1	对向量的介绍	8
1.1	Vector	
1.2	Vector Space	
1.3	向量运算	
1.4	内积	
1.5	Cauchy-Schwartz Inequality	
1.6	浮点运算	
2	Linear Function	14
2.1	Linear Function	
2.2	泰勒展开	
2.3	Regression Model	
3	Norm and Distance	19
3.1	Vector Norm	
3.2	Root Mean Square Value (RMS)	
3.3	Chebyshev's Inequality	
3.4	Distance	
3.5	Standard Derivation	
3.6	Angle	
4	优化问题初步	24
4.1	优化问题引入	
4.2	Convex Set	
4.3	向量偏导	
4.4	投影问题: 标量优化问题	
4.5	Clustering	
5	Linear Independence	30
5.1	线性相关、线性无关	
5.2	Basis	
5.3	标准正交向量	
5.4	Gram-Schmidt Algorithm	

1. 对向量的介绍

1.1 Vector

Definition 1.1.1 — Vector. 一个有序的数字列表.

$$\begin{bmatrix} -1.1 \\ 0.0 \\ 3.6 \\ -7.2 \end{bmatrix} \text{ 或者 } \begin{pmatrix} -1.1 \\ 0.0 \\ 3.6 \\ -7.2 \end{pmatrix} \text{ 或者 } (-1.1, 0, 3.6, -7.2)$$

表中的数字是元素 (项、系数、分量). 元素的数量是向量的大小 (维数, 长度). 大小为 n 的向量称为 n 维向量. 向量中的数字通常被称作标量.

用符号来表示向量, 比如 a, b , 一般小写字母表示. 其它表示形式 \mathbf{g}, \vec{a}

Definition 1.1.2 — n 维向量 a 的第 i 元素. n 维向量 a 的第 i 元素表示为 a_i .

有时 i 指的是向量列表中的第 i 个向量.

Definition 1.1.3 — $a = b$. 对于所有 i , 如果有 $a_i = b_i$, 则称两个相同大小的向量 a 和 b 是相等的, 可写成 $a = b$

Definition 1.1.4 — stacked vector. 假设 b, c, d 是大小为 m, n, p 的向量

$$a = \begin{bmatrix} b \\ c \\ d \end{bmatrix}$$

$$a = (b_1, b_2, \dots, b_m, c_1, c_2, \dots, c_n, d_1, d_2, \dots, d_p)$$

Definition 1.1.5 — 零向量. 所有项为 0 的 n 维向量表示为 0_n 或者 0 .

Definition 1.1.6 — 全一向量. 所有项为 1 的 n 维向量表示为 $\mathbf{1}_n$ 或者 $\mathbf{1}$.

Definition 1.1.7 — 单位向量. 当第 i 项为 1, 其余项为 0 时表示为 e_i

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Definition 1.1.8 — 稀疏向量. 如果一个向量的许多项都是 0, 该向量为稀疏 (Sparse) 的. 稀疏向量能在计算机上高效地存储和操作.

$\text{nnz}(x)$ 是指向量 x 中非零的项数 (number of non-zeros), 有时用 ℓ_0 表示.

向量 $x = (x_1, x_2)$ 可以在二维中表示一个位置或一个位移、图像、单词统计、颜色等.

1.2 Vector Space

Definition 1.2.1 — 向量空间 V . 设 V 是非空子集, P 是一数域, 向量空间 V 满足:

1. 向量加法: $V + V \rightarrow V$, 记作 $\forall x, y \in V$, 则 $x + y \in V$ (加法封闭)
 2. 标量乘法: $P \times V \rightarrow V$, 记作 $\forall x \in V, \lambda \in P$, 则 $\lambda x \in V$ (乘法封闭)
- 上述两个运算满足下列八条规则 ($\forall x, y, z \in V, \lambda, \mu \in P$)
1. $x + y = y + x$ (交换律)
 2. $x + (y + z) = (x + y) + z$ (结合律)
 3. V 存在一个零元素, 记作 0 , $x + 0 = x$
 4. 存在 x 的负元素, 记作 $-x$, 满足 $x + (-x) = 0$
 5. $\forall x \in V$, 都有 $1x = x, 1 \in P$
 6. $\lambda(\mu x) = (\lambda\mu)x$
 7. $(\lambda + \mu)x = \lambda x + \mu x$
 8. $\lambda(x + y) = \lambda x + \lambda y$

Corollary 1.2.1 向量空间也称为线性空间.

Corollary 1.2.2 如果 $x, y \in \mathbb{R}^2$, 则 $x + y \in \mathbb{R}^2, \lambda x \in \mathbb{R}^2 (\lambda \in \mathbb{R})$.

Definition 1.2.2 — 数域. 数的非空集合 P , 且其中任意两个数的和、差、积、商 (除数不为零) 仍属于该集合, 则称数集 P 为一个数域.

■ **Example 1.1** 有理数 \mathbb{Q} ■

■ **Example 1.2** 实数 \mathbb{R} ■

$x, y \in \mathbb{R}, x = 1, y = 2 \quad x + y \in \mathbb{R}, x \times y \in \mathbb{R}$ ■

■ **Example 1.3** 复数 \mathbb{C} ■

1.3 向量运算

Definition 1.3.1 — 向量加法. n 维向量 a 和 b 可以相加, 求和形式表示为 $a + b$.

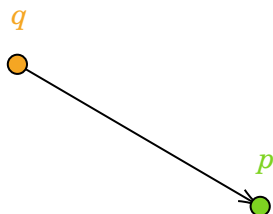
设向量 a, b, c 是向量空间 V 的元素, 即 $a, b, c \in V$.

1. 交换律: $a + b = b + a$
2. 结合律: $(a + b) + c = a + (b + c)$ (因此可写成 $a + b + c$)
3. $a + 0 = 0 + a = a$
4. $a - a = 0$

Corollary 1.3.1 — 向量位移相加. 如果二维向量 a 和 b 都表示位移, 则它们的位移之和为 $a + b$

■ **Example 1.4** 点 q 到点 p 的位移是 $p - q$.

Figure 1.1: The translation from q to p



Definition 1.3.2 — 标量与向量的乘法.

$$\beta a = \begin{bmatrix} \beta a_1 \\ \vdots \\ \beta a_n \end{bmatrix}$$

标量 β, γ 与向量 a, b 进行乘法, 有如下性质:

1. 结合律: $(\beta\gamma)a = \beta(\gamma a)$
2. 左分配律: $(\beta + \gamma)a = \beta a + \gamma a$
3. 右分配律: $\beta(a + b) = \beta a + \beta b$

Definition 1.3.3 — 线性组合. 对于向量 a_1, \dots, a_m 和标量 β_1, \dots, β_m ,

$$\beta_1 a_1 + \dots + \beta_m a_m$$

是向量的线性组合. β_1, \dots, β_m 是该向量的系数.

■ **Example 1.5** 对于任何向量 $b \in \mathbb{R}^n$, 有如下等式

$$b = b_1 e_1 + \dots + b_n e_n, b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

1.4 内积

Definition 1.4.1 — 内积. 在数域 \mathbb{R} 上的向量空间 V , 定义函数 $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$, 满足:

1. $\langle a, a \rangle \geq 0, \forall a \in V$, 当且仅当 $a = 0$ 时 $\langle a, a \rangle = 0$
2. $\langle \alpha a + \beta b, c \rangle = \alpha \langle a, c \rangle + \beta \langle b, c \rangle, \forall \alpha, \beta \in \mathbb{R}$, 且 $a, b, c \in V$
3. $\langle a, b \rangle = \langle b, a \rangle, \forall a, b \in V$

函数 $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ 成为内积.

■ **Example 1.6** 在向量空间 \mathbb{R}^n 上, 计算两个向量对应项相乘之后求和函数

$$\langle a, b \rangle = a_1 b_1 + a_2 b_2 + \dots + a_n b_n = a_b^T$$

$$\text{where } a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^n. \quad \blacksquare$$

Proof. $\langle a, a \rangle = a_1 a_1 + a_2 a_2 + \cdots + a_n a_n = \sum_{i=1}^n a_i^2 \geq 0$, $\langle a, a \rangle = 0$, 则 $a = 0$

$$\begin{aligned} \langle \alpha a + \beta b, c \rangle &= (\alpha a_1 + \beta b_1) c_1 + (\alpha a_2 + \beta b_2) c_2 + \cdots + (\alpha a_n + \beta b_n) c_n \\ &= \alpha \sum_{i=1}^n a_i c_i + \beta \sum_{i=1}^n b_i c_i \\ &= \alpha \langle a, c \rangle + \beta \langle b, c \rangle \\ \langle a, b \rangle &= a^T b = b^T a = \langle b, a \rangle \end{aligned} \quad \blacksquare$$

内积的性质：交换律、结合律、分配律.

交换律： $a^T b = b^T a$

结合律： $(\gamma a)^T b = \gamma (a^T b)$

分配律： $(a + b)^T c = a^T c + b^T c$

1.4.1 常用的内积等式

Corollary 1.4.1 — 选出第 i 项.

$$e_i^T a = a_i$$

Corollary 1.4.2 — 向量每一项之和.

$$\mathbf{1}^T a = a_1 + \cdots + a_n$$

Corollary 1.4.3 — 向量每一项的平方和.

$$a^T a = a_1^2 + \cdots + a_n^2$$

Corollary 1.4.4 — 向量元素的平均值.

$$(\mathbf{1}/n)^T a = (a_1 + \cdots + a_n) / n$$

Corollary 1.4.5 — Selective sum. Let b be a vector all of whose entries are either 0 or 1. Then

$$b^T a$$

is the sum of the elements in a for which $b_i = 1$.

Definition 1.4.2 — The sum of block vectors. If the vectors a and b are block vectors, and the corresponding blocks have the same sizes (in which case we say they conform), then

$$a^T b = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}^T \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} = a_1^T b_1 + \cdots + a_k^T b_k$$

内积用途很广.

■ **Example 1.7** — 计算同时出现的项目数.

$$a = (0, 1, 1, 1, 1, 1, 1), \quad b = (1, 0, 1, 0, 1, 0, 0)$$

Here we have $a^T b = 2$, which is the number of objects in both A and B (i.e., objects 3 and 5). ■

■ **Example 1.8** — **Weights, features, and score.** When the vector f represents a set of *features* of an object, and w is a vector of the same size (often called a *weight vector*), the inner product $w^T f$ is the sum of the feature values, scaled (or weighted) by the weights, and is sometimes called a *score*. ■

■ **Example 1.9**

$$p(x) = c_1 + c_2 x + \cdots + c_{n-1} x^{n-2} + c_n x^{n-1}$$

Let t be a number, $z = (1, t, t^2, \dots, t^{n-1})$ be the n -vector of powers of t . Then

$$c^T z = p(t)$$

■

1.5 Cauchy-Schwartz Inequality

Theorem 1.5.1 — **Cauchy-Schwartz Inequality.** 设 $\langle \cdot, \cdot \rangle$ 是向量空间 V 上的内积, $\forall x, y \in V$, 则有

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

Proof. 令 $\lambda \in \mathbb{R}$, 则有

$$0 \leq \langle x + \lambda y, x + \lambda y \rangle = \langle x, x \rangle + \lambda \langle y, x \rangle + \lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle = \langle x, x \rangle + 2\lambda \langle y, x \rangle + \lambda^2 \langle y, y \rangle$$

则

$$\lambda^2 \langle y, y \rangle + 2\lambda \langle y, x \rangle + \langle x, x \rangle \geq 0, \forall \lambda \in \mathbb{R}$$

所以

$$\Delta = (2\langle y, x \rangle)^2 - 4\langle y, y \rangle \langle x, x \rangle \leq 0$$

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

当 $|\langle x, y \rangle|^2 = \langle x, x \rangle \langle y, y \rangle$ 时, 有

$$\langle x, x \rangle^2 + 2\lambda \langle y, x \rangle + \lambda^2 \langle y, y \rangle = 0$$

也即

$$\langle x + \lambda y, x + \lambda y \rangle = 0$$

因此 $x + \lambda y = 0$, 即 $x = -\lambda y$. ■

Theorem 1.5.2

$$\left(\sum_{i=1}^n u_i v_i \right)^2 \leq \left(\sum_{i=1}^n u_i^2 \right) \left(\sum_{i=1}^n v_i^2 \right)$$

Proof. The Cauchy-Schwarz inequality can be proved using only ideas from elementary algebra in this case. Consider the following quadratic polynomial in x

$$0 \leq (u_1 x + v_1)^2 + \cdots + (u_n x + v_n)^2 = \left(\sum_i u_i^2 \right) x^2 + 2 \left(\sum_i u_i v_i \right) x + \sum_i v_i^2$$

Since it is nonnegative, it has at most one real root for x , hence its discriminant is less than or equal to zero. That is,

$$\left(\sum_i u_i v_i \right)^2 - \left(\sum_i u_i^2 \right) \left(\sum_i v_i^2 \right) \leq 0$$

which yields the Cauchy-Schwarz inequality. ■

1.6 浮点运算

计算机以浮点格式存储(实)数值.

基本的算术运算(加法, 乘法等)被称为浮点运算(flop).

算法或操作的时间复杂度: 作为输入维数的函数所需要的浮点运算总数.

算法复杂度通常以非常粗略地近似估算.

(程序)执行时间的粗略估计: 计算机速度/flops

目前的计算机大约是 1Gflops/秒 (10^9 flops/秒)

Corollary 1.6.1 假设有 n 维向量 x 和 y :

- $x + y$ 需要 n 次加法, 所以时间复杂度为 (n) flops.
- $x^T y$ 需要 n 次乘法和 $n - 1$ 次加法, 所以时间复杂度为 $(2n - 1)$ flops.
- 对于 $x^T y$, 通常将其时间复杂度简化为 $2n$, 甚至为 n .
- 当 x 或 y 是稀疏的时候, 算法的实际运算时间会比理论时间更少.

2. Linear Function

2.1 Linear Function

Definition 2.1.1 — Linear Function. f 是一个将 n 维向量映射成数的函数.

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

线性函数 f 满足以下两个性质 ($k \in \mathbb{R}, x, y \in \mathbb{R}^n$) :

- 齐次性 (homogeneity): $f(kx) = kf(x)$
- 叠加性 (Additivity): $f(x + y) = f(x) + f(y)$

■ **Example 2.1** 求平均值: $f(x) = \frac{1}{n} \sum_{i=1}^n x_i$ 为线性函数. ■

■ **Example 2.2** 求最大值: $f(x) = \max \{x_1, x_2, \dots, x_n\}$ 并不是线性函数. ■

Proof. 令 $x = (1, -1), y = (-1, 1), \alpha = 0.5, \beta = 0.5$, 有

$$f(\alpha x + \beta y) = 0 \neq \alpha f(x) + \beta f(y) = 1$$

$$\begin{aligned} f(x + y) &= \max \{x_1 + y_1, x_2 + y_2, \dots, x_n + y_n\} \\ &\leq \max \{x_1, x_2, \dots, x_n\} + \max \{y_1, y_2, \dots, y_n\} \\ &\leq f(x) + f(y) \end{aligned}$$

■

Theorem 2.1.1 设 $\alpha_1, \dots, \alpha_m \in \mathbb{R}, u_1, \dots, u_m \in \mathbb{R}^n$, 则线性函数 f 满足

$$\begin{aligned}
f(\alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_m u_m) &= f(\alpha_1 u_1) + f(\alpha_2 u_2 + \dots + \alpha_m u_m) \\
&= \alpha_1 f(u_1) + f(\alpha_2 u_2 + \dots + \alpha_m u_m) \\
&= \alpha_1 f(u_1) + \alpha_2 f(u_2) + \dots + \alpha_m f(u_m)
\end{aligned}$$

Definition 2.1.2 — 内积函数 (inner product function). 对于 n 维向量 a , 满足以下形式的函数被称为内积函数

$$f(x) = a^T x = a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

上述 $f(x)$ 可以看作是每项 x_i 的加权之和.

Corollary 2.1.2 内积函数都是线性的.

Proof.

$$\begin{aligned}
f(\alpha x + \beta y) &= a^T (\alpha x + \beta y) \\
&= a^T (\alpha x) + a^T (\beta y) \\
&= \alpha (a^T x) + \beta (a^T y) \\
&= \alpha f(x) + \beta f(y)
\end{aligned}$$

■

Corollary 2.1.3 所有线性函数都是内积.

$$\begin{aligned}
f(x) &= f(x_1 e_1 + x_2 e_2 + \dots + x_n e_n) \\
&= x_1 f(e_1) + x_2 f(e_2) + \dots + x_n f(e_n)
\end{aligned}$$

Proof. 假设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是线性函数, 那么可用 $f(x) = a^T x$ 来表示, a 为常量.

$$\begin{aligned}
f(x) &= f(x_1 e_1 + x_2 e_2 + \dots + x_n e_n) \\
&= x_1 f(e_1) + x_2 f(e_2) + \dots + x_n f(e_n)
\end{aligned}$$

■

Definition 2.1.3 — 仿射函数 (affine function). 其一般形式为 $f(x) = a^T x + b$, 其中 $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ 为标量.

Theorem 2.1.4 函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为仿射函数需要满足

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y), \alpha + \beta = 1, \alpha, \beta \in \mathbb{R}, x, y \in \mathbb{R}^n$$

2.2 泰勒展开

Definition 2.2.1 — 函数 f 第 i 个分量的一阶偏导数. 假设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 函数 f 在 z 点可微

$$\begin{aligned}\frac{\partial f}{\partial z_i}(z) &= \lim_{t \rightarrow 0} \frac{f(z_1, \dots, z_{i-1}, z_i + t, z_{i+1}, \dots, z_n) - f(z)}{t} \\ &= \lim_{t \rightarrow 0} \frac{f(z + te_i) - f(z)}{t}\end{aligned}$$

Definition 2.2.2 — f 在点 z 的梯度.

$$\nabla f(z) = \begin{bmatrix} \frac{\partial f}{\partial z_1}(z) \\ \vdots \\ \frac{\partial f}{\partial z_n}(z) \end{bmatrix}$$

Definition 2.2.3 — **Taylor's Approximation.** 假设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 函数 f 在 z 点充分光滑, 即处处可导.

$$\begin{aligned}f(x) &= f(z) + \frac{\partial f}{\partial x_1}(z)(x_1 - z_1) + \frac{\partial f}{\partial x_2}(z)(x_2 - z_2) + \dots + \frac{\partial f}{\partial x_n}(z)(x_n - z_n) \\ &\quad + \frac{1}{2!} \sum_{j=1}^n \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(z)(x_i - z_i)(x_j - z_j) + \dots\end{aligned}$$

■ **Example 2.3** 泰勒公式利用多项式在一点附近逼近函数

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots + (-1)^{k-1} \frac{x^{2k-1}}{(2k-1)!} + \frac{\sin \left[\xi + (2k+1)\frac{\pi}{2} \right]}{(2k+1)!} x^{2k+1}$$

一次逼近: $\sin x \approx x$

三次逼近: $\sin x \approx x - \frac{x^3}{3!}$ ■

Proof.

$$f(x) = P_n(x) + R_n(x)$$

$$P_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \dots + a_n(x - x_0)^n$$

$$R_n(x) = o(x - x_0)^n$$

$$f(x) \approx P_n(x)$$

$$\therefore P_n(x_0) = f(x_0), P'_n(x_0) = f'(x_0), P''_n(x_0) = f''(x_0), \dots, P_n^{(n)}(x_0) = f^{(n)}(x_0)$$

$$\text{要求 } P_n(x_0) = f(x_0) \Rightarrow a_0 = f(x_0)$$

$$P'_n(x) = a_1 + 2a_2(x - x_0) + \dots + na_n(x - x_0)^{n-1} \Rightarrow a_1 = f'(x_0)$$

$$\text{依此类推. } a_n = \frac{f^{(n)}(x_0)}{n!}$$
 ■

Corollary 2.2.1 — n 阶泰勒多项式.

$$\begin{aligned}P_n(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 \\ &\quad + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n\end{aligned}$$

where $a_n = \frac{f^{(n)}(x_0)}{n!}$

Corollary 2.2.2 — 对于高阶余项的公式. 带拉格朗日余项的泰勒公式

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}$$

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1} \quad (\xi \text{ 在 } x_0 \text{ 与 } x \text{ 之间})$$

Corollary 2.2.3 — 麦克劳林 (Maclaurin) 公式. 在零点展开麦克劳林 (Maclaurin) 公式

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \frac{f^{(n+1)}(\theta x)}{(n+1)!}x^{n+1} \quad (0 < \theta < 1)$$

Definition 2.2.4 — 一阶泰勒公式. 假设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 函数 f 在 z 点可导

$$\hat{f}(x) = f(z) + \frac{\partial f}{\partial x_1}(z)(x_1 - z_1) + \cdots + \frac{\partial f}{\partial x_n}(z)(x_n - z_n)$$

当 x 非常接近 z 时, $\hat{f}(x)$ 也非常接近 $f(z)$. $\hat{f}(x)$ 是关于 x 的一个仿射函数.

Corollary 2.2.4 — 一阶泰勒公式的内积形式.

$$\hat{f}(x) = f(z) + \nabla f(z)^T(x - z) \quad (\nabla f(z) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(z) \\ \vdots \\ \frac{\partial f}{\partial x_n}(z) \end{bmatrix})$$

一维时, $\hat{f}(x) = f(z) + f'(z)(x - z)$.

■ **Example 2.4**

$$f(x) = x_1 - 3x_2 + e^{2x_1+x_2-1}$$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \end{bmatrix} = \begin{bmatrix} 1 + 2e^{2x_1+x_2-1} \\ -3 + e^{2x_1+x_2-1} \end{bmatrix}$$

函数 f 在 o 点的一阶泰勒公式为:

$$\hat{f}(x) = f(0) + \nabla f(0)^T(x - 0) = e^{-1} + (1 + 2e^{-1})x_1 + (-3 + e^{-1})x_2$$

■

2.3 Regression Model

Definition 2.3.1 — Regression Model. 回归模型 (regression model) 为关于 x 的仿射函数

$$\hat{y} = x^T \beta + v$$

x 是特征向量 (*feature vector*), 它的元素 x_i 称为回归元 (*regressors*). n 维向量 β 是权重向量 (*weight vector*). 标量 v 是偏移量 (*offset*). 标量 \hat{y} 是预测值 (*prediction*). 表示某个实际结果或因变量, 用 y 表示.

3. Norm and Distance

3.1 Vector Norm

Definition 3.1.1 — Vector Norm. 在向量空间中存在一个函数 $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$, 且满足以下条件

- 齐次性: $\|\alpha x\| = |\alpha| \|x\|$, $\alpha \in \mathbb{R}$ 且 $x \in \mathbb{R}^n$;
- 三角不等式: $\|x + y\| \leq \|x\| + \|y\|$, $x, y \in \mathbb{R}^n$;
- 非负性: $\|x\| \geq 0$, $x \in \mathbb{R}^n$ 且 $\|x\| = 0 \Leftrightarrow x = 0$;

则称 $\|\cdot\|$ 为向量范数.

■ **Example 3.1 — ℓ_1 -范数 (曼哈顿范数, Manhattan norm)** .

$$\|x\|_1 = |x_1| + |x_2| + \dots + |x_n| \quad x, y \in \mathbb{R}^n, \alpha \in \mathbb{R}$$

Proof.

$$\|\alpha x\|_1 = |\alpha x_1| + |\alpha x_2| + \dots + |\alpha x_n| = |\alpha| \|x\|_1 \geq 0$$

$$\|x + y\|_1 = |x_1 + y_1| + \dots + |x_n + y_n| \leq |x_1| + |y_1| + \dots + |x_n| + |y_n| = \|x\|_1 + \|y\|_1$$

■ **Example 3.2 — ℓ_2 -范数 (欧几里得范数, Euclidean norm)** .

$$\|x\|_2 = \sqrt{(x_1^2 + x_2^2 + \dots + x_n^2)} = \sqrt{x^T x} = (\langle x, x \rangle)^{\frac{1}{2}}$$

Proof.

$$\|\alpha x\|_2 = (\langle \alpha x, \alpha x \rangle)^{\frac{1}{2}} = |\alpha| (\langle x, x \rangle)^{\frac{1}{2}} = |\alpha| \|x\|_2$$

$$\begin{aligned}
\|x+y\|_2^2 &= \langle x+y, x+y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\
&= \|x\|_2^2 + 2\langle x, y \rangle + \|y\|_2^2 \leq \|x\|_2^2 + 2\|x\|_2\|y\|_2 + \|y\|_2^2 \\
&= (\|x\|_2 + \|y\|_2)^2
\end{aligned}$$

$$\|x+y\|_2 \leq \|x\|_2 + \|y\|_2$$

■

Corollary 3.1.1 — 柯西—施瓦茨不等式.

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle = \|x\|_2^2 \|y\|_2^2$$

Definition 3.1.2 — ℓ_∞ -范数.

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|, x \in \mathbb{R}^n$$

Proof.

$$\begin{aligned}
\max_{1 \leq i \leq n} |x_i| &\leq (|x_1|^p + \cdots + |x_i|^p + \cdots + |x_n|^p)^{1/p} \\
&\leq \left(n \max_{1 \leq i \leq n} |x_i|^p \right)^{1/p} \\
&= n^{1/p} \max_{1 \leq i \leq n} |x_i| \\
&\rightarrow \max_{1 \leq i \leq n} |x_i| \quad (p \rightarrow \infty)
\end{aligned}$$

■

Definition 3.1.3 — ℓ_p -范数.

$$\|x\|_p = (x_1^p + x_2^p + \cdots + x_n^p)^{\frac{1}{p}}, \quad x \in \mathbb{R}^n, p \geq 1$$

ℓ_1 范数 $\|x\|_1$, ℓ_2 -范数 $\|x\|_2$, ℓ_∞ -范数是 ℓ_p -范数的特例.

证明可以使用以下两条不等式

Theorem 3.1.2 — Minkowshi Inequality.

$$\left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}}, p \geq 1, x, y \in \mathbb{R}^n$$

Theorem 3.1.3 — Hölder Inequality.

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \left(\sum_{i=1}^n |y_i|^q \right)^{1/q}, \frac{1}{p} + \frac{1}{q} = 1, 1 < p, q < \infty$$

3.2 Root Mean Square Value (RMS)

Definition 3.2.1 — 向量 x 的均方值 (mean-square value). 向量 $x \in \mathbb{R}^n$ 的均方值 (mean-square value)

$$\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n} = \frac{\|x\|_2^2}{n}$$

Definition 3.2.2 — n 维向量 x 的均方根 (root-mean-square value, RMS).

$$\text{rms}(x) = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}} = \frac{\|x\|_2}{\sqrt{n}}$$

$\text{rms}(x)$ 给出了 $|x_i|$ 的“典型” (typical) 值. 例如, $\text{rms}(\mathbf{1}) = 1$ (与 n 无关). 均方根 (RMS) 值对于比较不同长度的向量大小是比较有用的.

3.3 Chebyshev's Inequality

Theorem 3.3.1 — Chebyshev's Inequality.

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

$$P(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

Theorem 3.3.2 — Chebyshev's Inequality. 假设 k 为向量 x 分量满足条件 $|x_i| \geq a$ 的个数, 即 $x_i^2 \geq a^2$ 的个数.

因此: $\|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_n^2 \geq ka^2$

将 a^2 移项, 可得到 $k \leq \frac{\|x\|_2^2}{a^2}$

满足 $|x_i| \geq a$ 的 x_i 数量不会超过 $\frac{\|x\|_2^2}{a^2}$

Corollary 3.3.3 — Chebyshev's Inequality Using RMS.

$$\text{rms}(x) = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}} = \frac{\|x\|_2}{\sqrt{n}}$$

$|x_i| \geq a$ 的项数占整体的比例不会超过 $\left(\frac{\text{rms}(x)}{a}\right)^2$, 即 $\frac{k}{n} \leq \left(\frac{\text{rms}(x)}{a}\right)^2$

3.4 Distance

Definition 3.4.1 — Euclidean distance. n 维向量 a 和 b 之间的欧氏距离

$$\text{dist}(a, b) = \|a - b\|_2$$

Definition 3.4.2 — RMS deviation. $\text{rms}(a - b)$ 是 a 和 b 之间的均方根偏差.

Theorem 3.4.1 — Triangular Inequality.

$$\|a - c\|_2 = \|(a - b) + (b - c)\|_2 \leq \|a - b\|_2 + \|b - c\|_2$$

3.4.1 Feature Distance and Nearest Neighbor

Definition 3.4.3 — Feature Distance. 如果 x 和 y 分别为两个实体的特征向量, 那么它们的特征距离 (feature distance) 为 $\|x - y\|_2$

Definition 3.4.4 — Nearest Neighbor. 给定向量 x , 一个组向量 Z_1, \dots, Z_m , 当 \hat{q}_j 满足:

$$\|x - z_j\|_2 \leq \|x - z_i\|_2, \quad i = 1, \dots, m$$

则称 z_j 是 x 的最近邻 (nearest neighbor)

3.5 Standard Deviation

Definition 3.5.1 — 算术平均值. 对于 n 维向量 x

$$\text{avg}(x) = \frac{\mathbf{1}^T x}{n}$$

Definition 3.5.2 — De-meaned Vector.

$$\tilde{x} = x - \text{avg}(x)\mathbf{1}$$

因此 $\text{avg } \mathbf{g}(\tilde{x}) = 0$

Definition 3.5.3 — x 的标准差.

$$\text{std}(x) = \text{rms}(\tilde{x}) = \frac{\|x - (1^T x/n) \mathbf{1}\|_2}{\sqrt{n}}$$

$\text{std}(x)$ 表示数据元素的变化程度. 对于常数 α , 当且仅当 $x = \alpha\mathbf{1}$ 时, $\text{std}(x) = 0$.

Theorem 3.5.1

$$\text{rms}(x)^2 = \text{avg}(x)^2 + \text{std}(x)^2$$

3.6 Angle

Definition 3.6.1 — 两个非零向量 a 和 b 之间的角 (angle).

$$\angle(a, b) = \arccos\left(\frac{a^T b}{\|a\|_2 \|b\|_2}\right)$$

$\angle(a, b)$ 的取值范围为 $[0, \pi]$, 且满足

$$a^T b = \|a\|_2 \|b\|_2 \cos(\angle(a, b))$$

在二维和三维向量之中, 这里的角与普通角度 (ordinary angle) 是一致的.

- $\theta = \frac{\pi}{2} = 90$: a 和 b 为正交, 写作 $a \perp b$ ($a^T b = 0$).
- $\theta = 0$: a 和 b 为同向的 ($a^T b = \|a\| \|b\|$).
- $\theta = \pi = 180$: a 和 b 为反向的 ($a^T b = -\|a\| \|b\|$).
- $\theta < \frac{\pi}{2} = 90$: a 和 b 成锐角 ($a^T b > 0$).

- $\theta > \frac{\pi}{2} = 90$: a 和 b 成钝角 ($a^T b < 0$).

Definition 3.6.2 — 球面的距离.

$$R\angle(a, b)$$

3.6.1 相关系数

给定向量 a 和 b , 其去均值向量为:

$$\tilde{a} = a - \text{avg}(a)\mathbf{1}, \tilde{b} = b - \text{avg}(b)\mathbf{1}$$

Definition 3.6.3 — a 和 b 的相关系数.

$$\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\|_2 \|\tilde{b}\|_2} = \cos \angle(\tilde{a}, \tilde{b})$$

where $\tilde{a} \neq 0, \tilde{b} \neq 0$.

■ **Example 3.3** 高度相关的向量:

- 邻近地区的降雨时间序列.
- 类型密切相关文档的单词计数向量.
- 同行业中类似公司的日收益.

比较不相关的向量:

- 无关的向量.
- 音频信号 (比如, 在多轨录音中的不同轨).

负相关的向量:

- 深圳与墨尔本的每天气温变化

■

4. 优化问题初步

4.1 优化问题引入

Problem 4.1 假设 N 个样本向量 $x_1, \dots, x_N \in \mathbb{R}^n$, 需要找到中心向量 z 满足

$$\min_{z \in \mathbb{R}^n} \sum_{i=1}^N \|x_i - z\|_2^2$$

Definition 4.1.1 — 高阶无穷小记号 o . 设 x, y 是同一变化过程中的无穷小, 即 $x \rightarrow 0, y \rightarrow 0$, 如果它们极限

$$\lim \frac{y}{x} = 0$$

则称 y 是 x 的高阶无穷小, 记作 $y = o(x)$.

Corollary 4.1.1

$$\lim \frac{y}{Cx} = \frac{1}{C} \lim \frac{y}{x} = 0$$

也即则称 y 是 Cx 的高阶无穷小, 记作 $y = o(Cx)$.

Proposition 4.1.2 — 优化求解的必要条件. 假设函数 f 在 \hat{x} 可微, 则有

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} f(x) \Rightarrow \nabla f(\hat{x}) = 0$$

Proof. 假设函数 f 在 \hat{x} 一阶泰勒展开, 有

$$f(x) = f(\hat{x}) + \langle \nabla f(\hat{x}), x - \hat{x} \rangle + o(\|x - \hat{x}\|_2)$$

假设 $\delta f(\hat{x}) \neq 0$, 则令 $\tilde{x} = \hat{x} - t \nabla f(\hat{x}), t > 0$, 可得

$$f(\tilde{x}) = f(\hat{x}) - t \|\nabla f(\hat{x})\|_2^2 + o(t \|\nabla f(\hat{x})\|_2)$$

当 $t \rightarrow 0$ 则 $t\|\nabla f(\hat{x})\|_2 \rightarrow 0$, 高阶无穷小 $o'(t\|\nabla f(\hat{x})\|_2) \rightarrow 0$
 当 t 足够小时, 存在 $t\|\nabla f(\hat{x})\|_2 \geq o(t\|\nabla f(\hat{x})\|_2)$, 即

$$-t\|\nabla f(\hat{x})\|_2^2 + o(t\|\nabla f(\hat{x})\|_2) \leq 0$$

$$f(\tilde{x}) = f(\hat{x}) - t\|\nabla f(\hat{x})\|_2^2 + o(t\|\nabla f(\hat{x})\|_2) \leq f(\hat{x})$$

与 $\hat{x} = \arg \min_{\mathbf{R}^n} f(x)$ 矛盾.

$\nabla f(\hat{x}) = 0$, 是最优问题解的必要条件. 通常 $\nabla f(\hat{x}) = 0 \Leftrightarrow \hat{x} = \arg \min_{\mathbf{R}^n} f(x)$. ■

■ **Example 4.1**

$$f(x) = -x^2, \quad x \in \mathbf{R}, \hat{x} = \operatorname{argmin}_{\mathbf{R}} f(x)$$

$\nabla f(\hat{x}) = 0$, 则有 $-2\hat{x} = 0$, 即 $\hat{x} = 0$

$$f(\hat{x}) = 0 \geq f(x), \quad x \in \mathbf{R}$$

(最大值!) ■

4.2 Convex Set

Definition 4.2.1 — 凸集. $\forall x, y \in \Omega, \alpha \in \mathbf{R}, 0 \leq \alpha \leq 1$ 有

$$\alpha x + (1 - \alpha)y \in \Omega$$

则定义域 $\Omega \in \mathbf{R}^n$ 称为凸的 (Convex) 集合

(域内两点连线之间都属于这个域)

Definition 4.2.2 — 凸函数. 设函数 $f(x)$ 定义于称为凸的定义域 $\Omega \in \mathbf{R}^n$ 满足

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y), \forall x, y \in \Omega, \alpha \in \mathbf{R}, 0 \leq \alpha \leq 1$$

称其为凸函数.

■ **Example 4.2**

$$f(x) = x^2, x \in \mathbf{R}$$

$$\begin{aligned} f(\alpha x + (1 - \alpha)y) &= (\alpha x + (1 - \alpha)y)^2 \\ &= \alpha^2 x^2 + 2\alpha(1 - \alpha)xy + (1 - \alpha)^2 y^2 \\ &= \alpha x^2 + (1 - \alpha)y^2 + (\alpha^2 - \alpha)x^2 + (\alpha^2 - \alpha)y^2 + 2\alpha(1 - \alpha)xy \\ &= \alpha x^2 + (1 - \alpha)y^2 - \alpha(1 - \alpha)(x - y)^2 \\ &\leq \alpha x^2 + (1 - \alpha)y^2 = \alpha f(x) + (1 - \alpha)f(y) \end{aligned}$$

■ **Example 4.3** $f(x) = \|x\|$, 其中 $\| \cdot \|$ 表示 \mathbf{R}^n 上的向量范数, $x \in \mathbf{R}^n$. ■

Proof.

$$\|\alpha x + (1 - \alpha)y\| \leq \|\alpha x\| + \|(1 - \alpha)y\| = |\alpha|\|x\| + |1 - \alpha|\|y\|$$

■

■ Example 4.4

$$f(x) = \|x\|_2^2, x \in \mathbb{R}^n$$

■

Theorem 4.2.1 — 可微函数 f 是凸函数的充要条件.

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y$$

Proof. 首先, 证明一维情况 $f: \mathbb{R} \rightarrow \mathbb{R}, \alpha \in [0, 1]$.

\Rightarrow 充分条件: $f(\alpha x + (1 - \alpha)y) = f(x + (1 - \alpha)(y - x)) \leq \alpha f(x) + (1 - \alpha)f(y)$, 有

$$f(y) \geq f(x) + \frac{f(x + (1 - \alpha)(y - x)) - f(x)}{(1 - \alpha)(y - x)}(y - x)$$

令 $\alpha \rightarrow 1^-$, 则有 $f(y) \geq f(x) + f'(x)(y - x)$.

\Leftarrow 必要条件: 令 $y \neq x, z = \alpha x + (1 - \alpha)y$ 则有

$$f(x) \geq f(z) + f'(z)(x - z), f(y) \geq f(z) + f'(z)(y - z)$$

可得

$$\begin{aligned} \alpha f(x) + (1 - \alpha)f(y) &\geq f(z) + \alpha f'(z)(x - z) + (1 - \alpha)f'(z)(y - z) \\ &= f(z) + f'(z)(\alpha x + (1 - \alpha)y - z) \\ &= f(z) \end{aligned}$$

证明 n 维情况 $f: \mathbb{R}^n \rightarrow \mathbb{R}$.

\Rightarrow 充分条件: 令 $g(t) = f(tx + (1 - t)y), t \in \mathbb{R}$, 则 $g'(t) = \langle \nabla f(tx + (1 - t)y), x - y \rangle$

由于 f 是凸函数, 证明 $g(t)$ 也是凸函数; 并可得 $g(0) \geq g(1) + g'(1)(-1)$, 得证.

\Leftarrow 必要条件: 与一维类似. (将 f' 改为 $\nabla f(z)^T$)

■

Theorem 4.2.2 如果可微函数 f 是凸函数, 则有

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} f(x) \Leftrightarrow \nabla f(\hat{x}) = 0$$

Proof. 已证 $\hat{x} = \arg \min_{x \in \mathbb{R}^n} f(x) \Rightarrow$ 可得 $\nabla f(\hat{x}) = 0$

只需证 $\nabla f(\hat{x}) = 0 \Rightarrow \hat{x} = \arg \min_{x \in \mathbb{R}^n} f(x)$.

由于函数 f 是可微凸的, 则有 $\forall x \in \mathbb{R}^n$,

$$\begin{aligned} f(x) &\geq f(\hat{x}) + \langle \nabla f(\hat{x}), x - \hat{x} \rangle \\ &\geq f(\hat{x}) + \langle 0, x - \hat{x} \rangle \geq f(\hat{x}) \end{aligned}$$

可得 $f(x) \geq f(\hat{x}), \hat{x} = \arg \min_{x \in \mathbb{R}^n} f(x)$.

■

4.3 向量偏导

Definition 4.3.1 — 向量对向量的导数.

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

$$\nabla f(z) = \begin{bmatrix} \frac{\partial f(z)}{\partial z_1} \\ \vdots \\ \frac{\partial f(z)}{\partial z_n} \end{bmatrix}$$

■ **Example 4.5**

$$f(z) = x^T z + z^T z = \sum_{i=1}^n \{x_i z_i + z_i^2\}$$

$$\nabla f(z) = \begin{bmatrix} \frac{\partial f(z)}{\partial z_1} \\ \vdots \\ \frac{\partial f(z)}{\partial z_n} \end{bmatrix} = \begin{bmatrix} x_1 + 2z_1 \\ \vdots \\ x_n + 2z_n \end{bmatrix} = x + 2z$$

问题4.1中已知目标函数是凸函数。(见4.2, 4.3, 4.2)
则可以求解

$$f(z) = \sum_{i=1}^N \|x_i - z\|_2^2 = \sum_{i=1}^N \langle x_i - z, x_i - z \rangle = \sum_{i=1}^N \{x_i^T x_i - 2x_i^T z + z_i^T z\}$$

利用等价条件4.2.1

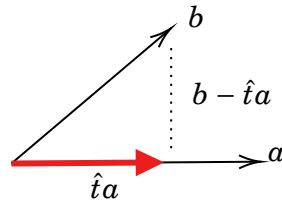
$$\nabla f(z) = \sum_{i=1}^N \{-2x_i + 2z\} = 0$$

(求导 4.3.1)

$$z = \frac{1}{N} \sum_{i=1}^N x_i$$

4.4 投影问题: 标量优化问题

Figure 4.1: Projection onto a line



Problem 4.2 假设 $a, b \in \mathbb{R}^n, a \neq 0, t \in \mathbb{R}$, 当 t 多大时, ta 到 b 之间的距离最小

$$\hat{t} = \min_t \|ta - b\|_2^2$$

定义

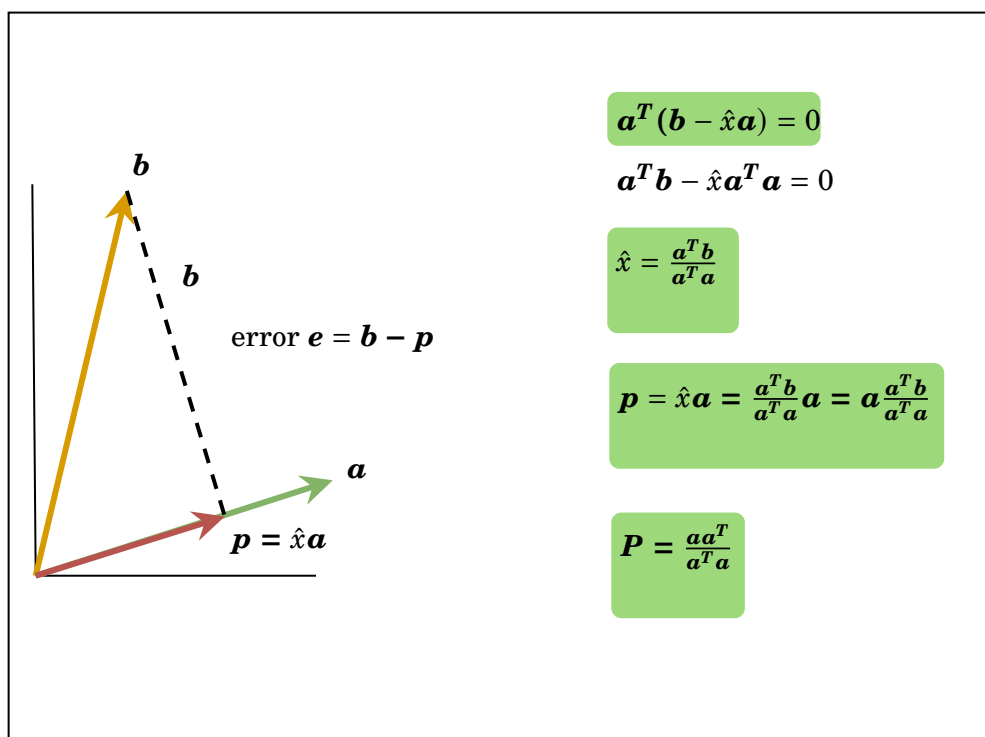
$$f(t) = \|ta - b\|_2^2 = \langle ta - b, ta - b \rangle = t^2 a^T a - 2ta^T b + b^T b$$

可以验证 $f(t)$ 满足凸函数的定义。

$$\nabla f(t) = 2ta^T a - 2a^T b = 0$$

$$\hat{t} = \frac{a^T b}{a^T a} = \frac{a^T b}{\|a\|_2^2}$$

Figure 4.2: Projection onto a line from the perspective of linear algebra.



4.5 Clustering

将物理或抽象对象的集合分成由类似特征组成的多个类的过程称为聚类 (*clustering*).

目标：分成 k 个集合, 尽量使得同一个集合中的向量彼此接近.

Notation 4.1. 给定 N 个 n 维向量 $x_1, \dots, x_N \in \mathbb{R}^n$

- 标签 $c_i \in \{1, 2, \dots, k\}$ 表示向量 x_i 所属类别, 例如 $c_i = 2$ 表示 x_i 属于第 2 类.
- 对于 $j = 1, \dots, k$, $G_j = \{i : c_i = j\}$ 表示属于第 j 类的向量 x_i 的下标集合.
- 向量 $z_j, j = 1, \dots, k$, 表示同属于 j 类的向量 $x_i, i \in G_j$ 的聚类中心.

聚类目标是找到向量 x_i 的“标签 c_i ”和“聚类中心 z_j ”

Problem 4.3

$$\min_{z_j} \sum_{i \in G_j} \|x_i - z_j\|_2^2, j = 1, \dots, k$$

$$c_i = \operatorname{argmin}_{j \in \{1, \dots, k\}} \|x_i - z_j\|_2^2, i = 1, 2, \dots, N$$

k -means 算法是将 N 向量 $x_i \in \mathbb{R}^n$ 划分成 k 类的迭代聚类算法.

Algorithm 1: k -means Algorithm

- 1 在 N 个点中随机选取 k 个点, 分别作为聚类中心 z_j
- 2 更新聚类标签 c_i : 计算每个点 x_i 到 k 个聚类中心 z_j 的距离, 并将其分配到最近的聚类中心 z_j 所在的聚类中 $c_i = j$
- 3 更新聚类中心 z_j : 重新计算每个聚类现在的质心, 并以其作为新的聚类中心, 根据更新标签 c_i , 更新属于第 j 类下标集合 $G_j = \{i : c_i = j\}$, 重新计算 c_i 类的聚类中心 z_j
- 4 重复步骤 2、3, 直到所有聚类中心不再变化

Proof. 更新聚类标签 c_i :

$$\|x_i - z_j\|_2^2 = \operatorname{argmin} \{ \|x_i - z_1\|_2^2, \|x_i - z_2\|_2^2, \dots, \|x_i - z_k\|_2^2 \}$$

更新聚类中心 z_j :

$$\nabla f_j(z_j) = \sum_{i \in G_j} 2(x_i - z_j) = 0$$

$$z_j = \frac{1}{|G_j|} \sum_{i \in G_j} x_i$$

$|G_j|$ 表示集合 G_j 中元素的数目. ■

在每一次迭代中目标函数 J 都会下降, 直到聚类中心 z_1, \dots, z_k 和划分聚类标签集合 G_1, \dots, G_k 不再变化.

但是 k -means 算法依赖于初始随机生成的聚类中心, 只可得到目标函数 J 的局部局部最优.

解决方案: 使用不同的 (随机的) 初始聚类中心运行 k -means 算法若干次, 取目标函数 J 值最小的一次作为最终的聚类结果.

5. Linear Independence

5.1 线性相关、线性无关

Definition 5.1.1 — 线性相关 (linearly dependent). 定义：对于向量 $a_1, \dots, a_m \in \mathbb{R}^n$, 如果存在不全为零的数 $\beta_1, \dots, \beta_m \in \mathbb{R}$, 使得

$$\beta_1 a_1 + \dots + \beta_m a_m = 0$$

则称向量 a_1, \dots, a_m 是线性相关 (linearly dependent).

线性相关等价于至少有一个向量 a_i 是其它向量的线性组合.

Corollary 5.1.1 向量集 $\{a_1\}$ 是线性相关的, 当且仅当 $a_1 = 0$.

Corollary 5.1.2 向量集 $\{a_1, a_2\}$ 是线性相关的, 当且仅当其中一个 $a_1 = \beta a_2, \beta \neq 0$.

Definition 5.1.2 — 线性独立 (linearly independent). 如果 n 维向量集 $\{a_1, \dots, a_m\}$ 不是线性相关的, 即线性独立 (linearly independent), 也称线性无关, 即:

$$\beta_1 a_1 + \dots + \beta_m a_m = 0$$

当且仅当 $\beta_1 = \dots = \beta_m = 0$, 上述等式成立.

线性无关等价于不存在一个向量 a_i 是其它向量的线性组合.

Corollary 5.1.3 注：一个 n 维向量集最多有 n 个线性无关的向量, 也就是说如果 n 维向量集有 $n + 1$ 个向量, 那它们必线性相关

■ **Example 5.1** n 维单位向量 e_1, \dots, e_n 是线性独立的. ■

■ **Example 5.2**

$$a_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \quad a_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\beta_1 a_1 + \beta_2 a_2 + \beta_3 a_3 = \begin{bmatrix} \beta_1 - \beta_2 \\ -2\beta_1 + \beta_3 \\ \beta_2 + \beta_3 \end{bmatrix} = 0$$

$$\beta_1 = \beta_2 = \beta_3 = 0$$

■

Theorem 5.1.4 假设 x 是线性无关向量 a_1, \dots, a_k 的线性组合:

$$x = \beta_1 a_1 + \dots + \beta_k a_k$$

则其系数 β_1, \dots, β_k 是唯一的, 即如果有:

$$x = \gamma_1 a_1 + \dots + \gamma_k a_k$$

则对于 $i = 1, \dots, k$, 有 $\beta_i = \gamma_i$.

Proof. 系数是唯一的原因:

$$(\beta_1 - \gamma_1) a_1 + \dots + (\beta_k - \gamma_k) a_k = x - x = 0$$

由于向量 a_1, \dots, a_k 线性无关, 有 $\beta_1 - \gamma_1 = \beta_k - \gamma_k = 0$.

■

5.2 Basis

■ **Definition 5.2.1** — 基 (Basis). n 个线性独立的 n 维向量 a_1, \dots, a_n 的集合

Definition 5.2.2 — 向量 b 在基底 a_1, \dots, a_n 下的分解. 任何一个 n 维向量 b 都可以用它们的线性组合来表示

$$b = \beta_1 a_1 + \dots + \beta_n a_n$$

Proof. 同一向量的系数是唯一的.

■

■ **Example 5.3** e_1, \dots, e_n 是一组基, 那么 b 在此基底下的分解为

$$b = b_1 e_1 + \dots + b_n e_n, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^n$$

■

5.3 标准正交向量

■ **Definition 5.3.1** — Orthogonal Vectors. 在 n 维向量集 a_1, \dots, a_k 中, 如果对于 $i \neq j$, 都有 $a_i \perp a_j$, 则称它们相互正交 (orthogonal).

Definition 5.3.2 — Orthonormal Vectors. 如果 n 维向量集 a_1, \dots, a_k 相互正交, 且每个向量的模长都为单位长度 1, 即对于 $i = 1, \dots, k$, 有 $\|a_i\|_2^2 = 1$, 则称它们是标准正交 (orthonormal) 的.

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Corollary 5.3.1 标准正交的向量集是线性无关的.

Corollary 5.3.2 根据线性无关的性质, 必有向量集向量个数 $k \leq n$.

Proof. The proof is by induction on the dimension n .

First consider a linearly independent collection a_1, \dots, a_k of 1-vectors. We must have $a_1 \neq 0$. This means that every element a_i of the collection can be expressed as a multiple $a_i = (a_i/a_1) a_1$ of the first element a_1 . This contradicts linear independence unless $k = 1$.

Next suppose $n \geq 2$ and the independence-dimension inequality holds for dimension $n - 1$.

Let a_1, \dots, a_k be a linearly independent list of n -vectors. We need to show that $k \leq n$. We partition the vectors as

$$a_i = \begin{bmatrix} b_i \\ \alpha_i \end{bmatrix}, \quad i = 1, \dots, k$$

where b_i is an $(n - 1)$ -vector and α_i is a scalar.

First suppose that $\alpha_1 = \dots = \alpha_k = 0$. Then the vectors b_1, \dots, b_k are linearly independent: $\sum_{i=1}^k \beta_i b_i = 0$ holds if and only if $\sum_{i=1}^k \beta_i a_i = 0$, which is only possible for $\beta_1 = \dots = \beta_k = 0$ because the vectors a_i are linearly independent. The vectors b_1, \dots, b_k therefore form a linearly independent collection of $(n - 1)$ -vectors. By the induction hypothesis we have $k \leq n - 1$, so certainly $k \leq n$.

Next suppose that the scalars α_i are not all zero. Assume $\alpha_j \neq 0$. We define a collection of $k - 1$ vectors c_i of length $n - 1$ as follows:

$$c_i = \begin{cases} b_i - \frac{\alpha_i}{\alpha_j} b_j, & i = 1, \dots, j - 1 \\ b_{i+1} - \frac{\alpha_{i+1}}{\alpha_j} b_j, & i = j, \dots, k - 1 \end{cases}$$

These $k - 1$ vectors are linearly independent: If $\sum_{i=1}^{k-1} \beta_i c_i = 0$ then

$$\sum_{i=1}^{j-1} \beta_i \begin{bmatrix} b_i \\ \alpha_i \end{bmatrix} + \gamma \begin{bmatrix} b_j \\ \alpha_j \end{bmatrix} + \sum_{i=j+1}^k \beta_{i-1} \begin{bmatrix} b_i \\ \alpha_i \end{bmatrix} = 0 \quad (5.1)$$

with

$$\gamma = -\frac{1}{\alpha_j} \left(\sum_{i=1}^{j-1} \beta_i \alpha_i + \sum_{i=j+1}^k \beta_{i-1} \alpha_i \right)$$

Since the vectors $a_i = (b_i, \alpha_i)$ are linearly independent, the eq. (5.1) only holds when all the coefficients β_i and γ are all zero. This in turns implies that the vectors c_1, \dots, c_{k-1} are linearly independent. By the induction hypothesis $k - 1 \leq n - 1$ so we have established that $k \leq n$ ■

Definition 5.3.3 — n 维向量的一个标准正交基. 当 $k = n$ 时, a_1, \dots, a_n 是 n 维向量的一个标准正交基.

Definition 5.3.4 — x 在标准正交基下的标准正交分解. 如果 a_1, \dots, a_n 是一个标准正交基, 对于任意维向量 x ;

$$x = \begin{pmatrix} a_1^T x \end{pmatrix} a_1 + \dots + \begin{pmatrix} a_n^T x \end{pmatrix} a_n$$

则称其为 x 在标准正交基下的标准正交分解.

这个分解可以用于计算不同标准正交基下的系数.

Proof. 由于正交向量的性质

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

所以

$$a_i^T x = \begin{pmatrix} a_1^T x \end{pmatrix} a_i^T a_1 + \dots + \begin{pmatrix} a_i^T x \end{pmatrix} a_i^T a_i + \dots + \begin{pmatrix} a_n^T x \end{pmatrix} a_i^T a_n = a_i^T x$$

■

5.4 Gram-Schmidt Algorithm

Algorithm 2: Gram-Schmidt Algorithm

Input: n 维向量 a_1, \dots, a_k

Output: 若这些向量线性无关时 q_1, \dots, q_k (标准正交基); 若线性相关时判断 a_j 是 a_1, \dots, a_{j-1} 的线性组合

```

1  $q_1 = a_1 / \|a_1\|_2$ 
2 while  $i = 2, \dots, k$  do
3   正交化:  $\tilde{q}_i = a_i - (q_1^T a_i) q_1 - \dots - (q_{i-1}^T a_i) q_{i-1}$ 
4   检验线性相关: 如果  $\tilde{q}_i = 0$ , 提前退出迭代
5   单位化:  $q_i = \tilde{q}_i / \|\tilde{q}_i\|_2$ 
6 end
```

如果步骤 2 中未提前结束迭代, 那么 a_1, \dots, a_k 是线性独立的, 而且 q_1, \dots, q_k 是标准正交基.

如果在第 j 次迭代中提前结束, 说明 a_j 是 a_1, \dots, a_{j-1} 的线性组合, 因此 a_1, \dots, a_k 是线性相关的.

Theorem 5.4.1 q_1, \dots, q_{i-1}, q_i 是标准正交的.

Proof. 假设第 $i-1$ 次迭代成立, 即: $q_r \perp q_s, \forall r, s < i$.

正交化步骤保证有以下关系成立

$$\tilde{q}_i = a_i - \begin{pmatrix} q_1^T a_i \end{pmatrix} q_1 - \dots - \begin{pmatrix} q_{i-1}^T a_i \end{pmatrix} q_{i-1}$$

等式两边同时乘以 $q_j^T, j = 1, \dots, i-1$

$$\begin{aligned} q_j^T \tilde{q}_i &= q_j^T a_i - \begin{pmatrix} q_1^T a_i \end{pmatrix} \begin{pmatrix} q_j^T q_1 \end{pmatrix} - \dots - \begin{pmatrix} q_{i-1}^T a_i \end{pmatrix} \begin{pmatrix} q_j^T q_{i-1} \end{pmatrix} \\ &= q_j^T a_i - q_j^T a_i = 0 \end{aligned}$$

$$\because q_j^T q_r = 0, j \neq r, q_j^T q_j = 1$$

$$\therefore \tilde{q}_i \perp q_1, \dots, \tilde{q}_i \perp q_{i-1}.$$

单位化步骤保证了 $q_i = \tilde{q}_i / \|\tilde{q}_i\|_2$, 即 q_1, \dots, q_i 是标准正交. ■

Algorithm 3: Gram-Schmidt Algorithm for Three Vectors

Input: Three independent vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$

Output: Three orthonormal vectors $\mathbf{q}_1 = \mathbf{A} / \|\mathbf{A}\|$, $\mathbf{q}_2 = \mathbf{B} / \|\mathbf{B}\|$, $\mathbf{q}_3 = \mathbf{C} / \|\mathbf{C}\|$.

1 Choose $\mathbf{A} = \mathbf{a}$

2

$$\mathbf{B} = \mathbf{b} - \frac{\mathbf{A}^T \mathbf{b}}{\mathbf{A}^T \mathbf{A}} \mathbf{A}$$

3

$$\mathbf{C} = \mathbf{c} - \frac{\mathbf{A}^T \mathbf{c}}{\mathbf{A}^T \mathbf{A}} \mathbf{A} - \frac{\mathbf{B}^T \mathbf{c}}{\mathbf{B}^T \mathbf{B}} \mathbf{B}$$

4 单位化

5.4.1 The Analysis of Gram-Schmidt Algorithm

假设 Gram-Schmidt 正交法未在第 i 次迭代提前终止:

Corollary 5.4.2 a_i 是 q_1, \dots, q_i 的一个线性组合.

$$a_i = \|\tilde{q}_i\|_2 q_i + (q_1^T a_i) q_1 + \dots + (q_{i-1}^T a_i) q_{i-1}$$

Proof.

$$\tilde{q}_i = a_i - (q_1^T a_i) q_1 - \dots - (q_{i-1}^T a_i) q_{i-1}$$

$$a_i = \tilde{q}_i + (q_1^T a_i) q_1 + \dots + (q_{i-1}^T a_i) q_{i-1}$$

注意有性质: $q_i = \tilde{q}_i / \|\tilde{q}_i\|_2$.

$$a_i = \|\tilde{q}_i\|_2 q_i + (q_1^T a_i) q_1 + \dots + (q_{i-1}^T a_i) q_{i-1}$$

■

则有

Corollary 5.4.3

$$q_i = \frac{a_i - (q_1^T a_i) q_1 - \dots - (q_{i-1}^T a_i) q_{i-1}}{\|\tilde{q}_i\|_2}$$

Corollary 5.4.4 q_i 是 a_1, \dots, a_i 的一个线性组合.

Proof. 归纳假设, 每个 q_{i-1} 都是 a_1, \dots, a_{i-1} 的线性组合:

$$\begin{aligned} q_2 &= \frac{a_2 - (q_1^T a_2) q_1}{\|\tilde{q}_2\|_2} \\ &= \frac{a_2 - (q_1^T a_2) \frac{a_1}{\|a_1\|_2}}{\|\tilde{q}_2\|_2} \\ q_3 &= \frac{a_3 - (q_1^T a_3) q_1 - (q_2^T a_3) q_2}{\|\tilde{q}_3\|_2} \end{aligned}$$

通过对 i 的归纳证明, 可得 q_i 是 a_1, \dots, a_i 的线性组合. ■

假设 Schmidt 正交法在第 j 次迭代提前终止:

Corollary 5.4.5 a_j 是 q_1, \dots, q_{j-1} 的一个线性组合.

$$a_j = (q_1^T a_j) q_1 + \dots + (q_{j-1}^T a_j) q_{j-1}$$

Proof.

$$\begin{aligned} \tilde{q}_i &= a_i - (q_1^T a_i) q_1 - \dots - (q_{i-1}^T a_i) q_{i-1} \\ 0 &= a_i - (q_1^T a_i) q_1 - \dots - (q_{i-1}^T a_i) q_{i-1} \\ a_i &= \|\tilde{q}_i\|_2 q_i + (q_1^T a_i) q_1 + \dots + (q_{i-1}^T a_i) q_{i-1} \end{aligned}$$

■

Corollary 5.4.6 a_j 是 a_1, \dots, a_{j-1} 的线性组合.

Proof. 每一个 q_1, \dots, q_{j-1} 都是 a_1, \dots, a_{j-1} 的线性组合.

因此 a_j 是 a_1, \dots, a_{j-1} 的线性组合. ■

Matrices

6	Matrices	37
6.1	Matrices	
6.2	矩阵运算	
6.3	Special Matrices and Matrices in Different Applications	
6.4	Gram 矩阵	
7	Matrices Norms	49
7.1	矩阵范数	
8	适定问题	53
8.1	The Definition of Well-posed Problem	
8.2	绝对误差的界限	
8.3	相对误差的界限	
9	Inverse of Matrices	55
9.1	Left Inverse, Right Inverse, Inverse	
9.2	Linear Equation Systems	
9.3	Fundamental Theorem of Linear Algebra	
9.4	Invertible Matrices	
9.5	转置和共轭转置的逆	
9.6	Gram Matrix 非奇异的性质	
9.7	伪逆	
10	Orthogonal Matrices	64
10.1	预备知识	
10.2	正交矩阵	
10.3	Permutation Matrices	
10.4	平面旋转	
10.5	Householder Matrix	
10.6	正交矩阵乘积	
10.7	具有正交矩阵的线性方程	
10.8	列标准正交的高矩阵	
10.9	值域范围、列空间	
11	QR 分解与 Householder 变换	72
11.1	Triangular Matrices	
11.2	QR Factorization	
11.3	QR 分解的应用	
11.4	QR Algorithm Using Gram-Schmidt Algorithm	
11.5	The Numerical Instability of QR Decomposition based on Gram-Schmidt Algorithm	
11.6	QR Decomposition Using Householder Transformation	
11.7	Householder 变换进行 QR 分解的 Q 因子	
11.8	Fast Orthogonalization	

6. Matrices

6.1 Matrices

Definition 6.1.1 — 矩阵. 矩阵是一个由数字构成的矩阵数组.

$$\begin{bmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{bmatrix} \quad \left(\begin{array}{cccc} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{array} \right)$$

上述矩阵大小 (size) 为 3×4 , 矩阵的每一个元素 (element) 又称为系数 (coefficient);

Notation 6.1. 设 B_{ij} 表示矩阵 B 中第 i 行第 j 的元素

实数域中大小为 $m \times n$ 的矩阵集合写为 $\mathbb{R}^{m \times n}$

复数域中大小为 $m \times n$ 的矩阵集合写为 $\mathbb{C}^{m \times n}$

Definition 6.1.2 — 标量. 不区分一个 1×1 矩阵和一个标量.

Definition 6.1.3 — 向量. 不区分一个 $n \times 1$ 矩阵和一个向量.

Definition 6.1.4 — 行向量, 列向量. 一个 $1 \times n$ 矩阵被称为一个行向量.
一个 $n \times 1$ 矩阵被称为一个列向量.

Definition 6.1.5 — 高形, 宽形和方形矩阵. 一个大小为 $m \times n$ 的矩阵为:

- 高的, 如果 $m > n$
- 宽的, 如果 $m < n$
- 方的, 如果 $m = n$

Definition 6.1.6 — 分块矩阵. 分块矩阵的每一个元都是一个矩阵.

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

其中 B, C, D, E 都是矩阵 (被称为矩阵 A 的子矩阵).

分块矩阵位于同一行的子矩阵行维度必须相等, 位于同一列的子矩阵列维度必须相等.

Definition 6.1.7 — 矩阵的列向量表示. 矩阵 $A \in \mathbb{R}^{m \times n}$, 可通过其列向量 (m -vector) 进行表示, 假设其列向量为 $a_1, \dots, a_n \in \mathbb{R}^m$, 则有

$$A = [a_1 \ \cdots \ a_n]$$

Definition 6.1.8 — 矩阵的行向量表示. 矩阵 $A \in \mathbb{R}^{m \times n}$ 通过其行向量 b_1, \dots, b_m 进行表示

$$A = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, b_i^T \in \mathbb{R}^n, i = 1, \dots, m$$

6.2 矩阵运算

Definition 6.2.1 — 矩阵数乘. 设矩阵 $A \in \mathbb{R}^{m \times n}$

$$\beta A = \begin{bmatrix} \beta A_{11} & \beta A_{12} & \cdots & \beta A_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta A_{m1} & \beta A_{m2} & \cdots & \beta A_{mn} \end{bmatrix}, \beta \in \mathbb{R}$$

Definition 6.2.2 — 矩阵加法. 矩阵 $A, B \in \mathbb{R}^{m \times n}$ 的和为

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \cdots & A_{1n} + B_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} + B_{m1} & A_{m2} + B_{m2} & \cdots & A_{mn} + B_{mn} \end{bmatrix}$$

Definition 6.2.3 — Transpose. 矩阵 A 的转置表示为 A^T

若 $A \in \mathbb{R}^{m \times n}$, 则 $A^T \in \mathbb{R}^{n \times m}$, 其被定义为: $(A^T)_{ij} = A_{ji}, i = 1, \dots, n; j = 1, \dots, m$

转置将原矩阵的行向量转化为列向量.

Corollary 6.2.1 — 转置的性质. 有如下性质:

- $(A^T)^T = A$
- 对称矩阵满足 $A^T = A$
- $(\beta A)^T = \beta A^T, (A + B)^T = A^T + B^T$

Definition 6.2.4 — 共轭转置. 矩阵 A 的共轭转置表示为 A^H

若 $A \in \mathbb{C}^{m \times n}$, 则 $A^H \in \mathbb{C}^{n \times m}$, 其被定义为: $(A^H)_{ij} = \bar{A}_{ji}, i = 1, \dots, n; j = 1, \dots, m$;

设矩阵 $A \in \mathbb{C}^{m \times n}$, 则其共轭转置为一个 $n \times m$ 矩阵

$$A^H = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{21} & \cdots & \bar{A}_{m1} \\ \bar{A}_{12} & \bar{A}_{22} & \cdots & \bar{A}_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{A}_{1n} & \bar{A}_{2n} & \cdots & \bar{A}_{mn} \end{bmatrix}$$

Corollary 6.2.2 — 共轭转置的性质. 有如下性质:

- $(A^H)^H = A$
- Hermitian 矩阵满足 $A = A^H$
- $(\beta A)^H = \beta A^H, (A + B)^H = A^H + B^H$

Definition 6.2.5 — 矩阵乘法. 设矩阵 $A \in \mathbb{R}^{m \times p}, B \in \mathbb{R}^{p \times n}$, 那么矩阵 A 与 B 的乘积, 记作 $C = AB$, 矩阵 $C \in \mathbb{R}^{m \times n}$ 的第 i 行第 j 列元素 C_{ij}

$$C_{ij} = \sum_{k=1}^p A_{ik} B_{kj}$$

注意: 矩阵 A 的列大小必须等于 B 的行大小.

Corollary 6.2.3 — 矩阵乘法性质. 有如下性质:

- 结合律: $(AB)C = A(BC)$
- 分配律: $A(B + C) = AB + AC$
-

$$(AB)^T = B^T A^T, \quad (AB)^H = B^H A^H$$

- 一般情况下 $AB \neq BA$
- 对于方阵 A 有, $IA = AI = A$

Definition 6.2.6 — 分块矩阵乘法.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} W & Y \\ X & Z \end{bmatrix} = \begin{bmatrix} AW + BX & AY + BZ \\ CW + DX & CY + DZ \end{bmatrix}$$

Definition 6.2.7 — 矩阵-向量乘积 Ax . 矩阵 $A \in \mathbb{R}^{m \times n}$ 和一个向量 $x \in \mathbb{R}^n$ 的积为

$$Ax = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n \end{bmatrix}$$

Corollary 6.2.4 Ax 是矩阵 A 列向量的线性组合.

$$Ax = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 a_1 + \cdots + x_n a_n$$

可以引出 A 的 Row Picture 和 Column Picture 的概念。

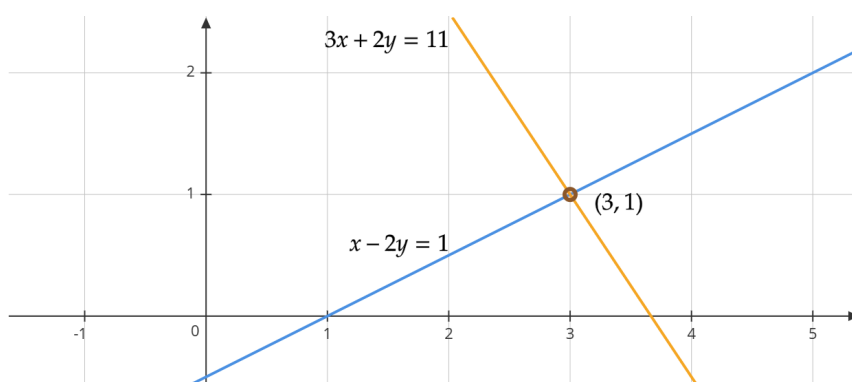
■ **Example 6.1**

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases}$$

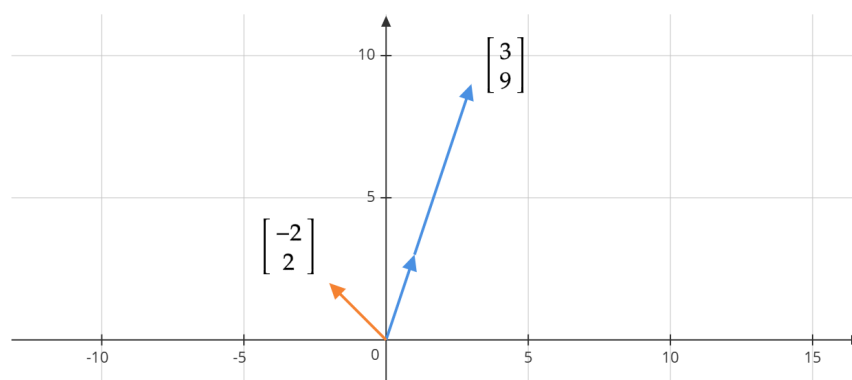
■

Figure 6.1: Row Picture and Column Picture for 6.1

(a) Row Picture



(b) Column Picture



Definition 6.2.8 — 矩阵-向量乘积函数 $f(x) = Ax$. 给定矩阵 $A \in \mathbb{R}^{m \times n}$, 定义函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m, f(x) = Ax$, 其中 $A = [f(e_1) \dots f(e_n)]$.

Proof. 该函数为一个线性函数:

$$A(\alpha x + \beta y) = \alpha(Ax) + \beta(Ay)$$

任意一个线性函数都可以写成矩阵-向量乘积函数的形式

$$\begin{aligned} f(x) &= f(x_1 e_1 + x_2 e_2 + \dots + x_n e_n) \\ &= x_1 f(e_1) + x_2 f(e_2) + \dots + x_n f(e_n) \\ &= \begin{bmatrix} f(e_1) & \dots & f(e_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

因此 $f(x) = Ax$, 其中 $A = [f(e_1) \dots f(e_n)]$ ■

6.2.1 Matrix Power

Definition 6.2.9 — **Matrix Power.** It makes sense to multiply a square matrix A by itself to form AA . We refer to this matrix as A^2 . Similarly, if k is a positive integer, then k copies of A multiplied together is denoted A^k . If k and l are positive integers, and A is square, then $A^k A^l = A^{k+l}$ and $(A^k)^l = A^{kl}$.

$$(A^{\ell+1})_{ij} = \sum_{k=1}^n A_{ik} (A^{\ell})_{kj}$$

By convention we take $A^0 = I$, which makes the formulas above hold for all nonnegative integer values of k and l .

■ **Example 6.2** — Paths in a directed graph.

$$A_{ij} = \begin{cases} 1 & \text{there is a edge from vertex } j \text{ to vertex } i \\ 0 & \text{otherwise} \end{cases}$$

■ **Example 6.3** — Linear dynamical system.

$$x_{t+\ell} = A^{\ell} x_t$$

6.2.2 矩阵乘法的算法复杂度

一般矩阵的乘法

$C = AB$ with A of size $m \times p$ and B of size $p \times n$

The product matrix C has size $m \times n$, so there are mn elements to compute.

The i, j element of C is the inner product of row i of A with column j of B . This is an inner product of vectors of length p and requires $2p - 1$ flops. Therefore the total is $mn(2p - 1)$ flops, which we approximate as $2mnp$ flops. $O(mnp)$

稀疏矩阵的乘法

Suppose that A is $m \times p$ and sparse, and B is $p \times n$, but not necessarily sparse.

The inner product of the i th row a_i^T of A with the j th column of B requires no more than $2 \text{nnz}(a_i^T)$ flops.

Summing over $i = 1, \dots, m$ and $j = 1, \dots, n$ we get $2 \text{nnz}(A)n$ flops.

If B is sparse, the total number of flops is no more than $2 \text{nnz}(B)m$ flops.



Note that these formulas agree with the one given above, $2mnp$, when the sparse matrices have all entries nonzero.

三重矩阵相乘

$$D = ABC$$

with A of size $m \times n$, B of size $n \times p$, and C of size $p \times q$.

The matrix D can be computed in two ways, as $(AB)C$ and as $A(BC)$.

In the first method we start with AB ($2mnp$ flops) and then form $D = (AB)C$ ($2mpq$ flops), for a total of $2mp(n + q)$ flops.

In the second method we compute the product BC ($2npq$ flops) and then form $D = A(BC)$ ($2mnq$ flops), for a total of $2nq(m + p)$ flops.



You might guess that the total number of flops required is the same with the two methods, but it turns out it is not. The first method is less expensive when $2mp(n + q) < 2nq(m + p)$, i.e., when

$$\frac{1}{n} + \frac{1}{q} < \frac{1}{m} + \frac{1}{p}$$

■ **Example 6.4** As a more specific example, consider the product

$$ab^T c$$

where a, b, c are n vectors.

If we first evaluate the outer product ab^T , the cost is n^2 flops, and we need to store n^2 values. We then multiply the vector c by this $n \times n$ matrix, which costs $2n^2$ flops. The total cost is $3n^2$ flops.

If we first evaluate the inner product $b^T c$, the cost is $2n$ flops, and we only need to store one number (the result). Multiplying the vector a by this number costs n flops, so the total cost is $3n$ flops. For n large, there is a dramatic difference between $3n$ and $3n^2$ flops.

(The storage requirements are also dramatically different for the two methods of evaluating $ab^T c$: 1 number versus n^2 numbers.) ■

6.2.3 矩阵向量乘积复杂度

矩阵 $A \in \mathbb{R}^{m \times n}$ 和向量 $x \in \mathbb{R}^n$ 的乘积 $y = Ax$, 需要 $(2n - 1)m$ flops;

乘积 $y \in \mathbb{R}^m$, 每个元素需要做向量内积, 需要 $2n - 1$ flops;

当 n 足够大时, 复杂度近似于 $2mn$;

特殊情况:

- A 为对角矩阵: n flops
- A 为下三角矩阵: n^2 flops
- A 为稀疏矩阵时: flops $\ll 2mn$

6.3 Special Matrices and Matrices in Different Applications

Definition 6.3.1 — Zero Matrix. 所有元素都为 0 的矩阵.

记作

$$0, 0_{m \times n}$$

Definition 6.3.2 — 单位矩阵. 为方形矩阵, 其中对角线元素为 1, 其它元素为 0.

记作 I 或者 I_n .

Corollary 6.3.1 I_n 的每一列是一个单位向量, 例如

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [e_1 \quad e_2 \quad e_3]$$

Definition 6.3.3 — Symmetric Matrices.

$$A_{ij} = A_{ji}, A^T = A$$

Definition 6.3.4 — Hermitian Matrices. $A_{ij} = \bar{A}_{ji}, A^H = A$ (共轭复数).

Definition 6.3.5 — Diagonal Matrices. 对角线上元素不全为 0, 其余元素全为 0.

对角矩阵用于膨胀 (dilation).

Definition 6.3.6 — 下三角矩阵. 方形矩阵且当 $i < j$ 时 $A_{ij} = 0$.

Definition 6.3.7 — 上三角矩阵. 方形矩阵且当 $i > j$ 时 $A_{ij} = 0$

通过 Gram-Schmidt 正交化算法可以化为上三角或者下三角矩阵。

6.3.1 $f(x) = Ax$ 中的 A

引入上节 $f(x) = Ax$ (矩阵-向量乘积函数) 的概念,

■ **Example 6.5 — Permutation Matrices.** f 颠倒向量 x 中的元素的顺序, 一个线性函数 $f(x) = Ax$

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

是一个置换矩阵. ■

Proof.

$$Ax = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_2 \\ x_1 \end{bmatrix}$$

■ **Example 6.6** f 对向量 x 中的元素进行升序排序, 非线性; ■

■ **Example 6.7** f 将向量 x 中的元素替换成相应的绝对值, 非线性; ■

■ **Example 6.8** — 反转矩阵.

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Proof.

$$Ax = \begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_2 \\ x_1 \end{bmatrix}$$

■ **Example 6.9** — 循环移位矩阵.

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Proof.

$$Ax = \begin{bmatrix} x_n \\ x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{bmatrix}$$

■ **Example 6.10** — 旋转矩阵.

$$A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

Ax 将向量 x 进行旋转, 角度为 θ .

■ **Example 6.11** — **Reflection Matrices.** Suppose that y is the vector obtained by reflecting x through the line that passes through the origin, inclined θ radians with respect to horizontal. Then

$$y = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} x$$

6.3.2 Selectors

■ **Definition 6.3.8** — Selector matrices.

■ **Definition 6.3.9** — Downsampling.

6.3.3 图论：节点弧关联矩阵

Definition 6.3.10 — 关联矩阵. 假设有向图 G 有 m 个顶点, n 条弧, 则关联矩阵 A 大小为 $m \times n$, 其中

$$A_{ij} = \begin{cases} 1 & \text{如果点 } i \text{ 是弧 } j \text{ 的终点} \\ -1 & \text{如果点 } i \text{ 是弧 } j \text{ 的起点} \\ 0 & \text{其它} \end{cases}$$

6.3.4 Convolution

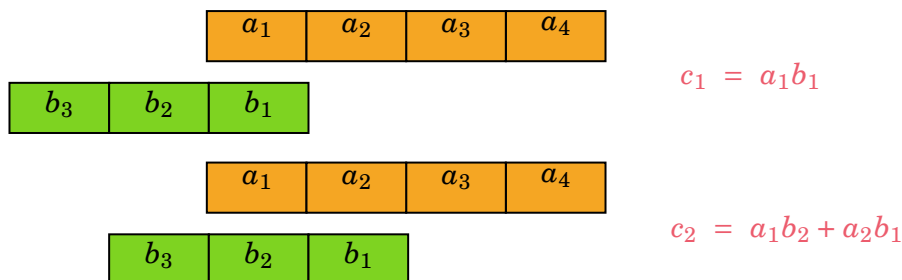
Definition 6.3.11 — 一维卷积. 向量 $a \in \mathbb{R}^n$ 和向量 $b \in \mathbb{R}^m$ 的卷积是一个 $(n + m - 1)$ 维向量 $c \in \mathbb{R}^{m+n-1}$

$$c_k = \sum_{i+j=k+1} a_i b_j, \quad k = 1, \dots, n + m - 1$$

记为 $c = a * b$

■ **Example 6.12** 设 $n = 4, m = 3$

Figure 6.2: An example of convolution



$$\begin{aligned} c_1 &= a_1 b_1 \\ c_2 &= a_1 b_2 + a_2 b_1 \\ c_3 &= a_1 b_3 + a_2 b_2 + a_3 b_1 \\ c_4 &= a_2 b_3 + a_3 b_2 + a_4 b_1 \\ c_5 &= a_3 b_3 + a_4 b_2 \\ c_6 &= a_4 b_3 \end{aligned}$$

Corollary 6.3.2 假设向量 a 和 b 分别是以下多项式的系数

$$p(x) = a_1 + a_2 x + \dots + a_n x^{n-1}, q(x) = b_1 + b_2 x + \dots + b_m x^{m-1}$$

则 $c = a * b$ 是多项式 $p(x)q(x)$ 的系数.

$$p(x)q(x) = c_1 + c_2 x + \dots + c_{m+n-1} x^{m+n-2}$$

Corollary 6.3.3 — 卷积性质. 有如下性质:

- 对称性: $a * b = b * a$
- 结合律: $(a * b) * c = a * (b * c)$
- 如果 $a * b = 0$, 则 $a = 0$, 或者 $b = 0$

Corollary 6.3.4 如果固定 a 或 b , 则 $c = a * b$ 是一个线性函数

■ **Example 6.13 — Toeplitz Matrix.** 4 维向量 a 和 3 维向量 b , 则 $c = a * b$

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{bmatrix} = \begin{bmatrix} a_1 & 0 & 0 \\ a_2 & a_1 & 0 \\ a_3 & a_2 & a_1 \\ a_4 & a_3 & a_2 \\ 0 & a_4 & a_3 \\ 0 & 0 & a_4 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 & 0 & 0 & 0 \\ b_2 & b_1 & 0 & 0 \\ b_3 & b_2 & b_1 & 0 \\ 0 & b_3 & b_2 & b_1 \\ 0 & 0 & b_3 & b_2 \\ 0 & 0 & 0 & b_3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

6.3.5 多项式

Definition 6.3.12 — 多项式. 多项式 $p(t)$, 度为 $n-1$, 系数为 x_1, x_2, \dots, x_n

$$p(t) = x_1 + x_2 t + x_3 t^2 + \dots + x_n t^{n-1}$$

Definition 6.3.13 — Vandermonde Matrices. $p(t)$ 在 m 个点中 t_1, t_2, \dots, t_m 的值为

$$\begin{bmatrix} p(t_1) \\ p(t_2) \\ \vdots \\ p(t_m) \end{bmatrix} = \begin{bmatrix} 1 & t_1 & \dots & t_1^{n-1} \\ 1 & t_2 & \dots & t_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_m & \dots & t_m^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = Ax$$

矩阵 A 被称为 *Vandermonde* 矩阵.

6.3.6 Fourier Transform

Definition 6.3.14 — Discrete Fourier Transform (DFT). DFT 将 n 维复向量 x 映射为 n 维复向量 y ($\mathbb{C}^n \rightarrow \mathbb{C}^n$)

$$y_k = \sum_{\ell=1}^n x_\ell e^{-i \frac{2\pi}{n} (k-1)(\ell-1)}, k = 1, \dots, n$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \dots & \omega^{-(n-1)} \\ 1 & \omega^{-2} & \omega^{-4} & \dots & \omega^{-2(n-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \omega^{-(n-1)} & \omega^{-2(n-1)} & \dots & \omega^{-(n-1)(n-1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

其中 $\omega = e^{2\pi i/n}$.

DFT 矩阵 W 的第 k 行第 l 列的元素为 $W_{kl} = \omega^{-(k-1)(l-1)}$.

Definition 6.3.15 — Discrete Inverse Fourier Transform.

$$x_\ell = \frac{1}{n} \sum_{k=1}^n y_k e^{i \frac{2\pi}{n} (k-1)(\ell-1)}, \ell = 1, \dots, n$$

6.3.7 Semi-Definite Matrices

Definition 6.3.16 — 半正定矩阵. 对称矩阵 $A \in \mathbb{R}^{n \times n}$ 称为半正定矩阵, 满足以下条件

$$x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n$$

Definition 6.3.17 — 正定矩阵. 对称矩阵 $A \in \mathbb{R}^{n \times n}$ 称为正定矩阵, 满足以下条件

$$x^T A x > 0 \quad \forall x \neq 0$$

Definition 6.3.18 — 二次型. 如果对称矩阵 $A \in \mathbb{R}^{n \times n}$, 则 $x^T A x$ 是二次型

Proof.

$$x^T A x = \sum_{i=1}^n \sum_{j=1}^n x_i A_{ij} x_j = \sum_{i=1}^n A_{ii} x_i^2 + 2 \sum_{i>j} A_{ij} x_i x_j$$

■

■ **Example 6.14**

$$A = \begin{bmatrix} 9 & 6 \\ 6 & a \end{bmatrix}$$

$$x^T A x = 9x_1^2 + 12x_1x_2 + ax_2^2 = (3x_1 + 2x_2)^2 + (a - 4)x_2^2$$

如果 $a > 4$, 矩阵 A 为正定矩阵:

$$x^T A x > 0 \quad \forall x \neq 0$$

如果 $a = 4$, 矩阵 A 为半正定矩阵, 但不是正定矩阵:

$$x^T A x \geq 0 \quad \forall x, \quad x^T A x = 0 \quad \exists x = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$

如果 $a < 4$, 矩阵 A 不是半正定矩阵:

$$x^T A x < 0 \quad \exists x = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$

■

Corollary 6.3.5 正定矩阵 A 都是非奇异的.

Proof.

$$Ax = 0 \Rightarrow x^T A x = 0 \Rightarrow x = 0$$

最后一步由正定性得到的. ($x^T A x > 0 \quad \forall x \neq 0$)

■

Theorem 6.3.6 — 正定矩阵对角元素性质. 正定矩阵 A 有正的对角元素.

$$A_{ii} = e_i^T A e_i > 0$$

Theorem 6.3.7 — 半正定矩阵对角元素性质. 每个半正定矩阵 A 都有非负的对角元素.

$$A_{ii} = e_i^T A e_i \geq 0$$

6.4 Gram 矩阵

Definition 6.4.1 — 实矩阵 A 的 Gram 矩阵.

$$G = A^T A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} [a_1, a_2, \dots, a_n] = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix}$$

Definition 6.4.2 — 复矩阵的 A 的 Gram 矩阵.

$$G = A^H A = \begin{bmatrix} a_1^H a_1 & a_1^H a_2 & \cdots & a_1^H a_n \\ a_2^H a_1 & a_2^H a_2 & \cdots & a_2^H a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^H a_1 & a_n^H a_2 & \cdots & a_n^H a_n \end{bmatrix}$$

Theorem 6.4.1 每个 Gram 矩阵都是半正定的.

Proof.

$$x^T A x = x^T B^T B x = \|Bx\|_2^2 \geq 0, \forall x$$

■

Theorem 6.4.2 如果 Gram 矩阵是正定的, 则要满足

$$x^T A x = x^T B^T B x = \|Bx\|_2^2 > 0 (\forall x \neq 0)$$

Corollary 6.4.3 如果 Gram 矩阵是正定的, 则 B 的列向量是线性无关的.

Proof.

$$\|Bx\|_2^2 > 0 (\forall x \neq 0)$$

所以 $\forall x \neq 0, Bx \neq 0$.

注意和线性无关 5.1.2 的定义进行参照.

■

7. Matrices Norms

7.1 矩阵范数

Definition 7.1.1 — Matrix Norm. 向量空间中存在一个函数 $\|\cdot\| : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$

且满足以下条件:

- 齐次性: $\|\alpha A\| = |\alpha| \|A\|, \alpha \in \mathbb{R} \text{ 且 } A \in \mathbb{R}^{m \times n};$
 - 三角不等式: $\|A + B\| \leq \|A\| + \|B\|, A, B \in \mathbb{R}^{m \times n};$
 - 非负性: $\|A\| \geq 0, A \in \mathbb{R}^{m \times n} \text{ 且 } \|A\| = 0 \Leftrightarrow A = 0;$
- 则称 $\|\cdot\|$ 为矩阵范数.

向量空间 $\mathbb{R}^{m \times n}$ 矩阵范数:

■ **Example 7.1 — F-范数 (Frobenius norm).**

$$\|A\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \right)^{\frac{1}{2}}$$

Proof.

$$\|A\|_F \geq 0$$

$$\|\alpha A\|_F = |\alpha| \|A\|_F, \alpha \in \mathbb{R}$$

$$\begin{aligned} \|A + B\|_F &= \left(\sum_{i=1}^n \sum_{j=1}^n (a_{ij} + b_{ij})^2 \right)^{\frac{1}{2}} \leq \left(\sum_{i=1}^n \sum_{j=1}^n (a_{ij})^2 \right)^{\frac{1}{2}} + \left(\sum_{i=1}^n \sum_{j=1}^n (b_{ij})^2 \right)^{\frac{1}{2}} \\ &= \|A\|_F + \|B\|_F \end{aligned}$$

Definition 7.1.2 — 从属于给定向量范数 $\|x\|_v$ 的矩阵范数. 设 $x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, \|\cdot\|_v$ 为一种向量范数. 则 $\frac{\|Ax\|_v}{\|x\|_v}$ 对所有 $x \neq 0$ 有最大值, 令

$$\|A\|_v = \max_{x \neq 0} \left\{ \frac{\|Ax\|_v}{\|x\|_v} \right\} = \max_{x \neq 0} \left\{ \left\| A \frac{x}{\|x\|_v} \right\|_v \right\} = \max_{\|y\|_v=1} \{\|Ay\|_v\}$$

即

$$\|A\|_v = \max_{x \neq 0} \left\{ \frac{\|Ax\|_v}{\|x\|_v} \right\}$$

$\|A\|_v$ 称为从属于给定向量范数 $\|x\|_v$ 的矩阵范数, 简称为从属范数或算子范数.

Proof. 可以验证 $\|A\|_v$ 满足矩阵范数定义.

$$\|A\|_v \geq 0$$

$$\|\alpha A\|_v = |\alpha| \|A\|_v, \alpha \in \mathbb{R}$$

$$\begin{aligned} \|A+B\|_v &= \max_{\|y\|_v=1} \|(A+B)y\|_v \\ &\leq \max_{\|y\|_v=1} \{\|Ay\|_v + \|By\|_v\} \\ &\leq \max_{\|y\|_v=1} \|Ay\|_v + \max_{\|y\|_v=1} \|By\|_v \\ &= \|A\|_v + \|B\|_v \end{aligned}$$

■



在本书中若未明确说明, $\|A\|$ 表示的是算子范数.

由定义 $\|A\|_v = \max_{x \neq 0} \left\{ \frac{\|Ax\|_v}{\|x\|_v} \right\}$ 可得

Definition 7.1.3 — 向量范数和算子范数相容.

$$\frac{\|Ax\|_v}{\|x\|_v} \leq \|A\|_v \Rightarrow \|Ax\|_v \leq \|A\|_v \|x\|_v$$

称向量范数和算子范数相容.

Theorem 7.1.1 — 算子范数服从乘法范数相容性. 对于 $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$

$$\begin{aligned} \|AB\|_v &= \max_{x \neq 0} \left\{ \frac{\|ABx\|_v}{\|x\|_v} \right\} \\ &\leq \max_{x \neq 0} \left\{ \frac{\|A\|_v \|Bx\|_v}{\|x\|_v} \right\} \\ &\leq \|A\|_v \max_{x \neq 0} \left\{ \frac{\|B\|_v \|x\|_v}{\|x\|_v} \right\} \\ &= \|A\|_v \|B\|_v \end{aligned}$$

算子范数服从乘法范数相容性.

根据向量的常用范数可以导出矩阵 $A \in \mathbb{R}^{m \times n}$ 的算子范数

Definition 7.1.4 — A 的列范数.

$$\|A\|_1 = \max_{x \neq 0} \left(\frac{\|Ax\|_1}{\|x\|_1} \right) = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$$

Definition 7.1.5 — A 的行范数.

$$\|A\|_\infty = \max_{x \neq 0} \left(\frac{\|Ax\|_\infty}{\|x\|_\infty} \right) = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$$

Definition 7.1.6 — A 的 2-范数.

$$\|A\|_2 = \max_{x \neq 0} \left(\frac{\|Ax\|_2}{\|x\|_2} \right) = \sqrt{\lambda_{\max}(A^T A)} \quad (7.1)$$

Proof. For any A Choose x to be the eigenvector of $A^T A$ with largest eigenvalue λ_{\max} . The ratio in equation 7.1 is $\mathbf{x}^T A^T A \mathbf{x} = \mathbf{x}^T (\lambda_{\max}) \mathbf{x}$ divided by $\mathbf{x}^T \mathbf{x}$. This is λ_{\max} .

No \mathbf{x} can give a larger ratio. The symmetric matrix $A^T A$ has eigenvalues $\lambda_1, \dots, \lambda_n$ and orthonormal eigenvectors $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$. Every \mathbf{x} is a combination of those vectors. Try this combination in the ratio and remember that $\mathbf{q}_i^T \mathbf{q}_j = 0$:

$$\frac{\mathbf{x}^T A^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \frac{(c_1 \mathbf{q}_1 + \dots + c_n \mathbf{q}_n)^T (c_1 \lambda_1 \mathbf{q}_1 + \dots + c_n \lambda_n \mathbf{q}_n)}{(c_1 \mathbf{q}_1 + \dots + c_n \mathbf{q}_n)^T (c_1 \mathbf{q}_1 + \dots + c_n \mathbf{q}_n)} = \frac{c_1^2 \lambda_1 + \dots + c_n^2 \lambda_n}{c_1^2 + \dots + c_n^2}$$

The maximum ratio λ_{\max} is when all c 's are zero, except the one that multiplies λ_{\max} . ■



The ratio in equation 7.1 is the Rayleigh quotient for the symmetric matrix $A^T A$. Its maximum is the largest eigenvalue $\lambda_{\max}(A^T A)$. The minimum ratio is $\lambda_{\min}(A^T A)$. If you substitute any vector \mathbf{x} into the Rayleigh quotient $\mathbf{x}^T A^T A \mathbf{x} / \mathbf{x}^T \mathbf{x}$, you are guaranteed to get a number between $\lambda_{\min}(A^T A)$ and $\lambda_{\max}(A^T A)$.



The norm $\|A\|$ equals the largest singular value σ_{\max} of A . The singular values $\sigma_1, \dots, \sigma_r$ are the square roots of the positive eigenvalues of $A^T A$. So certainly $\sigma_{\max} = (\lambda_{\max})^{1/2}$. Since U and V are orthogonal in $A = U \Sigma V^T$, the norm is $\|A\| = \sigma_{\max}$.

■ **Example 7.2** 求矩阵 A 的各种常用范数

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

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| = \max_{1 \leq j \leq n} \{2, 5, 2\} = 5$$

$$\|A\|_{\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| = \max_{1 \leq i \leq n} \{3, 4, 2\} = 4$$

由于 $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$, 因此先求 $A^T A$ 的特征值

$$A^T A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 2 & 1 \\ 0 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 0 \\ -1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 9 & -1 \\ 1 & -1 & 2 \end{pmatrix}$$

特征方程为

$$\det(\lambda I - A^T A) = \begin{vmatrix} \lambda - 2 & 0 & -1 \\ 0 & \lambda - 9 & 1 \\ -1 & 1 & \lambda - 2 \end{vmatrix} = 0$$

可得 $A^T A$ 的特征值

$$\lambda_1 = 9.1428, \lambda_2 = 2.9211, \lambda_3 = 0.9361$$

■



对于 $\|A\|_2$ 需要计算 $\lambda_{\max}(A^T A)$, 直接根据特征方程计算特征值的算法复杂度太高.

8. 适定问题

8.1 The Definition of Well-posed Problem

In 1923, the French mathematician Hadamard introduced the notion of well-posed (适定) problem:

- A solution for the problem exists;
- The solution is unique;
- Perturbations in the data should cause small perturbations in the solution.

One of these conditions is not satisfied, the problem is said to be ill-posed (病态) and demands a special consideration.

■ **Example 8.1** 假设 A 是非奇异矩阵

$$Ax = b$$

如果将 b 为 $b + \Delta b$, 方程新的解 $x + \Delta x$, 则有:

$$A(x + \Delta x) = b + \Delta b$$

即

$$\Delta x = A^{-1} \Delta b$$

如果小的变化 Δb 导致小变化 Δx , 则称解是稳定的. 如果小的变化 Δb 导致大变化 Δx , 则称解不稳定的.

设

$$A = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 + 10^{-10} & 1 - 10^{-10} \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} 1 - 10^{10} & 10^{10} \\ 1 + 10^{10} & -10^{10} \end{bmatrix}$$

若 $b = (1, 1)$, 方程 Ax 的解 $x = (1, 1)$. 如果将 b 改为 $b + \Delta b$, 那么 x 的变化量为

$$\Delta x = A^{-1} \Delta b = \begin{bmatrix} \Delta b_1 - 10^{10} (\Delta b_1 - \Delta b_2) \\ \Delta b_1 + 10^{10} (\Delta b_1 - \Delta b_2) \end{bmatrix}$$

很小变化 Δb 会导致非常大变化 Δx ! 由矩阵 A 定义的问题, 称为适定问题或病态问题.

■

8.2 绝对误差的界限

假设 A 是非奇异的, 并给出定义:

Notation 8.1.

$$x = A^{-1}b$$

$$\Delta x = A^{-1}\Delta b$$

$\|\Delta x\|$ 的上界为:

$$\|\Delta x\|_2 \leq \|A^{-1}\|_2 \|\Delta b\|_2$$

矩阵范数 $\|A^{-1}\|_2$ 小时, 当 $\|\Delta b\|_2$ 变化很小, $\|\Delta x\|_2$ 也很小; $\|A^{-1}\|_2$ 大时, $\|\Delta x\|_2$ 可能很大, 即使 $\|\Delta b\|_2$ 很小.

8.3 相对误差的界限

假设 $b \neq 0$; 因此 $x \neq 0$

$\|\Delta x\|_2/\|x\|_2$ 的上界为:

$$\begin{aligned} \|\Delta x\|_2 &= \|A^{-1}\Delta b\|_2 \leq \|A^{-1}\|_2 \|\Delta b\|_2 \text{ (向量范数和算子范数相容)} \\ \Rightarrow \frac{\|\Delta x\|_2}{\|x\|_2} &\leq \frac{\|A^{-1}\|_2 \|\Delta b\|_2}{\|x\|_2} = \frac{\|A\|_2 \|A^{-1}\|_2 \|\Delta b\|_2}{\|x\|_2 \|A\|_2} \leq \frac{\|A\|_2 \|A^{-1}\|_2 \|\Delta b\|_2}{\|b\|_2} \end{aligned}$$

由 $\|b\|_2 = \|Ax\|_2 \leq \|A\|_2 \|x\|_2$, 可得

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq \|A\|_2 \|A^{-1}\|_2 \frac{\|\Delta b\|_2}{\|b\|_2}$$

$\|A\|_2 \|A^{-1}\|_2$ 小, 当 $\frac{\|\Delta b\|_2}{\|b\|_2}$ 相对变化很小时, $\frac{\|\Delta x\|_2}{\|x\|_2}$ 也变化很小;

$\|A\|_2 \|A^{-1}\|_2$ 大, $\frac{\|\Delta x\|_2}{\|x\|_2}$ 可远远大于 $\frac{\|\Delta b\|_2}{\|b\|_2}$.

Definition 8.3.1 — 非奇异矩阵 A 的条件数 (condition number).

$$\kappa(A) = \|A\|_2 \|A^{-1}\|_2$$

Corollary 8.3.1 — 非奇异矩阵 A 的条件数 (condition number) 性质. 有如下性质:

- 对于所有 A , 有 $\kappa(A) \geq 1$;
- 如果 $\kappa(A)$ 比较小 (接近 1), x 的相对误差接近 b 的相对误差;
- 如果 $\kappa(A)$ 比较大 (超过 100), x 的相对误差比 b 的相对误差大得多.

9. Inverse of Matrices

9.1 Left Inverse, Right Inverse, Inverse

Definition 9.1.1 — A 的左逆. 当一个矩阵 X 满足

$$XA = I$$

X 被称为 A 的左逆; 当左逆存在时, 则称 A 是可左逆的;

如果左逆矩阵存在, 则左逆矩阵有无穷多个.

■ **Example 9.1**

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix}$$

矩阵 A 是可左逆的, 其左逆矩阵有两个

$$B = \frac{1}{9} \begin{bmatrix} -11 & -10 & 16 \\ 7 & 8 & -11 \end{bmatrix} \quad C = \frac{1}{2} \begin{bmatrix} 0 & -1 & 6 \\ 0 & 1 & -4 \end{bmatrix}$$

■ **Definition 9.1.2** — A 的右逆. 当左逆存在时, 则称 A 是可左逆的;

如果右逆矩阵存在, 则右逆矩阵有无穷多个.

■ **Example 9.2**

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

矩阵 B 可右逆, 以下矩阵都是 B 的右逆

$$D = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}, E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, G = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

一个大小为 $m \times n$ 的矩阵, 其左逆或右逆的维度为 $n \times m$.

Theorem 9.1.1 A 的左逆为 X 当且仅当 X^T 是 A^T 的右逆.

Proof.

$$A^T X^T = (XA)^T = I$$

Theorem 9.1.2 A 的右逆为 X 当且仅当 X^T 是 A^T 的左逆.

Proof.

$$X^T A^T = (AX)^T = I$$

Theorem 9.1.3 如果矩阵 A 存在左逆和右逆, 则左逆和右逆一定相等

Proof.

$$\begin{aligned} XA &= I, AY = I \\ \Rightarrow X &= XI = X(AY) = (XA)Y = Y \\ \Rightarrow X &= Y \end{aligned}$$

Definition 9.1.3 — 逆 A^{-1} . 如果矩阵 A 存在左逆和右逆, 此时 X 称为矩阵的逆, 记作 A^{-1} 当矩阵的逆存在时, 则称矩阵 A 可逆.

9.2 Linear Equation Systems

Definition 9.2.1 — Linear Equation Systems. 有 n 个变量的 m 个方程为

$$\begin{cases} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n = b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n = b_2 \\ \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n = b_m \end{cases}$$

写成矩阵形式为: $Ax = b$. 其中 A 为系数矩阵, x 为 n 维列向量.

该方程组可能无解, 有唯一解和无穷解.

9.2.1 线性方程组求解

Theorem 9.2.1 如果矩阵 A 可左逆, 假设 X 是矩阵 A 的左逆, 则至多一个解, 如有解则 $x = Xb$.

Proof.

$$Ax = b \Rightarrow x = XAx = Xb$$

列满秩时 (下面证明), 列向量线性无关, 所以其零空间中只有零解, 方程 $Ax = b$ 可能有一个唯一解 (b 在 A 的列空间中, 此特解就是全部解, 因为通常的特解可以通过零空间中的向量扩展出一组解集, 而此时零空间只有 0 向量), 也可能无解 (b 不在 A 的列空间中)

Theorem 9.2.2 如果矩阵 A 可右逆, 假设 Y 是矩阵 A 的右逆, 则至少一个解, 即 $x = Yb$.

Proof. 设 $x = Yb$

$$x = Yb \Rightarrow Ax = AYb = b$$

右逆就是研究 $m \times n$ 矩阵 A 行满秩的情况, 此时 $n > m = \text{rank}(A)$. 对称的, 其左零空间中仅有零向量, 即没有行向量的线性组合能够得到零向量. ($N(A^T) = \{0\}$) ■

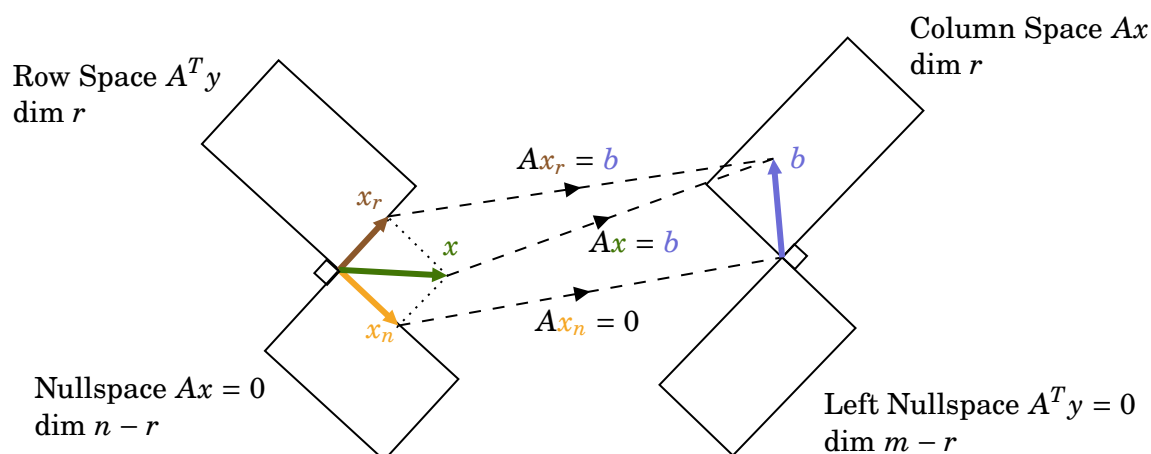
Theorem 9.2.3 如果矩阵 A 可逆的, 假设 X 是矩阵 A 的逆, 则

$$Ax = b \Rightarrow x = A^{-1}b$$

唯一解.

9.3 Fundamental Theorem of Linear Algebra

Figure 9.1: Four Subspace of Matrix A



$r = m$	$r = n$	Square and invertible	$Ax = b$ has 1 solution
$r = m$	$r < n$	Short and wide	$Ax = b$ has ∞ solutions
$r < m$	$r = n$	Tall and thin	$Ax = b$ has 0 or 1 solution
$r < m$	$r < n$	Not full rank	$Ax = b$ has 0 or ∞ solutions

The set of linear equations is called *over-determined* if $m > n$, *under-determined* if $m < n$, and *square* if $m = n$.

A set of equations with zero right-hand side, $Ax = 0$, is called a *homogeneous* set of equations. Any homogeneous set of equations has $x = 0$ as a solution.

9.4 Invertible Matrices

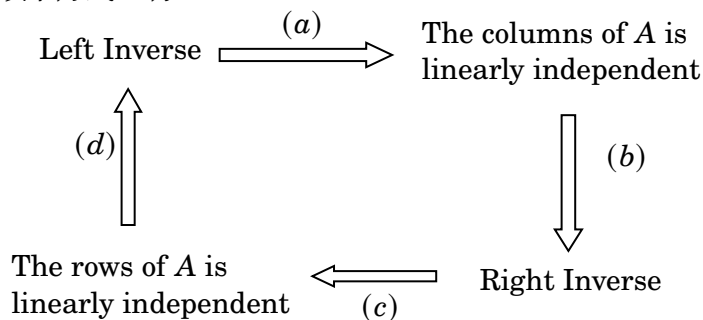
Theorem 9.4.1 对于方阵 $A \in \mathbb{R}^{n \times n}$, 以下条件都是等价的:

1. A 可左逆

2. A 的列向量线性无关
3. A 可右逆
4. A 的行向量线性无关
5. A 可逆

此时矩阵 A 为非奇异矩阵, 由条件 1 与 3, 可得 A 为可逆矩阵.

Proof. 可以通过以下方式证明:



- 性质 (a) 对任意矩阵 $A \in \mathbb{R}^{m \times n}$ 都成立
- 性质 (b) 对方阵矩阵 $A \in \mathbb{R}^{n \times n}$ 都成立
- 对于性质 (c) 与 (d), 可利用 A^T 证明

■

Theorem 9.4.2 (a): A 可左逆, 则 A 列向量线性无关.

Proof. 假设 A 的左逆是 B , 则

$$\begin{aligned} Ax = 0 &\Rightarrow BAx = 0 \\ &\Rightarrow Ix = 0 \end{aligned}$$

假设 A 的列向量 $A = [a_1, a_2, \dots, a_n]$

$$Ax = x_1 a_1 + x_2 a_2 + \dots + x_n a_n = 0$$

则当该等式 $Ax = 0$ 成立时, 其解 $x = 0$, 则 A 的列向量线性无关.

Corollary 9.4.3 如果 $A \in \mathbb{R}^{m \times n}$ 有左逆, 则有 $m \geq n = r$.

即 A 是高或方的矩阵, 如 $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$.

此时 A 的行向量可能线性相关, 而 A 的列向量线性无关. $N(A) = \{0\}$.

假设 A 的列向量 $A = [a_1, a_2, \dots, a_n]$

$$Ax = x_1 a_1 + x_2 a_2 + \dots + x_n a_n = b$$

$$Ay = y_1 a_1 + y_2 a_2 + \dots + y_n a_n = b$$

$$Ax - Ay = A(x - y) = 0 \Rightarrow x = y$$

当 $b \in \mathbb{R}^m, b \notin \{y \mid y = Ax, x \in \mathbb{R}^n\}$ 时 (即 b 不在 A 的列空间, $m \geq n$ 时), 线性方程组无解. $Ax = b$ 至多一个解, 如有解则 $x = Xb$. ■

Theorem 9.4.4 矩阵的行秩等于列秩.

Proof. 令 A 是一个 $m \times n$ 的矩阵, 其列秩为 r . 因此矩阵 A 的列空间的维度是 r .

令 c_1, c_2, \dots, c_r 是 A 的列空间的一组基, 构成 $m \times r$ 矩阵 C 的列向量 $C = [c_1, c_2, \dots, c_r]$, 并使得 A 的每个列向量是 C 的 r 个列向量的线性组合.

由矩阵乘法的定义, 存在一个 $r \times n$ 矩阵 R , 使得 $A = CR$. (A 的 (i, j) 元素是 c_i 与 R 的第 j 个行向量的点积.)

现在, 由于 $A = CR$, A 的每个行向量是 R 的行向量的线性组合, 这意味着 A 的行向量空间被包含于 R 的行向量空间之中. 因此 A 的行秩 $\leq R$ 的行秩. 但 R 仅有 r 行, 所以 R 的行秩 $\leq r = A$ 的列秩. 这就证明了 A 的行秩 $\leq A$ 的列秩.

把上述证明过程中的“行”与“列”交换, 利用对偶性质同样可证 A 的列秩 $\leq A$ 的行秩. 更简单的方法是考虑 A 的转置矩阵 A^T , 则 A 的列秩 $= A^T$ 的行秩 $\leq A^T$ 的列秩 $= A$ 的行秩. 这证明了 A 的列秩等于 A 的行秩. 证毕. ■

Theorem 9.4.5 (c): 矩阵 $A \in \mathbb{R}^{m \times n}$ 有右逆 X , 则 A 行向量线性无关.

Proof.

$$X^T A^T = (AX)^T = I$$

则有 X^T 是 A^T 的左逆, A^T 的列向量线性无关.

即 $A^T \in \mathbb{R}^{n \times m}$.

Corollary 9.4.6 如果 $A \in \mathbb{R}^{m \times n}$ 有左逆, 则有 $r = m \leq n$.

即 A 是宽或方的矩阵.

此时 A 的列向量可能线性相关, 而 A 的行向量线性无关.

$N(A^T) = \{0\}$, $\dim N(A) = n - r$, $r = m$. ($Ax = b$ 有无穷解)

根据定理“矩阵的行秩等于列秩”, A^T 的列向量线性无关, 则矩阵 A 有 m 个线性无关列向量 (行向量), 即通过 Gram-Schmidt 正交化可得 m 个正交基.

$\forall b \in \mathbb{R}^m$, 有 $b \in \{y \mid y = Ax, x \in \mathbb{R}^n\}$ ($m \leq n$), 方程 $Ax = b$ 有解, 其解为 $x = Xb$.

Theorem 9.4.7 (b): 若方阵 A 列向量线性无关, 则 A 可右逆.

Proof. 假设 $A \in \mathbb{R}^{n \times n}$ 为方阵且列向量线性无关

$$A = [a_1, a_2, \dots, a_n]$$

则对于任意向量 $b \in \mathbb{R}^n$, 则向量组 $[a_1, a_2, \dots, a_n, b]$ 线性相关, 存在不全为 0 的系数, 使得以下等式成立

$$x_1 a_1 + x_2 a_2 + \dots + x_n a_n + x_{n+1} b = 0$$

因为 A 列向量线性无关, 则 $x_{n+1} \neq 0$ (假设 $x_{n+1} = 0$ 会推出违反线性无关假设的结论), 即 b 是 A 列向量的线性组合;

$$b = -\frac{x_1}{x_{n+1}} a_1 - \frac{x_2}{x_{n+1}} a_2 - \dots - \frac{x_n}{x_{n+1}} a_n$$

存在向量 $c_1, \dots, c_n \in \mathbb{R}^n$, 使得

$$Ac_1 = e_1$$

$$Ac_2 = e_2$$

...

$$Ac_n = e_n$$

则矩阵 $C = [c_1 c_2 \dots c_n]$ 是矩阵 A 的右逆, $AC = I$.

■

9.5 转置和共轭转置的逆

Theorem 9.5.1 — 转置 A^T 和共轭转置 A^H . 如果矩阵 A 为非奇异矩阵, 则其转置 A^T 和共轭转置 A^H 都为非奇异矩阵, 则有

$$(A^T)^{-1} = (A^{-1})^T, \quad (A^H)^{-1} = (A^{-1})^H$$

Proof.

$$(AA^{-1})^T = I \Rightarrow \underbrace{(A^{-1})^T}_{\text{the inverse of } A^T} A^T = I$$

■

Corollary 9.5.2 如果矩阵 A 和矩阵 B 都为非奇异矩阵, 则乘积 AB 也为非奇异矩阵

$$(AB)^{-1} = B^{-1}A^{-1}$$

Proof.

$$(AB) \underbrace{B^{-1}A^{-1}}_{\text{the inverse of } AB} = I$$

■

9.6 Gram Matrix 非奇异的性质

Gram 矩阵的定义见 6.4.1.

Corollary 9.6.1 — Gram Matrix 可逆等价于 A 列线性无关. 矩阵 $A \in \mathbb{R}^{m \times n}$, $G = A^T A$ 矩阵 A 列向量线性无关 \Leftrightarrow Gram 矩阵 G 非奇异.

Proof. " \Rightarrow ":

假设矩阵 A 列向量线性无关, $A^T A$ 奇异. 则存在 $A^T A x = 0, x \neq 0$, 可得 $x^T A^T A x = \|Ax\|_2^2 = 0$, 即 $Ax = 0$ 与列向量线性无矛盾.

" \Leftarrow ":

假设 $A^T A$ 非奇异, 矩阵 A 列向量线性相关. 则有 $Ax = 0, x \neq 0$, 可得 $A^T A x = 0$, 即 $A^T A$ 是奇异矩阵. ■

9.7 伪逆

Definition 9.7.1 — Pseudo-inverse.

$$A^\dagger = A^T (AA^T)^{-1}$$

$$A^\dagger = V\Sigma^+U^T = [v_1 \cdots v_r \cdots v_n] \begin{bmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \end{bmatrix} [u_1 \cdots u_r \cdots u_m]^T$$

Theorem 9.7.1 伪逆 A^\dagger 为 A 的右逆

Proof.

$$AA^\dagger = AA^T (AA^T)^{-1} = (AA^T)^{-1} (AA^T) = I$$

■

Theorem 9.7.2 当 A 为方阵时, 右逆等于矩阵的逆

Proof.

$$A^\dagger = A^T (AA^T)^{-1} = A^T A^{-T} A^{-1} = A^{-1}$$

■

Theorem 9.7.3

Corollary 9.7.4 以下三个结论为等价的, 对于实矩阵 A

- A 是可左逆的
- A 的列向量线性无关
- $A^T A$ 为非奇异矩阵

Corollary 9.7.5 以下三个结论为等价的, 对于实矩阵 A

- A 是可右逆的
- A 的行向量线性无关
- AA^T 为非奇异矩阵

By choosing good bases, A multiplies \mathbf{v}_i in the row space to give $\sigma_i \mathbf{u}_i$ in the column space. A^{-1} must do the opposite!

If $A\mathbf{v} = \sigma\mathbf{u}$ then $A^{-1}\mathbf{u} = \mathbf{v}/\sigma$. The singular values of A^{-1} are $1/\sigma$, just as the eigenvalues of A^{-1} are $1/\lambda$. The bases are reversed. The \mathbf{u} 's are in the row space of A^{-1} , the \mathbf{v} 's are in the column space.

The pseudoinverse A^+ is an n by m matrix. **If A^{-1} exists, then A^+ is the same as A^{-1} .** In that case $m = n = r$ and we are inverting $U\Sigma V^T$ to get $V\Sigma^{-1}U^T$.

The new symbol A^+ is needed when $r < m$ or $r < n$. Then A has no two-sided inverse, but it has a *pseudoinverse* A^+ with that same rank r :

$$A^+ \mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{v}_i \quad \text{for } i \leq r \quad \text{and} \quad A^+ \mathbf{u}_i = \mathbf{0} \quad \text{for } i > r$$

The vectors $\mathbf{u}_1, \dots, \mathbf{u}_r$ in the column space of A go back to $\mathbf{v}_1, \dots, \mathbf{v}_r$ in the row space.

The other vectors $\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ are in the left nullspace, and A^+ sends them to zero. When we know what happens to all those basis vectors, we know A^+ .

Notice the pseudoinverse of the diagonal matrix Σ . Each σ in Σ is replaced by σ^{-1} in Σ^+ . The product $\Sigma^+\Sigma$ is as near to the identity as we can get. It is a projection matrix, $\Sigma^+\Sigma$ is partly I and otherwise zero. We can invert the σ 's, but we can't do anything about the zero rows and columns.

Figure 9.2: $A\mathbf{x}^\dagger$ in the column space goes back to $A^\dagger A\mathbf{x}^\dagger = \mathbf{x}^\dagger$ in the row space

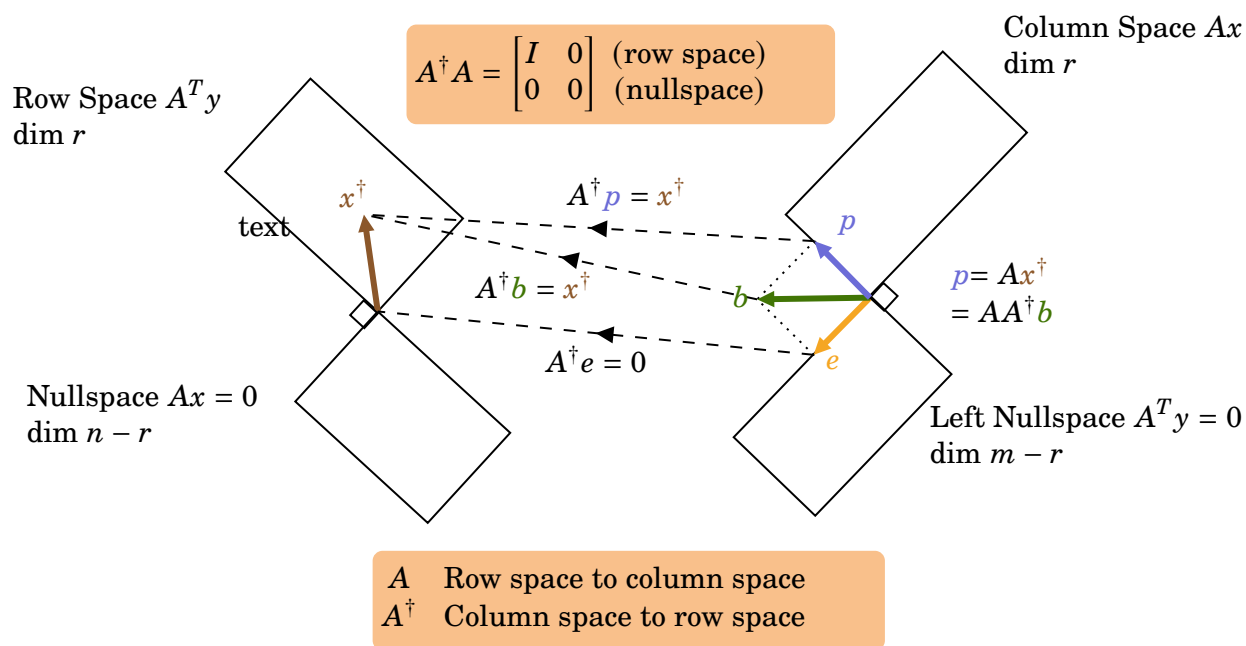
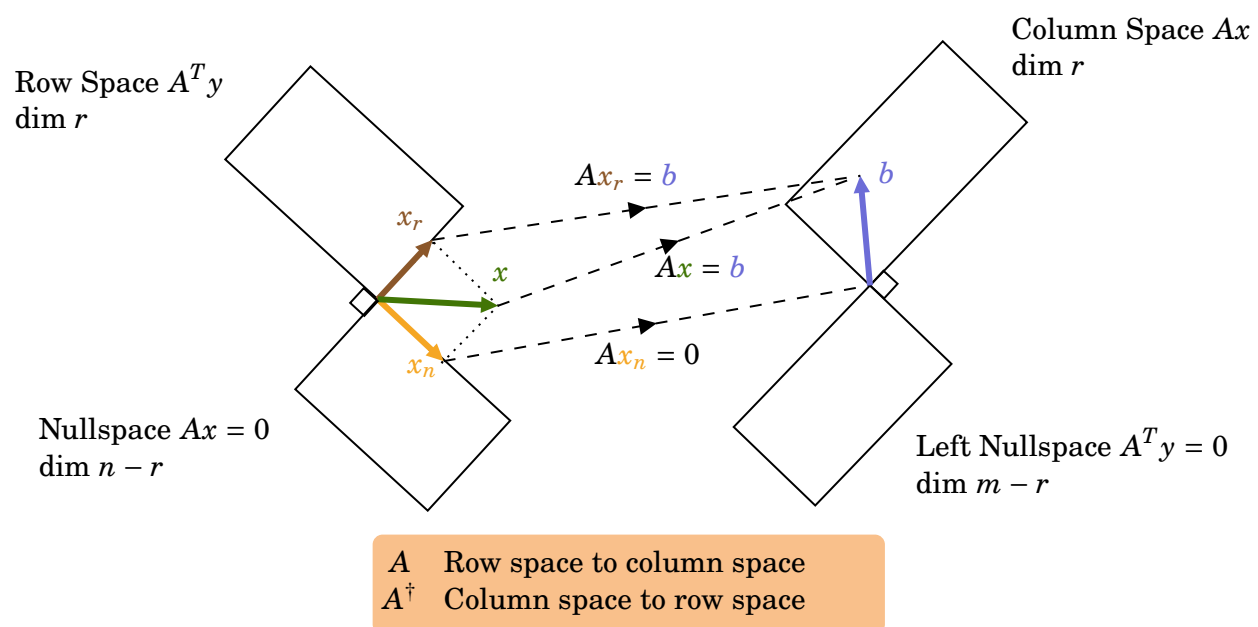


Figure 9.3: Projection from row space to column space



10. Orthogonal Matrices

10.1 预备知识

10.1.1 标准正交向量

参见 5.3.2。

10.1.2 Gram 矩阵与标准正交的关系

For the definition of Gram matrices, refer to 6.4.1. 关于它与非奇异的性质，参见 9.6.

Theorem 10.1.1 如果 A 的 Gram 矩阵为单位矩阵，则 $A \in \mathbb{R}^{m \times n}$ 具有标准正交列.

Proof.

$$\begin{aligned} A^T A &= \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \\ &= \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \\ &= I \end{aligned}$$

■

10.1.3 矩阵-向量乘积与标准正交的关系

如果 $A \in \mathbb{R}^{m \times n}$ 具有标准正交列，则线性函数 $f(x) = Ax$

Theorem 10.1.2 保持原内积.

$$\langle Ax, Ay \rangle = x^T y$$

Proof.

$$\langle Ax, Ay \rangle = (Ax)^T (Ay) = x^T A^T A y = x^T y$$

■

Theorem 10.1.3 保持原范数.

$$\|Ax\|_2 = \|x\|_2$$

Proof.

$$\|Ax\|_2 = \left((Ax)^T (Ax) \right)^{1/2} = \left(x^T x \right)^{1/2} = \|x\|_2$$

■

Theorem 10.1.4 保持原距离.

$$\|Ax - Ay\|_2 = \|x - y\|_2$$

Proof.

$$\|Ax - Ay\|_2 = \left((Ax - Ay)^T (Ax - Ay) \right)^{1/2} = \left((x - y)^T (x - y) \right)^{1/2} = \|x - y\|_2$$

■

Theorem 10.1.5 保持原角度.

$$\angle(Ax, Ay) = \angle(x, y)$$

Proof.

$$\angle(Ax, Ay) = \arccos \left(\frac{(Ax)^T (Ay)}{\|Ax\|_2 \|Ay\|_2} \right) = \arccos \left(\frac{x^T y}{\|x\|_2 \|y\|_2} \right) = \angle(x, y)$$

■

10.1.4 左可逆性与正交的关系

如果矩阵 $A \in \mathbb{R}^{m \times n}$ 有标准正交列，则：

Theorem 10.1.6 A 是左可逆的，其左逆为 A^T .

Proof. 根据定义：

$$A^T A = I$$

■

Theorem 10.1.7 A 有线性无关的列向量.

Proof.

$$Ax = 0 \Rightarrow A^T Ax = x = 0$$

■

Theorem 10.1.8 A 是高的或者方的, 即 $m \geq n$.

Proof. 列向量 $a_1, a_2, \dots, a_n \in \mathbb{R}^m$, 由维度定理可得 $n \leq m$ 。 ■

10.2 正交矩阵

Definition 10.2.1 — 正交矩阵. 所有列两两相互正交的方形实矩阵。

Theorem 10.2.1 — 正交矩阵满足非奇异性. 即如果矩阵 A 是正交的, 则 A 是可逆的, 左逆等于右逆, 且它的逆为 A^T .

$$\left. \begin{array}{l} A^T A = I \\ A \text{ 是方的} \end{array} \right\} \Rightarrow AA^T = I$$

Corollary 10.2.2 A^T 也是一个正交矩阵。

Corollary 10.2.3 A 的行是标准正交的, 即 a_i 范数为 1 且相互正交。



如果 $A \in \mathbb{R}^{m \times n}$ 有标准正交列以及 $m > n$, 则 $AA^T \neq I$ 。

10.3 Permutation Matrices

Notation 10.1. $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ 为 $(1, 2, \dots, n)$ 的一个重新排序的排列。

将 π 与一个置换矩阵 $A \in \mathbb{R}^{n \times n}$ 联系起来:

$$A_{i\pi_i} = 1, \quad A_{ij} = 0 \text{ 如果 } j \neq \pi_i$$

Definition 10.3.1 — 置换. Ax 是 x 的一个置换:

$$Ax = (x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_n})$$

A 在每一行和每一列中都有一个等于 1 的元素。

Theorem 10.3.1 置换矩阵满足正交性, 即所有置换矩阵都是正交的。

Corollary 10.3.2

$$A^T A = I$$

Proof. 因为 A 的每一行有一个元素等于 1

$$(A^T A)_{ij} = \sum_{k=1}^n A_{ki} A_{kj} = \begin{cases} 1 & i = j \\ 0 & \text{其它} \end{cases}$$

■

Corollary 10.3.3 $A^T = A^{-1}$ 是逆置换矩阵.

■ **Example 10.1** 若 $\{1, 2, 3, 4\}$ 的置换为:

$$(\pi_1, \pi_2, \pi_3, \pi_4) = (2, 4, 1, 3)$$

相应的置换矩阵及其逆矩阵为

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, A^{-1} = A^T = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

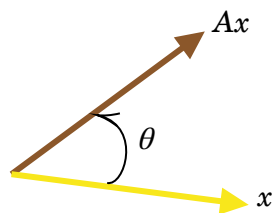
A^T 是与置换相关的置换矩阵

$$(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4) = (3, 1, 4, 2)$$

■

10.4 平面旋转

Figure 10.1: An example of rotation



■ **Example 10.2 — Rotation Matrices in \mathbb{R}^2 .** 在一个 2 维平面的旋转可以用矩阵表示为

$$A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

■

■ **Example 10.3 — Rotation Matrices in \mathbb{R}^3 .**

$$A = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix}$$

描述了在 \mathbb{R}^3 中 (x_1, x_3) 平面的旋转。

■

10.5 Householder Matrix

Definition 10.5.1 — Householder matrix, reflector matrix.

$$A = I - 2aa^T$$

其中，向量 a 满足 $\|a\|_2 = 1$ 。

Theorem 10.5.1 反射矩阵 (reflector matrix) 是对称的。

$$A^T = A$$

Theorem 10.5.2 反射矩阵 (reflector matrix) 是正交的。

Proof.

$$A^T A = (I - 2aa^T)(I - 2aa^T) = I - 4aa^T + 4aa^T aa^T = I$$

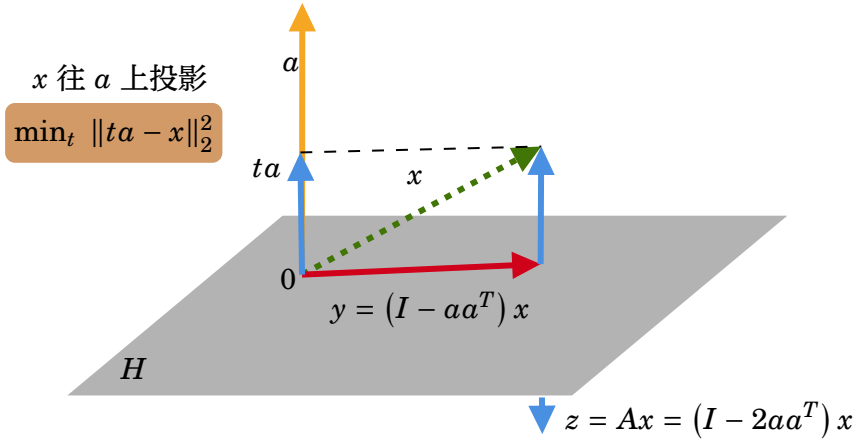
■

Theorem 10.5.3 反射矩阵是对合矩阵 (Involutory matrix)，即它的逆是它本身。

$$H = H^{-1}$$

10.5.1 The Geometry of Householder Transformation

Figure 10.2: Elementary Reflection



$H = \{u \mid a^T u = 0\}$ 是与 a 正交的向量的(超)平面。

Corollary 10.5.4 如果 $\|a\|_2 = 1$ ， x 在 H 上的投影为

$$y = x - (a^T x) a = x - a (a^T x) = (I - aa^T) x$$

Proof. 1. $y \in H$.

$$a^T y = a^T (x - a (a^T x)) = a^T x - (a^T a) (a^T x) = a^T x - a^T x = 0$$

2. 考虑任意 $z \in H (z \neq y)$, 证明 $\|x - z\| > \|x - y\|$

$$\begin{aligned}
\|x - z\|_2^2 &= \|x - y + y - z\|_2^2 \\
&= \|x - y\|_2^2 + 2(x - y)^T(y - z) + \|y - z\|_2^2 \\
&= \|x - y\|_2^2 + 2(a^T x) a^T(y - z) + \|y - z\|_2^2 \\
&= \|x - y\|_2^2 + \|y - z\|_2^2 \quad (\text{因为 } a^T y = a^T z = 0) \\
&\geq \|x - y\|_2^2
\end{aligned}$$

■

Corollary 10.5.5 x 通过超平面的反射由反射算子的乘积给出

$$z = y + (y - x) = (I - 2aa^T)x$$

10.6 正交矩阵乘积

若 $A_1, \dots, A_k \in \mathbb{R}^{n \times n}$ 是正交矩阵, 那么它们的乘积为:

$$A = A_1 A_2 \cdots A_k$$

Corollary 10.6.1 — 正交矩阵乘积的正交性.

$$\begin{aligned}
A^T A &= (A_1 A_2 \cdots A_k)^T (A_1 A_2 \cdots A_k) \\
&= A_k^T \cdots A_2^T A_1^T A_1 A_2 \cdots A_k \\
&= I
\end{aligned}$$

10.7 具有正交矩阵的线性方程

系数正交矩阵 $A \in \mathbb{R}^{n \times n}$ 的线性方程:

$$Ax = b$$

解为:

$$x = A^{-1}b = A^T b$$

10.7.1 The Complexity of the Multiplication Ax of Orthogonal Matrix A

可以在 $2n^2$ 个 flop 内计算矩阵向量乘法。

如果 A 有特殊性质, 代价将会小于 n^2 。例如,

- 置换矩阵: 0 flop。
- 反射算子 (给定 a): $4n$ flops。
- 平面旋转: $O(1)$ flop。

10.8 列标准正交的高矩阵

Theorem 10.8.1 假设矩阵 $A \in \mathbb{R}^{m \times n}$ 是高的 ($m > n$), 具有标准正交列, 则有 A^T 具有标准正交行。

Theorem 10.8.2 A^T 是 A 的一个左逆。

$$A^T A = I$$

Theorem 10.8.3 A 没有右逆。

10.9 值域范围、列空间

Definition 10.9.1 — 向量集合张成的空间。一个向量集合张成的空间是其所有线性组合的集合。

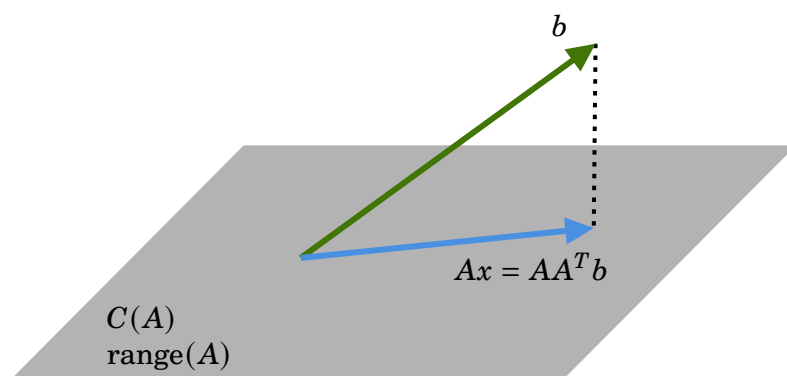
$$\text{span}(a_1, a_2, \dots, a_n) = \{x_1 a_1 + x_2 a_2 + \dots + x_n a_n \mid x \in \mathbb{R}^n\}$$

Definition 10.9.2 — A 的范围 (列空间)。

$$\text{range}(A) = \{Ax \mid x \in \mathbb{R}^n\}$$

10.9.1 投影到列标准正交的矩阵 A 的列空间

Figure 10.3: Projection onto the column space of A , A has orthonormal columns



Problem 10.1 假设矩阵 $A \in \mathbb{R}^{m \times n}$ 具有标准正交列, 求投影 b 在 $C(A)$ 上的投影。
即向量 Ax 与 b 有最短距离

$$\min_x \|Ax - b\|_2^2$$

$$\begin{aligned} f(x) &= \|Ax - b\|_2^2 \\ &= (Ax - b)^T (Ax - b) = x^T A^T Ax - 2x^T A^T b + b^T b \\ &= x^T x - 2x^T A^T b + b^T b \quad (\because A^T A = I) \end{aligned}$$

$$\nabla f(x) = 2x - 2A^T b = 0 \Rightarrow x = A^T b$$

$AA^T b$ 称为向量 $b \in \mathbb{R}^m$ 在 $\text{range}(A)$ 上的正交投影。

Theorem 10.9.1

$$Ax = AA^T b \in \text{range}(A)$$

且是 b 在 $C(A)$ 上的投影。

Proof. 1.

$$Ax = AA^T b \in \text{range}(A)$$

2. 可以证明 $\hat{x} = A^T b$ 满足 $\|A\hat{x} - b\| < \|Ax - b\|$, 对于所有 $x \neq \hat{x}$ 。
 b 到 $\text{range}(A)$ 内任意点 Ax 的距离的平方和为:

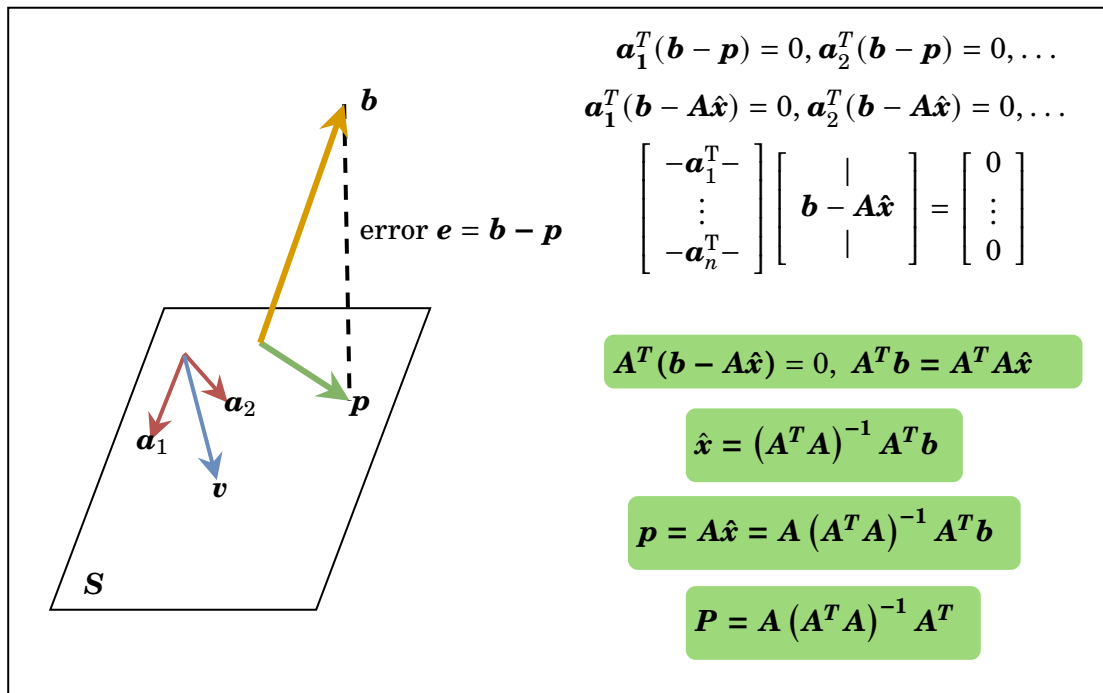
$$\begin{aligned} \|Ax - b\|_2^2 &= \|A(x - \hat{x}) + A\hat{x} - b\|_2^2 \quad \left(\text{其中 } \hat{x} = A^T b \right) \\ &= \|A(x - \hat{x})\|_2^2 + \|A\hat{x} - b\|_2^2 + 2(x - \hat{x})^T A^T (A\hat{x} - b) \\ &= \|A(x - \hat{x})\|_2^2 + \|A\hat{x} - b\|_2^2 \\ &= \|(x - \hat{x})\|_2^2 + \|A\hat{x} - b\|_2^2 \\ &\geq \|A\hat{x} - b\|_2^2 \end{aligned}$$

当且仅当 $x = \hat{x}$, 等号成立。

第 3 行成立是因为 $A^T(A\hat{x} - b) = \hat{x} - A^T b = 0$ 。

■

Figure 10.4: Projection of b into the column space of A , A is any matrix. Sourced from [Strang1993IntroductionTL]



11. QR 分解与 Householder 变换

11.1 Triangular Matrices

Definition 11.1.1 — Lower Triangular Matrices. 矩阵 $A \in \mathbb{R}^{n \times n}$ 为下三角 (*Lower Triangular*) 矩阵, $A_{ij} = 0, j > i$ 。

$$A = \begin{bmatrix} A_{11} & 0 & \cdots & 0 & 0 \\ A_{21} & A_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_{n-1,1} & A_{n-1,2} & \cdots & A_{n-1,n-1} & 0 \\ A_{n1} & A_{n2} & \cdots & A_{n,n-1} & A_{nn} \end{bmatrix}$$

Definition 11.1.2 — Upper Triangular Matrices. A^T 为上三角 (*Upper Triangular*) 矩阵。

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1,n-1} & A_{1,n} \\ 0 & A_{22} & \cdots & A_{2,n-1} & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{n-1,n-1} & A_{n-1,n} \\ 0 & 0 & \cdots & 0 & A_{nn} \end{bmatrix}$$

Definition 11.1.3 — 单位上三角矩阵, 单位下三角矩阵. 对角元素 a_{ii} 都等于 1 的上 (下) 三角矩阵。

11.1.1 高斯消元法

Problem 11.1 当 A 是具有非零对角元素的下三角矩阵时, 解 $Ax = b$ 。

使用前向回代 (Forward Substitution) 算法求解。

时间复杂度: $1 + 3 + 5 + \dots + (2n - 1) = n^2$ flops

Problem 11.2 当 A 是具有非零对角元素的上三角矩阵, 解 $Ax = b$ 。

使用后向回代 (Back Substitution) 算法来求解。

Algorithm 4: Forward Substitution

$$\begin{aligned}
1 \quad x_1 &= \frac{b_1}{A_{11}} \\
2 \quad x_2 &= \frac{b_2 - A_{21}x_1}{A_{22}} \\
3 \quad x_3 &= \frac{b_3 - A_{31}x_1 - A_{32}x_2}{A_{33}} \\
4 \quad \cdots \\
5 \quad x_n &= \frac{b_n - A_{n1}x_1 - A_{n2}x_2 - \cdots - A_{n,n-1}x_{n-1}}{A_{nn}}
\end{aligned}$$

Algorithm 5: Backward Substitution

$$\begin{aligned}
1 \quad x_n &= \frac{b_n}{A_{nn}} \\
2 \quad x_{n-1} &= \frac{b_{n-1} - A_{n-1,n}x_n}{A_{n-1,n-1}} \\
3 \quad x_{n-2} &= \frac{b_{n-2} - A_{n-2,n-1}x_{n-1} - A_{n-2,n}x_n}{A_{n-2,n-2}} \\
4 \quad \cdots \\
5 \quad x_1 &= \frac{b_1 - A_{12}x_2 - A_{13}x_3 - \cdots - A_{1n}x_n}{A_{11}}
\end{aligned}$$

时间复杂度: $1 + 3 + \dots + 2n - 1 = n^2$ flops

11.1.2 The Inverses of Triangular Matrices

Theorem 11.1.1 对角元素非零的三角矩阵 A 是非奇异的, 即:

$$Ax = 0 \quad \Rightarrow \quad x = 0$$

Theorem 11.1.2 — 高斯消元法. A 的逆可以通过逐列解方程 $AX = I$ 来计算得到

$$A \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix}$$

Theorem 11.1.3 下三角矩阵的逆是下三角矩阵, 上三角矩阵的逆是上三角矩阵。

上/下三角矩阵 $A \in \mathbb{R}^{n \times n}$ 逆的复杂度

$$n^2 + (n-1)^2 + \cdots + 1 \approx \frac{1}{3}n^3 \text{ flops}$$

11.2 QR Factorization

如果矩阵 $A \in \mathbb{R}^{m \times n}$ 的列向量线性无关, 则可以将其分解为

Theorem 11.2.1 — QR Factorization.

$$\begin{aligned}
A &= \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \\
&= \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ 0 & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{nn} \end{bmatrix} \\
&= QR
\end{aligned}$$

向量 $q_1, \dots, q_n \in \mathbb{R}^m$ 是标准正交向量:

$$\|q_i\|_2 = 1, \quad q_i^T q_j = 0, \text{ if } i \neq j$$

对角元素 R_{ii} 是非零的。若 $R_{ii} < 0$, 改变 R_{ii}, \dots, R_{in} 和向量 q_i 的符号。大多数定义要求 $R_{ii} > 0$, 使得 Q 和 R 是唯一的。

Corollary 11.2.2 $Q \in \mathbb{R}^{m \times n}$ 具有标准正交列 ($Q^T Q = I$).

Corollary 11.2.3 如果 A 是方阵 ($m = n$), 则 Q 是正交的 ($Q^T Q = Q Q^T = I$).

Corollary 11.2.4 $R \in \mathbb{R}^{n \times n}$ 的上三角矩阵.

Corollary 11.2.5 R 是非奇异的 (对角元素是非零的).

Corollary 11.2.6

$$R = Q^{-1}A \Rightarrow R = Q^T A$$

QR 分解可通过 Gram-Schmidt 正交化法 (参见5.4)、Householder 变换等进行。

Algorithm 6: QR Decomposition Using Gram-Schmidt Algorithm

- 1 设矩阵 A 的列向量依次为 a_1, a_2, \dots, a_n , 由于 A 为非奇异矩阵, 则列向量线性无关
- 2 对列向量 a_1, a_2, \dots, a_n 按照 Gram-Schmidt 方法进行正交化, 然后单位化
- 3 单位化得到的标准正交向量 q_1, q_2, \dots, q_n , 即得到标准正交矩阵 Q
- 4 根据 $R = Q^{-1}A \Rightarrow R = Q^T A$, 得到上三角矩阵 R
- 5 QR 分解 $A = QR$

■ **Example 11.1** 矩阵 A 的 QR 分解过程

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

令 $a_1 = (1, 1, 0)^T, a_2 = (1, -1, 0)^T, a_3 = (0, 1, 2)^T$, 由 Schmidt 方法正交单位化后, 得到 $q_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)^T, q_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)^T, q_3 = (0, 0, 1)^T$ 。

所以 $a_1 = \sqrt{2}q_1$, $a_2 = \sqrt{2}q_2$, $a_3 = \frac{1}{\sqrt{2}}q_1 - \frac{1}{\sqrt{2}}q_2 + 2q_3$ 。

$$A = QR = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & 0 & 2 \end{bmatrix}$$

■ **Example 11.2**

$$\begin{aligned} \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix} &= \begin{bmatrix} -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 8 \\ 0 & 0 & 4 \end{bmatrix} \\ &= \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix} \\ &= QR \end{aligned}$$

Theorem 11.2.7 — 满秩矩阵 A 的 QR 分解的存在唯一性. 设 $A \in \mathbb{C}^{m \times n} (m \geq n)$. 则存在一个单位列正交矩阵 $Q \in \mathbb{C}^{m \times n}$ (即 $Q^H Q = I_{n \times n}$) 和一个上三角矩阵 $R \in \mathbb{C}^{n \times n}$, 使得

$$A = QR$$

若 A 列满秩, 则存在一个具有正对角线元素的上三角矩阵 R 使得上式成立, 且此时 QR 分解唯一, 即 Q 和 R 都唯一.

Proof. 存在性:

由于 A 列满秩, 由 Gram-Schmidt 正交化过程可知, 存在上三角矩阵 $R = [r_{ij}]_{n \times n}$ 满足 $r_{jj} > 0$, 使得 $A = QR$, 其中 Q 单位列正交.

唯一性:

假设 A 存在 QR 分解

$$A = Q_1 R_1 = Q_2 R_2$$

其中 $Q_1, Q_2 \in \mathbb{C}^{m \times n}$ 单位列正交, $R_1, R_2 \in \mathbb{C}^{n \times n}$ 为具有正对角元素的上三角矩阵. 则有

$$Q_1 = Q_2 R_2 R_1^{-1} \quad (11.1)$$

由于 R_1, R_2 均为上三角矩阵, 所以 $R_2 R_1^{-1}$ 也是上三角矩阵, 且其对角线元素为 $R_2(i, i)/R_1(i, i), i = 1, 2, \dots, n$. 由 eq. (11.1) 可得

$$1 = \|Q_1\|_2 = \|Q_2 R_2 R_1^{-1}\|_2 = \|R_2 R_1^{-1}\|_2$$

所以

$$\frac{R_2(i, i)}{R_1(i, i)} \leq 1, \quad i = 1, 2, \dots, n$$

同理可证 $R_1(i, i)/R_2(i, i) \leq 1$. 所以

$$R_1(i, i) = R_2(i, i), \quad i = 1, 2, \dots, n$$

又 $\|Q_1\|_F^2 = \text{tr}(Q_1^T Q_1) = n$, 所以由 eq. (11.1) 可知

$$\|R_2 R_1^{-1}\|_F^2 = \|Q_2 R_2 R_1^{-1}\|_F^2 = \|Q_1\|_F^2 = n$$

由于 $R_2 R_1^{-1}$ 的对角线元素都是 1, 所以 $R_2 R_1^{-1}$ 只能是单位矩阵, 即 $R_2 = R_1$. 因此 $Q_2 = A R_2^{-1} = A R_1^{-1} = Q_1$, 即 A 的 QR 分解是唯一的。 ■

11.2.1 复矩阵的 QR 分解

Theorem 11.2.8 如果 $A \in \mathbb{C}^{m \times n}$ 的列向量是线性无关的, 则可以将其分解为

$$A = QR$$

$Q \in \mathbb{C}^{m \times n}$ 具有正交列。($Q^H Q = I$)
 $R \in \mathbb{C}^{n \times n}$ 具有实非零对角元素的上三角矩阵。

大多数情况下, 会优先选择对角线元素 R_{ii} 为正数。
 如果没有特别说明, 之后默认矩阵 A 都是实数的。

11.3 QR 分解的应用

可用 QR 分解求解以下问题:

- 线性方程
- 最小二乘问题
- 带约束的最小二乘问题

11.3.1 QR 分解和求解线性方程组 $Ax = b$

QR 分解的思想可以用于加快求逆矩阵速度。

Corollary 11.3.1

$$Ax = b \Rightarrow x = A^{-1}b$$

$$QRx = b \Rightarrow x = R^{-1}(Q^T b)$$

Algorithm 8: Solving linear equations via QR factorization

Input: $n \times n$ invertible matrix A

- 1 QR factorization. Compute the QR factorization $A = QR$
- 2 Compute $Q^T b$
- 3 Back substitution. Solve the triangular equation $Rx = Q^T b$ using back substitution

对于普通矩阵使用 QR 分解求解线性方程组, 正交分解需要 $2n^3$ flops, 计算 $Q^T b$ 需要 $2n^2$ flops, 第三步回代求解 $Rx = Q^T b$ 需要 n^2 flops.(总时间复杂度是 $O(n^3)$).

对于稀疏矩阵求解线性方程组, 时间复杂度接近 $\text{nnz}(A)$ 。内存使用和复杂度在很大程度上取决于系数矩阵的稀疏模式。内存使用量通常是 $\text{nnz}(A) + n$ 的适度倍数, $\text{nnz}(A) + n$ 是指定问题数据 A 和 b 所需的标量数量, 通常远小于 $n^2 + n$ (如果 A 和 b 不是稀疏时存储它们所需的标量数)。求解稀疏线性方程的 flop 数通常也更接近 $\text{nnz}(A)$, 而不是 n^3 (矩阵 A 不稀疏时的阶数)。

11.3.2 QR 分解和求解伪逆 A^\dagger 、逆 A^{-1}

Definition 11.3.1 — 线性无关列向量的矩阵 A 的伪逆.

$$A^\dagger = \left(A^T A \right)^{-1} A^T$$

Theorem 11.3.2

$$\begin{aligned} A^\dagger &= \left((QR)^T (QR) \right)^{-1} (QR)^T \\ &= \left(R^T Q^T Q R \right)^{-1} R^T Q^T \\ &= \left(R^T R \right)^{-1} R^T Q^T \quad \left(Q^T Q = I \right) \\ &= R^{-1} R^{-T} R^T Q^T \quad (R \text{ 是非奇异的}) \\ &= R^{-1} Q^T \end{aligned}$$

Corollary 11.3.3 对于方阵非奇异矩阵 A , 其逆为

$$A^{-1} = (QR)^{-1} = R^{-1} Q^T$$

Corollary 11.3.4 对于方阵非奇异矩阵 $A = QR$

$$RA^{-1} = Q^T$$

Proof.

$$\begin{aligned} A &= QR \\ \Rightarrow AA^{-1} &= QRA^{-1} \\ \Rightarrow I &= Q \underbrace{RA^{-1}}_{Q^{-1}} \end{aligned}$$

■

Algorithm 9: Computing the inverse via QR factorization

Input: $n \times n$ invertible matrix A

- 1 QR factorization. Compute the QR factorization $A = QR$
- 2 **for** $i = 1, \dots, n$ **do**
- 3 Solve the triangular equation $Rb_i = \tilde{q}_i$ using back substitution.
- 4 **end**

对于通过 QR 分解求 A^{-1} , QR 分解需要 $2n^3$ flops, n 次回代需要 n^3 flops, 时间复杂度是 $O(3n^3)$.

11.3.3 A 的列空间和 Q 的列空间相同

矩阵 $A \in \mathbb{R}^{m \times n}$ 的值域范围定义为:

$$\text{range}(A) = \{Ax \mid x \in \mathbf{R}^n\}$$

Theorem 11.3.5 假设 A 有线性无关的列向量, 且其 QR 因子为 Q, R , 则 Q 和 A 的值域范围相同 (有相同的列空间)。

即 Q 的列向量是标准正交的, 并且和 A 的列向量张成相同的空间。

Proof.

$$\begin{aligned} y \in \text{range}(A) &\Leftrightarrow y = Ax, x \in \mathbb{R}^n \\ &\Leftrightarrow y = QRx, z = Rx \\ &\Leftrightarrow y = Qz, z \in \mathbb{R}^n \\ &\Leftrightarrow y \in \text{range}(Q) \end{aligned}$$

■

11.3.4 往 A 列空间上的投影也是往 Q 列空间上的投影

结合 $A = QR$ 和 $A^\dagger = R^{-1}Q^T$, 可得:

Theorem 11.3.6

$$AA^\dagger = QRR^{-1}Q^T = QQ^T$$

Theorem 11.3.7

$$R^T R \hat{x} = R^T Q^T b \text{ or } R \hat{x} = Q^T b \text{ or } \hat{x} = R^{-1} Q^T b$$

可以用于求解最小二乘法, 而不用直接求解 $Ax = b$ (当 A 不可逆时)



注意在 AA^\dagger 中乘积的顺序与 $A^\dagger A = I$ 的差异。

$$\begin{aligned} &\min_y \|Qy - x\|_2^2 \\ &\Rightarrow Q^T(Qy - x) = 0 \\ &\Rightarrow Q^T Qy = Q^T x \\ &\Rightarrow y = Q^T x \end{aligned}$$

$QQ^T x$ 是 x 在 Q 值域 (列空间) 上的投影, 也是往 A 的列空间上的投影 (见10.9.1)。

The pseudoinverse A^\dagger is the n by m matrix that makes AA^\dagger and $A^\dagger A$ into projections.



Trying for $AA^{-1} = A^{-1}A = I$, AA^\dagger = projection matrix onto the column space of A (refer to ??)

$A^\dagger A$ = projection matrix onto the row space of A

11.4 QR Algorithm Using Gram-Schmidt Algorithm

Gram-Schmidt QR 算法将逐列计算 Q 和 R 。

Figure 11.1: Projecting onto the column space of A is also projecting onto the column space of Q

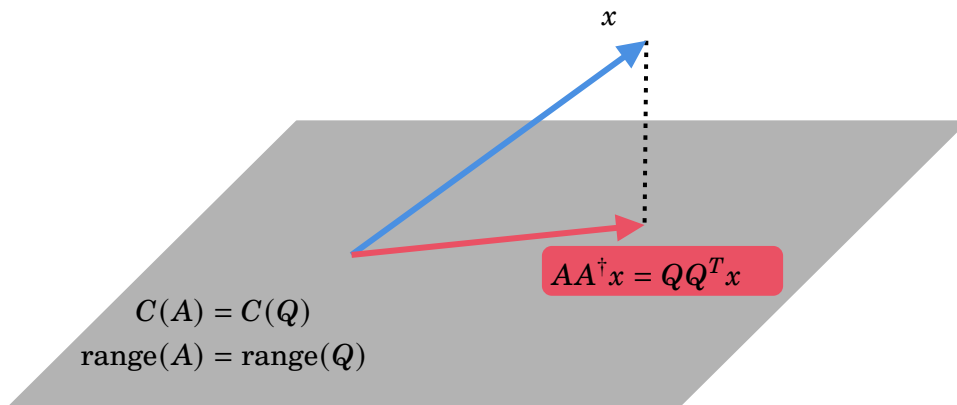
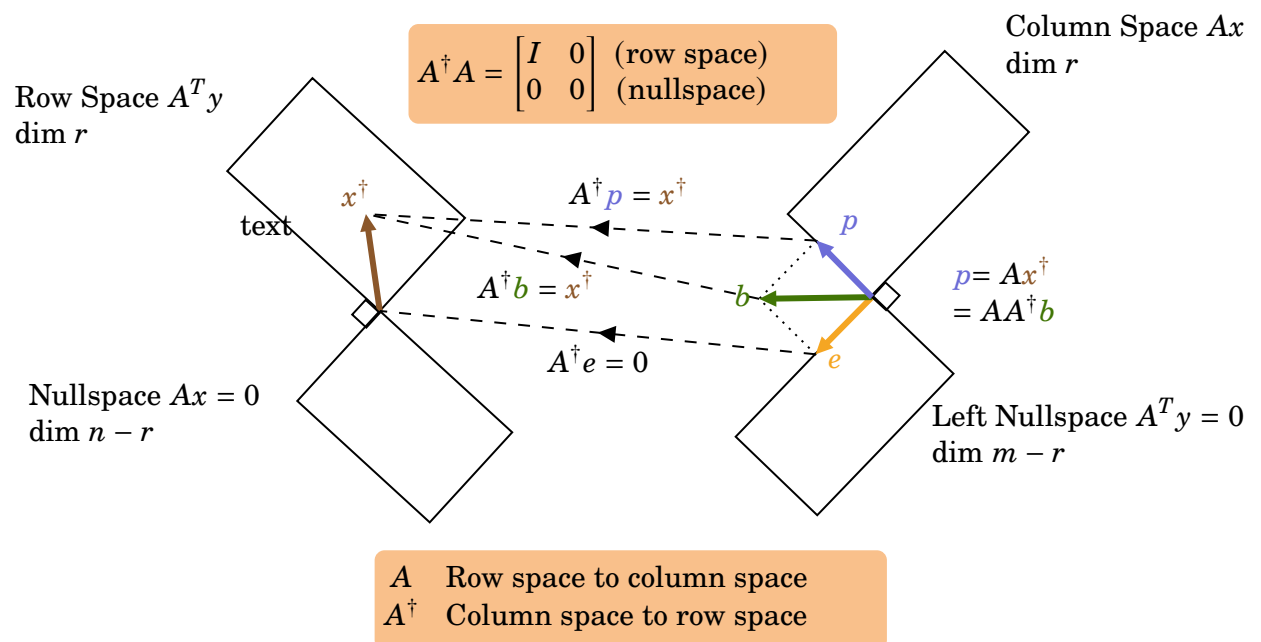


Figure 11.2: Ax^\dagger in the column space goes back to $A^\dagger Ax^\dagger = x^\dagger$ in the row space



k 步后我们得到了 QR 的部分分解:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_k \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1k} \\ 0 & R_{22} & \cdots & R_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{kk} \end{bmatrix}$$

Corollary 11.4.1 QR 的部分分解列向量 q_1, \dots, q_k 是标准正交的。

Corollary 11.4.2 对角线元素 $R_{11}, R_{22}, \dots, R_{kk}$ 是正的。

Corollary 11.4.3 列向量 q_1, \dots, q_k 和 a_1, \dots, a_k 张成的空间相同。

Theorem 11.4.4 $A = QR$ 矩阵中的 R 矩阵为

$$R_{1k} = q_1^T a_k, R_{2k} = q_2^T a_k, \dots, R_{k-1,k} = q_{k-1}^T a_k$$



Gram-Schmidt 正交化的两条基本公式

$$\tilde{q}_i = a_i - (q_1^T a_i) q_1 - \cdots - (q_{i-1}^T a_i) q_{i-1}$$

$$\begin{aligned} a_i &= (q_1^T a_i) q_1 + \cdots + (q_{i-1}^T a_i) q_{i-1} + \underbrace{\|\tilde{q}_i\|_2}_{\tilde{q}_i} q_i \\ &= R_{1i} q_1 + \cdots + R_{ii} q_i \end{aligned}$$

Proof. 假设已经实现 $k-1$ 列的因数分解, 方程 $A = QR$ 的第 k 列可以计算为:

$$a_k = R_{1k} q_1 + R_{2k} q_2 + \cdots + R_{k-1,k} q_{k-1} + R_{kk} q_k$$

无论如何选择 $R_{1k}, \dots, R_{k-1,k}$, 向量

$$\tilde{q}_k = a_k - R_{1k} q_1 - R_{2k} q_2 - \cdots - R_{k-1,k} q_{k-1} \neq 0$$

都将是非零的。(由于 a_1, \dots, a_k 是线性无关的, 即 q_1, \dots, q_{i-1}, q_i 是线性无关的, 假设 Gram-Schmidt 算法过程中没有出现中途退出的状况)

因此

$$a_k \notin \text{span}\{q_1, \dots, q_{k-1}\} = \text{span}\{a_1, \dots, a_{k-1}\}$$

q_k 是 \tilde{q}_k 的单位化: 选择 $R_{kk} = \|\tilde{q}_k\|_2$, 以及 $q_k = \left(\frac{1}{R_{kk}}\right) \tilde{q}_k$ 。

\tilde{q}_k 和 q_k 正交于 q_1, \dots, q_{k-1} , 则 $R_{1k}, \dots, R_{k-1,k}$ 为:

$$R_{1k} = q_1^T a_k, R_{2k} = q_2^T a_k, \dots, R_{k-1,k} = q_{k-1}^T a_k$$

■

Algorithm 10: QR Decomposition Using Gram-Schmidt Algorithm

Input: 矩阵 $A \in \mathbb{R}^{m \times n}$, 列向量 a_1, \dots, a_n 线性无关

```

1  $R_{11} = \|a_1\|_2$ 
2  $q_1 = \frac{1}{R_{11}}a_1$ 
3 for  $k = 2$  to  $n$  do
4    $R_{1k} = q_1^T a_k$ 
5    $R_{2k} = q_2^T a_k$ 
6    $\vdots$ 
7    $R_{k-1,k} = q_{k-1}^T a_k$ 
8    $\tilde{q}_k = a_k - (R_{1k}q_1 + R_{2k}q_2 + \dots + R_{k-1,k}q_{k-1})$ 
9    $R_{kk} = \|\tilde{q}_k\|_2$ 
10   $q_k = \frac{1}{R_{kk}}\tilde{q}_k$ 
11 end
```

11.4.1 Gram-Schmidt Algorithm**■ Example 11.3**

$$\begin{aligned}
 \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} &= \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix} \\
 &= \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}
 \end{aligned}$$

Q 和 R 的第一列:

$$\tilde{q}_1 = a_1 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, R_{11} = \|\tilde{q}_1\| = 2, q_1 = \frac{1}{R_{11}}\tilde{q}_1 = \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix}$$

Q 和 R 的第二列: 计算得到 $R_{12} = q_1^T a_2 = 4$ 。

正交化计算:

$$\tilde{q}_2 = a_2 - R_{12}q_1 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 3 \end{bmatrix} - 4 \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

将其单位化得到:

$$R_{22} = \|\tilde{q}_2\| = 2, \quad q_2 = \frac{1}{R_{22}}\tilde{q}_2 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

Q 和 R 的第三列: 计算得到 $R_{13} = q_1^T a_3 = 2$ 以及 $R_{23} = q_2^T a_3 = 8$ 。

$$\text{正交化计算: } \tilde{q}_3 = a_3 - R_{13}q_1 - R_{23}q_2 = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} - 2 \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} - 8 \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 2 \\ 2 \end{bmatrix}$$

将其单位化得到:

$$R_{33} = \|\tilde{q}_3\| = 4, \quad q_3 = \frac{1}{R_{33}}\tilde{q}_3 = \begin{bmatrix} -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

最终结果:

$$\begin{aligned} \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix} &= \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix} \\ &= \begin{bmatrix} -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 8 \\ 0 & 0 & 4 \end{bmatrix} \end{aligned}$$

■

11.4.2 基于 Gram-Schmidt 方法进行 QR 分解的时间复杂度

Gram-Schmidt 方法第 k 次循环的复杂度:

- a_k 有 $k-1$ 个 $q_i^T a_k$ 内积操作: $(k-1)(2m-1)$ flops
- 计算 \tilde{q}_k : $2(k-1)m$ flops.

$$\tilde{q}_i = a_i - \left(q_1^T a_i\right) q_1 - \cdots - \left(q_{i-1}^T a_i\right) q_{i-1}$$

- 计算 R_{kk} 和 q_k : $3m$ flops。 $R_{kk} = \|\tilde{q}_k\|_2, q_k = \tilde{q}_k / R_{kk}$
- 第 k 次循环的总和: $(4m-1)(k-1) + 3m$ flops
- $A \in \mathbb{R}^{m \times n}$ 分解的复杂度:

$$\begin{aligned} \sum_{k=1}^n ((4m-1)(k-1) + 3m) &= (4m-1) \frac{n(n-1)}{2} + 3mn \\ &\approx 2mn^2 \text{ flops} \end{aligned}$$

对于稀疏矩阵 (空间存储小于 $m \times n$) 的 QR 分解, 时间复杂度可以低于 $O(2mn^2)$ 。

11.5 The Numerical Instability of QR Decomposition based on Gram-Schmidt Algorithm

Gram-Schmidt 算法复杂度为 $2mn^2$ flops。在实际情况中**不推荐使用** (容易被舍入误差影响)。

修正 Gram-Schmidt 算法复杂度为 $2mn^2$ flops, 有更好的数值计算性能。修正之处在于求解投影时每次只减去一个投影。

Householder 算法复杂度为 $2mn^2 - (2/3)n^3$ flops。将 Q 表示为初等正交矩阵的乘积。是最广泛使用的算法 (在 MATLAB 和 JULIA 中的 QR 函数使用该算法)。

本书中认为 QR 分解的复杂度为 $O(2mn^2)$ 。

■ Example 11.4

```

1 [m,n]=size(A);
2 Q = zeros(m,n);
3 R = zeros(n,n);
4 for k = 1:n
5     R(1:k-1,k)=Q(:,1:k-1)'*A(:,k);
6     v= A(:,k)-Q(:,1:k-1)*R(1:k-1,k);
7     R(k,k) = norm(v);
8     Q(:,k)=v/R(k,k);
9 end;
10

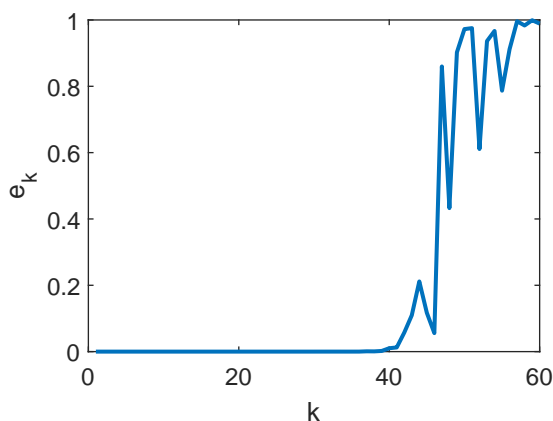
```

Listing 11.1: Gram Schmidt

构造一个矩阵 $A = USV$, 其中 U 和 V 是正交矩阵, S 是对角矩阵

$$S_{ii} = 10^{-10(i-1)/(n-1)}, \quad i = 1, \dots, n$$

把 Gram-Schmidt 算法应用到一个大小为 $m = n = 50$ 的方形矩阵 A 上。

Figure 11.3: $\max_{1 \leq i < k} |q_i^T q_k|$ 

图中显示了 q_k 与前面列之间的正交性的偏差:

$$e_k = \max_{1 \leq i < k} |q_i^T q_k|, \quad k = 2, \dots, n$$

失去正交性是由于浮点数存储的舍入误差。

```

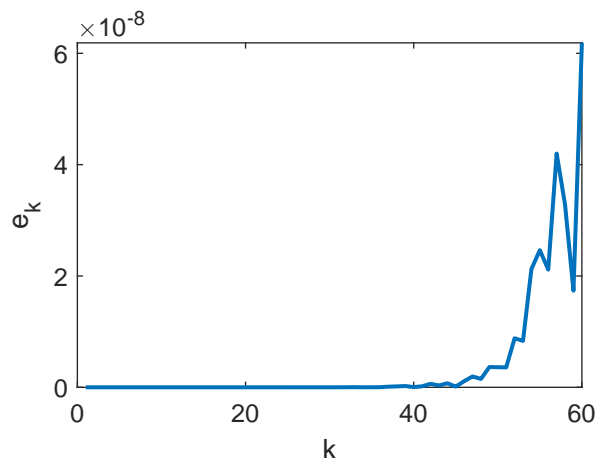
1 for j = 1:n
2     v=A(:,j);
3     for i=1:j-1
4         R(i,j)=Q(:,i)'*v;
5         v=v-R(i,j)*Q(:,i);
6     end
7     R(j,j)= norm(v);
8     Q(:,j)=v/R(j,j);
9 end;
10

```

Listing 11.2: modified Gram-Schmidt

修正 Gram-Schmidt 算法的误差更小。 ■

Figure 11.4: 在同一组数据中使用修正 Gram-Schmidt 算法求得的误差, $e_k = \max_{1 \leq i < k} |q_i^T q_k|, k = 2, \dots, n$



11.6 QR Decomposition Using Householder Transformation

Householder 算法是 QR 分解常用的算法 (MATLAB 和 Julia 中的 qr 函数)。与 Gram-Schmidt 相比, 对舍入误差更有鲁棒性。

Householder 算法计算一个“完整的”QR 因数分解:

$$A = \begin{bmatrix} Q & \tilde{Q} \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix}, \quad \begin{bmatrix} Q & \tilde{Q} \end{bmatrix} \text{ 是正交的矩阵}$$

Proof.

$$\begin{aligned} A &= \begin{bmatrix} Q & \tilde{Q} \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} \\ &= QR + \tilde{Q}0 \\ &= QR \end{aligned}$$

■

完整的 Q 因子被构造成正交矩阵的乘积:

$$\begin{bmatrix} Q & \tilde{Q} \end{bmatrix} = H_1 H_2 \cdots H_n$$

每个 $H_i \in \mathbb{R}^{m \times m}$ 是对称正交的 Householder 矩阵。

11.6.1 Householder Matrix

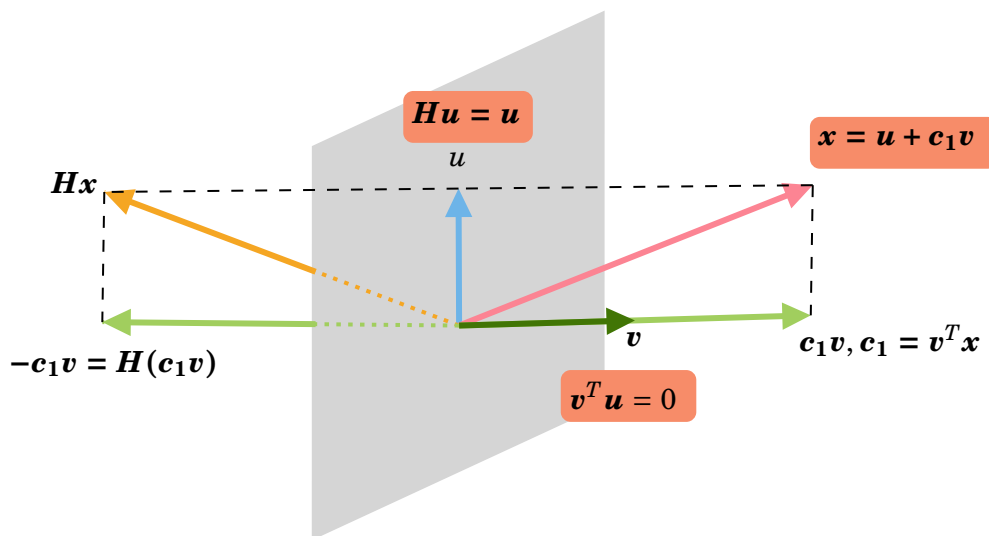
Theorem 11.6.1 $H = I - 2vv^T$, 其中 $\|v\|_2 = 1$, Hx 是 x 关于超平面 $\{u \mid v^T u = 0\}$ 反对称.

H 是对称的

$$H^T = H$$

H 是正交的

$$H^T H = I$$

Figure 11.5: Reflection of x 

Proof.

$$\begin{aligned}
 Hv &= (I - 2vv^T)(c_1v) \\
 &= c_1v - 2vv^T c_1v \\
 &= c_1v - 2c_1vv^T v \\
 &= -c_1v
 \end{aligned}$$

$$\begin{aligned}
 Hu &= (I - 2vv^T)u \\
 &= u - 2vv^T u \\
 &= u \quad (v^T u = 0)
 \end{aligned}$$

$$\begin{aligned}
 Hx &= H(u + c_1v) \\
 &= u - c_1v \\
 &= x - 2(v^T x)v
 \end{aligned}$$

■

Theorem 11.6.2 矩阵向量积 Hx 能化简为

$$Hx = x - 2(v^T x)v$$

Hx 的算法复杂度

如果 v 和 x 的长度是 p , 复杂度是 $4p$ flops。

11.6.2 构造反射算子

给定非零 p 维向量 $y = (y_1, y_2, \dots, y_p)$, 定义

Definition 11.6.1

$$w = \begin{bmatrix} y_1 + \text{sign}(y_1) \|y\|_2 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

$$v = \frac{1}{\|w\|_2} w$$

$\text{sign}(x)$ 是符号函数, $\text{sign}(0) = 0$ 。

Theorem 11.6.3 向量 w 满足

$$\|w\|_2^2 = 2y^T w$$

$$\begin{aligned} \|w\|_2^2 &= w^T w \\ &= 2 \left(\|y\|_2^2 + |y_1| \|y\|_2 \right) \\ &= 2y^T (y + \text{sign}(y_1) \|y\|_2 e_1) \\ &= 2y^T w \end{aligned}$$

Definition 11.6.2 — Constructed Householder Matrix. 将 y 变换为单位基向量 $e = [1, 0, 0, \dots, 0]^T$ 乘以一个常数的 Householder 矩阵为

$$H = I - 2 \frac{ww^T}{\|w\|_2^2}$$

Proof.

$$H = I - 2vv^T = I - 2 \frac{ww^T}{\|w\|_2^2} \quad (v = \frac{1}{\|w\|_2} w)$$

■

Theorem 11.6.4 Householder 矩阵 $H = I - 2vv^T = I - 2 \frac{ww^T}{\|w\|_2^2}$ 将 y 映射为

$$Hy = \begin{bmatrix} -\text{sign}(y_1) \|y\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = r_1$$

Proof.

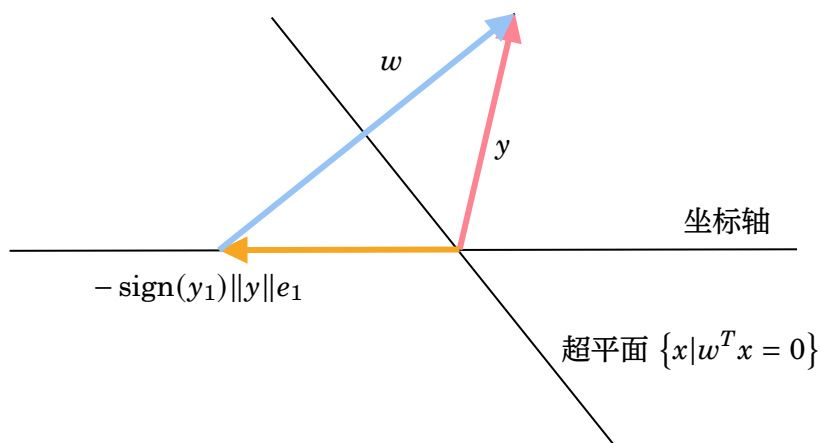
$$\begin{aligned} Hy &= y - \frac{2(w^T y)}{\|w\|_2^2} w \\ &= y - w \\ &= -\text{sign}(y_1) \|y\|_2 e_1 \\ &= \begin{bmatrix} -\text{sign}(y_1) \|y\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{aligned}$$



即 Householder 变换的结果 Hy 只保留第一个元素的值, 其余元素的值变为 0. 又因为 H 是正交、对称的, 它与 QR 分解求解 R 有一定联系。

构造的 Householder 矩阵几何意义

Figure 11.6: 构造的 Householder 矩阵的几何图示



关于超平面 $\{x \mid w^T x = 0\}$, 其法向量:

$$v = \frac{w}{\|w\|_2}, w = y + \text{sign}(y_1) \|y\|_2 e_1$$

反射算子 H 将 y 映射到向量 $-\text{sign}(y_1) \|y\|_2 e_1$ 。

11.6.3 Householder 三角化

计算反射算子 H_1, \dots, H_n 将 A 简化为上三角矩阵形式:

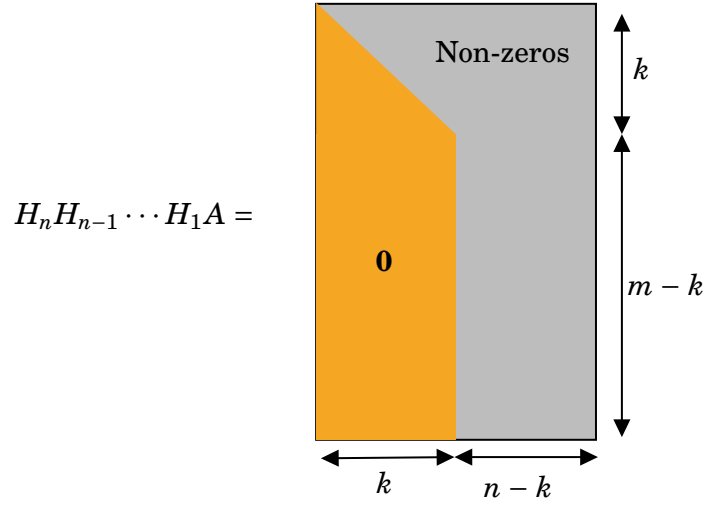
$$H_n H_{n-1} \cdots H_1 A = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

第 k 个步骤之后, 矩阵 $H_k H_{k-1} \cdots H_1 A$ 具有以下结构:

对于 $i > j$ 和 $j \leq k$, 第 i, j 个位置的元素为零。

$$A = \begin{bmatrix} X & X & X & X & X & X & X & X \\ X & X & X & X & X & X & X & X \\ X & X & X & X & X & X & X & X \\ X & X & X & X & X & X & X & X \\ X & X & X & X & X & X & X & X \\ X & X & X & X & X & X & X & X \\ X & X & X & X & X & X & X & X \\ X & X & X & X & X & X & X & X \end{bmatrix}$$

在第一次处理之后, 第一列只剩下第一个元素不为 0.

Figure 11.7: The structure of $H_k H_{k-1} \dots H_1 A$ 

$$H_1 A = A_1 = \begin{bmatrix} X & X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X & X \end{bmatrix}$$

第二次处理对于 $H_1 A_{2:m,1:n}$ ($A_{12:m,1:n}$) 进行处理。

$$H_2 A_1 = A_2 = \begin{bmatrix} X & X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X & X \\ 0 & 0 & X & X & X & X & X & X \\ 0 & 0 & X & X & X & X & X & X \\ 0 & 0 & X & X & X & X & X & X \\ 0 & 0 & X & X & X & X & X & X \\ 0 & 0 & X & X & X & X & X & X \\ 0 & 0 & X & X & X & X & X & X \end{bmatrix}$$

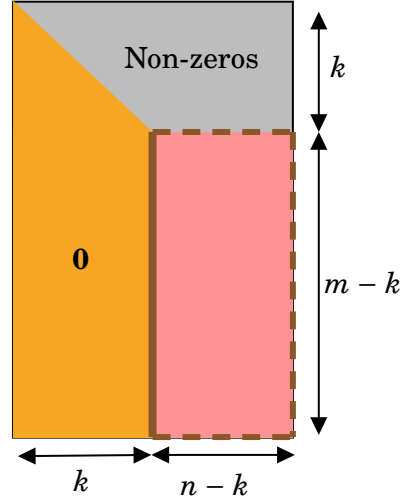
下面的算法用 $\begin{bmatrix} R \\ 0 \end{bmatrix}$ 来代替 $A \in \mathbb{R}^{m \times n}$.

Proof. 在步骤 2 中, 将 $A_{k:m,k:n}$ 与反射算子 $I - 2v_k v_k^T$ 相乘

$$(I - 2v_k v_k^T) A_{k:m,k:n} = A_{k:m,k:n} - 2v_k (v_k^T A_{k:m,k:n})$$

等价于用 $H_k \in \mathbb{R}^{m \times m}$ 乘以 A

$$H_k = \begin{bmatrix} I & 0 \\ 0 & I - 2v_k v_k^T \end{bmatrix} = I - 2 \begin{bmatrix} 0 \\ v_k \end{bmatrix} \begin{bmatrix} 0 & v_k \end{bmatrix}^T$$

Figure 11.8: 迭代计算过程中 A 的结构**Algorithm 11:** QR Decomposition Using Householder Transformation**Input:** Matrix A **Output:** v_1, \dots, v_n (Q), $A = \begin{bmatrix} R \\ 0 \end{bmatrix}$ (R)**1 for** k **in** $1 : n$ **do****2** 令 $y = A_{k:m,k} \in \mathbb{R}^{m-k+1}$, 计算向量 v_k

$$w = y + \text{sign}(y_1) \|y\| e_1$$

$$v_k = \frac{1}{\|w\|} w$$

3 将 $A_{k:m,k:n} \in \mathbb{R}^{(m-k+1) \times (n-k+1)}$ 与反射矩阵 $I - 2v_k v_k^T$ 相乘

$$A_{k:m,k:n} := A_{k:m,k:n} - 2v_k \left(v_k^T A_{k:m,k:n} \right)$$

4 end

算法将下列矩阵来代替 A

$$\begin{bmatrix} R \\ 0 \end{bmatrix}$$

返回向量 v_1, \dots, v_n , 其中 v_k 的长度为 $m - k + 1$ 。 ■

11.6.4 An Example for Householder Algorithm

Problem 11.3

$$A = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix} = H_1 H_2 H_3 \begin{bmatrix} R \\ 0 \end{bmatrix}$$

计算反射算子 H_1, H_2, H_3 来将矩阵 A 三角化

$$H_3 H_2 H_1 A = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \\ 0 & 0 & 0 \end{bmatrix}$$

R 的第一列: 计算将 A 的第一列映射到 e_1 乘积的反射算子

$$y = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad w = y - \|y\|_2 e_1 = \begin{bmatrix} -3 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad v_1 = \frac{1}{\|w\|_2} w = \frac{1}{2\sqrt{3}} \begin{bmatrix} -3 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

用 $I - 2v_1 v_1^T$ 和 A 的乘积代替 A :

$$A := (I - 2v_1 v_1^T) A = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 4/3 & 8/3 \\ 0 & 2/3 & 16/3 \\ 0 & 4/3 & 20/3 \end{bmatrix}$$

R 的第二列: 计算将 $A_{2:4,2}$ 映射到 e_1 乘积的反射算子

$$y = \begin{bmatrix} 4/3 \\ 2/3 \\ 4/3 \end{bmatrix}, \quad w = y + \|y\|_2 e_1 = \begin{bmatrix} 10/3 \\ 2/3 \\ 4/3 \end{bmatrix}, \quad v_2 = \frac{1}{\|w\|_2} w = \frac{1}{\sqrt{30}} \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}$$

用 $I - 2v_2 v_2^T$ 和 $A_{2:4,2:3}$ 的乘积代替 $A_{2:4,2:3}$:

$$A := \begin{bmatrix} 1 & 0 \\ 0 & I - 2v_2 v_2^T \end{bmatrix} A = \begin{bmatrix} 2 & 4 & 2 \\ 0 & -2 & -8 \\ 0 & 0 & 16/5 \\ 0 & 0 & 12/5 \end{bmatrix}$$

R 的第三列: 计算将 $A_{3:4,3}$ 映射到 e_1 乘积的反射算子

$$y = \begin{bmatrix} 16/5 \\ 12/5 \end{bmatrix}, \quad w = y + \|y\|_2 e_1 = \begin{bmatrix} 36/5 \\ 12/5 \end{bmatrix}, \quad v_3 = \frac{1}{\|w\|_2} w = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

用 $I - 2v_3v_3^T$ 和 $A_{3:4,3}$ 的乘积代替 $A_{3:4,3}$:

$$\begin{aligned}
 A &:= \begin{bmatrix} I & 0 \\ 0 & I - 2v_3v_3^T \end{bmatrix} A = \begin{bmatrix} 2 & 4 & 2 \\ 0 & -2 & -8 \\ 0 & 0 & -4 \\ 0 & 0 & 0 \end{bmatrix} \\
 H_3H_2H_1A &= \begin{bmatrix} I_2 & 0 \\ 0 & I_2 - 2v_3v_3^T \end{bmatrix} \begin{bmatrix} I_1 & 0 \\ 0 & I_3 - 2v_2v_2^T \end{bmatrix} (I_4 - 2v_1v_1^T) A \\
 &= \begin{bmatrix} I_2 & 0 \\ 0 & I_2 - 2v_3v_3^T \end{bmatrix} \begin{bmatrix} I_1 & 0 \\ 0 & I_3 - 2v_2v_2^T \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 4/3 & 8/3 \\ 0 & 2/3 & 16/3 \\ 0 & 4/3 & 20/3 \end{bmatrix} \\
 &= \begin{bmatrix} I_2 & 0 \\ 0 & I_2 - 2v_3v_3^T \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & -2 & -8 \\ 0 & 0 & 16/5 \\ 0 & 0 & 12/5 \end{bmatrix} \\
 &= \begin{bmatrix} 2 & 4 & 2 \\ 0 & -2 & -8 \\ 0 & 0 & -4 \\ 0 & 0 & 0 \end{bmatrix}
 \end{aligned}$$

11.6.5 Complexity of Householder Algorithm

Householder 方法第 k 次循环的复杂度:

- $v_k^T A_{k:m,k:n}$ 的乘积: $(2(m-k+1)-1)(n-k+1)$ flops
- v_k 的外积: $(m-k+1)(n-k+1)$ flops
- $A_{k:m,k:n}$ 的减法: $(m-k+1)(n-k+1)$ flops

第 k 次循环的总和: $4(m-k+1)(n-k+1)$ flops

计算 R 和 v_1, \dots, v_n 的总复杂度

$$\begin{aligned}
 \sum_{k=1}^n 4(m-k+1)(n-k+1) &\approx \int_0^n 4(m-t)(n-t+1)dt \\
 &\approx 2mn^2 - \frac{2}{3}n^3 \text{ flops}
 \end{aligned}$$

11.7 Householder 变换进行 QR 分解的 Q 因子

Householder 算法返回向量 v_1, \dots, v_n , 其定义为:

Definition 11.7.1 — v_1, \dots, v_n 的完整表示.

$$[Q \quad \tilde{Q}] = H_1 H_2 \cdots H_n$$

通常不需计算矩阵 $[Q \quad \tilde{Q}]$ 。向量 v_1, \dots, v_n 是 $[Q \quad \tilde{Q}]$ 简单表示 (economical representation)。

Theorem 11.7.1 $\begin{bmatrix} Q & \tilde{Q} \end{bmatrix}$ 或其转置的乘积可以计算为:

$$\begin{aligned} \begin{bmatrix} Q & \tilde{Q} \end{bmatrix} x &= H_1 H_2 \cdots H_n x \\ \begin{bmatrix} Q & \tilde{Q} \end{bmatrix}^T y &= H_n H_{n-1} \cdots H_1 y \end{aligned}$$

Definition 11.7.2 — 矩阵-向量积 $H_k x$. 矩阵-向量积 $H_k x$ 定义为:

$$H_k x = \begin{bmatrix} I_{k-1} & 0 \\ 0 & I - 2v_k v_k^T \end{bmatrix} \begin{bmatrix} x_{1:k-1} \\ x_{k:m} \end{bmatrix} = \begin{bmatrix} x_{1:k-1} \\ x_{k:m} - 2(v_k^T x_{k:m}) v_k \end{bmatrix}$$

11.7.1 矩阵-向量积 $H_k x$ 算法复杂度

$H_k x$ 乘积的复杂度为: $4(m - k + 1)$ flops。

$H_1 H_2, \dots, H_n$ 或其转置的乘积的复杂度为:

$$\sum_{k=1}^n 4(m - k + 1) \approx 4mn - 2n^2 \text{ flops}$$

其复杂度约等于 $m \times n$ 矩阵的矩阵-向量乘积 ($2mn$ flops)。

11.8 Fast Orthogonalization



This section is quoted from [Strang1993IntroductionTL].

There are three ways to reach the important factorization $A = QR$.

Gram-Schmidt works to find the orthonormal vectors in Q . Then R is upper triangular because of the order of Gram-Schmidt steps.

Now we look at better methods (Householder and Givens), which use a product of specially simple Q 's that we know are orthogonal.

Elimination gives $A = LU$, orthogonalization gives $A = QR$. We don't want a triangular L , we want an orthogonal Q . L is a product of E' 's from elimination, with 1's on the diagonal and the multiplier ℓ_{ij} below. **Q will be a product of orthogonal matrices.**

There are two simple orthogonal matrices to take the place of the E 's. The *reflection matrices* $I - 2uu^T$ are named after *Householder*. The *plane rotation matrices* are named after *Givens*. The simple matrix that rotates the xy plane by θ is Q_{21} :

Definition 11.8.1 — Givens Rotation in the 1-2 plane.

$$Q_{21} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Use Q_{21} the way you used E_{21} , to produce a zero in the $(2, 1)$ position. That determines the angle θ . Bill Hager gives this example in Applied Numerical Linear Algebra:

■ **Example 11.5** — 使用 Givens 进行 QR 分解.

$$Q_{21}A = \begin{bmatrix} .6 & .8 & 0 \\ -.8 & .6 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 90 & -153 & 114 \\ 120 & -79 & -223 \\ 200 & -40 & 395 \end{bmatrix} = \begin{bmatrix} 150 & -155 & -110 \\ 0 & 75 & -225 \\ 200 & -40 & 395 \end{bmatrix}$$

The zero came from $-8(90) + .6(120)$. No need to find θ , what we needed was $\cos \theta$

$$\cos \theta = \frac{90}{\sqrt{90^2 + 120^2}} \quad \text{and} \quad \sin \theta = \frac{-120}{\sqrt{90^2 + 120^2}}$$

Now we attack the $(3, 1)$ entry. The rotation will be in rows and columns 3 and 1. The numbers $\cos \theta$ and $\sin \theta$ are determined from 150 and 200, instead of 90 and 120.

$$Q_{31}Q_{21}A = \begin{bmatrix} .6 & 0 & .8 \\ 0 & 1 & 0 \\ -.8 & 0 & .6 \end{bmatrix} \begin{bmatrix} 150 & \cdot & \cdot \\ 0 & \cdot & \cdot \\ 200 & \cdot & \cdot \end{bmatrix} = \begin{bmatrix} 250 & -125 & 250 \\ 0 & 75 & -225 \\ 0 & 100 & 325 \end{bmatrix}$$

One more step to R . The $(3, 2)$ entry has to go. The numbers $\cos \theta$ and $\sin \theta$ now come from 75 and 100. The rotation is now in rows and columns 2 and 3:

$$Q_{32}Q_{31}Q_{21}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & .6 & .8 \\ 0 & -.8 & .6 \end{bmatrix} \begin{bmatrix} 250 & -125 & \cdot \\ 0 & 75 & \cdot \\ 0 & 100 & \cdot \end{bmatrix} = \begin{bmatrix} 250 & -125 & 250 \\ 0 & 125 & 125 \\ 0 & 0 & 375 \end{bmatrix}$$

We have reached the upper triangular R . What is Q ? Move the plane rotations Q_{ij} to the other side to find $A = QR$ —just as you moved the elimination matrices E_{ij} to the other side to find $A = LU$:

Theorem 11.8.1

$$Q_{32}Q_{31}Q_{21}A = R \quad \text{means} \quad A = \left(Q_{21}^{-1}Q_{31}^{-1}Q_{32}^{-1}\right)R = QR$$

The inverse of each Q_{ij} is Q_{ij}^T (rotation through $-\theta$). The inverse of E_{ij} was not an orthogonal matrix! **LU and QR are similar but L and Q are not the same.** ■

Householder reflections are faster than rotations because each one clears out a whole column below the diagonal. Watch how the first column a_1 of A becomes column r_1 of R :

■ **Example 11.6** — Reflection by H_1 .

$$H_1 = I - 2u_1u_1^T$$

$$H_1a_1 = \begin{bmatrix} \|a_1\| \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} -\|a_1\| \\ 0 \\ \vdots \\ 0 \end{bmatrix} = r_1$$

The length was not changed, and u_1 is in the direction of $a_1 - r_1$. We have $n - 1$ entries in the unit vector u_1 to get $n - 1$ zeros in r_1 . (Rotations had one angle θ to get one zero.) When we reach column k , we have $n - k$ available choices in the unit vector u_k . This leads to $n - k$ zeros in r_k . We just store the u 's and r 's to know the final Q and R :

Theorem 11.8.2 — H 的逆是它本身.

$$(H_{n-1} \dots H_1) A = R \quad \text{means} \quad A = (H_1 \dots H_{n-1}) R = QR$$

■



Extensive Reading

12	Fourier Series, Fourier Transform	96
12.1	基本概念	
12.2	Fourier Series	
12.3	Fourier transform	
12.4	Discrete Fourier Transform	
13	Factorization of Matrices	104
13.1	主要的矩阵分解	
13.2	$A = LU$	
14	List Of Definitions	106

12. Fourier Series, Fourier Transform

12.1 基本概念

Definition 12.1.1 — Complex Numbers.

$$C = R + jI$$

where R and I are real numbers and $j = \sqrt{-1}$. Here, R denotes the real part of the complex number and I its imaginary part.

Real numbers are a subset of complex numbers in which $I = 0$.

Definition 12.1.2 — Complex Number in Polar Coordinates.

$$C = |C|(\cos \theta + j \sin \theta) = |C|e^{j\theta}$$

where $|C| = \sqrt{R^2 + I^2}$ is the length of the vector extending from the origin of the complex plane to point (R, I) , and θ is the angle between the vector and the real axis.

上式使用了欧拉公式

Theorem 12.1.1 — Euler's Formular.

$$e^{j\theta} = \cos \theta + j \sin \theta$$

Corollary 12.1.2 $e^{2\pi it}$ 可以表示每秒 1 圈的旋转.

Definition 12.1.3 — 正弦函数.

$$y = A \sin(\omega t + \varphi)$$

就是一个以 $\frac{2\pi}{\omega}$ 为周期的正弦函数, 其中 y 表示动点的位置, t 表示时间, A 为振幅, ω 为角频率, φ 为初相.

如何深入研究非正弦周期函数呢? 将周期函数展开成由简单的周期函数例如三角函数组成的级数. 具体地说, 将周期为 $T (= \frac{2\pi}{\omega})$ 的周期函数用一系列以 T 为周期的正弦函数 $A_n \sin(n\omega t + \varphi_n)$ 组成的级数来表示, 记为

$$f(t) = A_0 + \sum_{n=1}^{\infty} A_n \sin(n\omega t + \varphi_n)$$

其中 $A_0, A_n, \varphi_n (n = 1, 2, 3, \dots)$ 都是常数.

将周期函数按上述方式展开, 它的物理意义是很明确的, 这就是把一个比较复杂的周期运动看成是许多不同频率的简谐振动的叠加. 在电工学上, 这种展开称为谐波分析, 其中常数项 A_0 称为 $f(t)$ 的直流分量, $A_1 \sin(\omega t + \varphi_1)$ 称为一次谐波 (又叫做基波), $A_2 \sin(2\omega t + \varphi_2), A_3 \sin(3\omega t + \varphi_3), \dots$ 依次称为二次谐波, 三次谐波, 等等.

Definition 12.1.4 — 三角级数. 令 $\frac{a_0}{2} = A_0, a_n = A_n \sin \varphi_n, b_n = A_n \cos \varphi_n, \omega = \frac{\pi}{l}$ (即 $T = 2l$), 则 $f(t) = A_0 + \sum_{n=1}^{\infty} A_n \sin(n\omega t + \varphi_n)$ 可以改写为

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi t}{l} + b_n \sin \frac{n\pi t}{l} \right)$$

其中 $a_0, a_n, b_n (n = 1, 2, 3, \dots)$ 都是常数.

Definition 12.1.5 — 三角函数系.

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \dots, \cos nx, \sin nx, \dots$$

在区间 $[-\pi, \pi]$ 上正交, 就是指在三角函数系 (7-4) 中任何不同的两个函数的乘积在区间 $[-\pi, \pi]$ 上的积分等于零, 即

$$\begin{aligned} \int_{-\pi}^{\pi} \cos nx \, dx &= 0 \quad (n = 1, 2, 3, \dots), \\ \int_{-\pi}^{\pi} \sin nx \, dx &= 0 \quad (n = 1, 2, 3, \dots), \\ &\dots \end{aligned}$$

Definition 12.1.6 — 复数域的 Gram 矩阵. 如果 $A \in \mathbb{C}^{m \times n}$ 的 Gram 矩阵为单位矩阵, 则 A 具有正交列:

$$\begin{aligned} A^H A &= \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^H \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \\ &= \begin{bmatrix} a_1^H a_1 & a_1^H a_2 & \cdots & a_1^H a_n \\ a_2^H a_1 & a_2^H a_2 & \cdots & a_2^H a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^H a_1 & a_n^H a_2 & \cdots & a_n^H a_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \\ &= I \end{aligned}$$

Gram 矩阵列有单位范数: $\|a_i\|_2^2 = a_i^H a_i = 1$ 。

Gram 矩阵列是相互正交的: 对于 $i \neq j, a_i^H a_j = 0$ 。

Definition 12.1.7 — 酉矩阵. 列正交的方形复数矩阵称为酉矩阵。

Definition 12.1.8 — 酉矩阵的逆.

$$\left. \begin{array}{l} A^H A = I \\ A \text{ 是方的} \end{array} \right\} \Rightarrow A A^H = I$$

酉矩阵是具有逆 A^H 的非奇异矩阵。如果 A 是酉矩阵, 那么 A^H 也是酉矩阵。

12.2 Fourier Series

Definition 12.2.1 — 傅里叶级数.

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx).$$

where

$$\begin{cases} a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx & (n = 0, 1, 2, 3, \dots), \\ b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx & (n = 1, 2, 3, \dots). \end{cases}$$

Proof. 设 $f(x)$ 是周期为 2π 的周期函数, 且能展开成三角级数

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx).$$

先求 a_0 . 上式一从 $-\pi$ 到 π 积分, 由于假设式右端级数可逐项积分, 因此有

$$\int_{-\pi}^{\pi} f(x) dx = \int_{-\pi}^{\pi} \frac{a_0}{2} dx + \sum_{k=1}^{\infty} \left[a_k \int_{-\pi}^{\pi} \cos kx \, dx + b_k \int_{-\pi}^{\pi} \sin kx \, dx \right]$$

根据三角函数系的正交性, 等式右端除第一项外, 其余各项均为零, 所以

$$\int_{-\pi}^{\pi} f(x) dx = \frac{a_0}{2} \cdot 2\pi$$

于是得

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx$$

其次求 a_n . 用 $\cos nx$ 乘式两端, 再从 $-\pi$ 到 π 积分, 得到

$$\begin{aligned} & \int_{-\pi}^{\pi} f(x) \cos nx \, dx \\ &= \frac{a_0}{2} \int_{-\pi}^{\pi} \cos nx \, dx + \sum_{k=1}^{\infty} \left[a_k \int_{-\pi}^{\pi} \cos kx \cos nx \, dx + b_k \int_{-\pi}^{\pi} \sin kx \cos nx \, dx \right] \end{aligned}$$

根据三角函数系的正交性, 等式右端除 $k = n$ 的一项外, 其余各项均为 0. 所以

$$\int_{-\pi}^{\pi} f(x) \cos nx \, dx = a_n \int_{-\pi}^{\pi} \cos^2 nx \, dx = a_n \pi$$

于是得

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx \quad (n = 1, 2, 3, \dots)$$

类似地, 用 $\sin nx$ 乘 (7-5) 式的两端, 再从 $-\pi$ 到 π 积分, 可得

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx \quad (n = 1, 2, 3, \dots)$$

■

Theorem 12.2.1 — 收敛定理, 狄利克雷 (Dirichlet) 充分条件. 设 $f(x)$ 是周期为 2π 的周期函数, 如果它满足:

1. 在一个周期内连续或只有有限个第一类间断点,
2. 在一个周期内至多只有有限个极值点

那么 $f(x)$ 的傅里叶级数收敛, 并且

- 当 x 是 $f(x)$ 的连续点时, 级数收敛于 $f(x)$;
- 当 x 是 $f(x)$ 的间断点时, 级数收敛于 $\frac{1}{2} [f(x^-) + f(x^+)]$.

收敛定理告诉我们: 只要函数在 $[-\pi, \pi]$ 上至多有有限个第一类间断点, 并且不作无限次振动, 函数的傅里叶级数在连续点处就收敛于该点的函数值, 在间断点处收敛于该点左极限与右极限的算术平均值. 可见, 函数展开成傅里叶级数的条件比展开成幂级数的条件低得多.

Corollary 12.2.2 记 $C = \{x \mid f(x) = \frac{1}{2} [f(x^-) + f(x^+)]\}$
在 C 上就成立 $f(x)$ 的傅里叶级数展开式

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx), x \in C$$

Definition 12.2.2 — 正弦级数. 当 $f(x)$ 为奇函数时, $f(x) \cos nx$ 是奇函数, $f(x) \sin nx$ 是偶函数, 故

$$\left. \begin{aligned} a_n &= 0 \quad (n = 0, 1, 2, \dots), \\ b_n &= \frac{2}{\pi} \int_0^{\pi} f(x) \sin nx \, dx \quad (n = 1, 2, 3, \dots). \end{aligned} \right\}$$

即知奇函数的傅里叶级数是只含有正弦项的正弦级数.

$$\sum_{n=1}^{\infty} b_n \sin nx$$

Definition 12.2.3 — 余弦级数. 当 $f(x)$ 为偶函数时, $f(x) \cos nx$ 是偶函数, $f(x) \sin nx$ 是奇函数, 故

$$\left. \begin{aligned} a_n &= \frac{2}{\pi} \int_0^{\pi} f(x) \cos nx \, dx \quad (n = 0, 1, 2, \dots) \\ b_n &= 0 \quad (n = 1, 2, 3, \dots) \end{aligned} \right\}$$

即知偶函数的傅里叶级数是只含常数项和余弦项的余弦级数

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx$$

在实际应用 (如研究某种波动问题, 热的传导、扩散问题) 中, 有时还需要把定义在区间 $[0, \pi]$ 上的函数 $f(x)$ 展开成正弦级数或余弦级数.

根据前面讨论的结果, 这类展开问题可以按如下的方法解决: 设函数 $f(x)$ 定义在区间 $[0, \pi]$ 上并且满足收敛定理的条件, 我们在开区间 $(-\pi, 0)$ 内补充函数 $f(x)$ 的定义, 得到定义在 $(-\pi, \pi]$ 上的函数 $F(x)$, 使它在 $(-\pi, \pi)$ 上成为奇函数 (偶函数). 按这种方式推广函数定义域的过程称为奇延拓 (偶延拓). 然后将奇延拓 (偶延拓) 后的函数展开成傅里叶级数, 这个级数必定是正弦级数 (余弦级数). 再限制 x 在 $(0, \pi]$ 上, 此时 $F(x) \equiv f(x)$, 这样便得到 $f(x)$ 的正弦级数 (余弦级数) 展开式.

Theorem 12.2.3 设周期为 $2l$ 的周期函数 $f(x)$ 满足收敛定理的条件, 则它的傅里叶级数展开式为

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{l} + b_n \sin \frac{n\pi x}{l} \right) \quad (x \in C), \quad (8-1)$$

其中

$$\left. \begin{aligned} a_n &= \frac{1}{l} \int_{-l}^l f(x) \cos \frac{n\pi x}{l} dx \quad (n = 0, 1, 2, \dots) \\ b_n &= \frac{1}{l} \int_{-l}^l f(x) \sin \frac{n\pi x}{l} dx \quad (n = 1, 2, 3, \dots), \\ C &= \{x \mid f(x) = \frac{1}{2} [f(x^-) + f(x^+)]\} \end{aligned} \right\}$$

当 $f(x)$ 为奇函数时,

$$f(x) = \sum_{n=1}^{\infty} b_n \sin \frac{n\pi x}{l} \quad (x \in C)$$

其中

$$b_n = \frac{2}{l} \int_0^l f(x) \sin \frac{n\pi x}{l} dx \quad (n = 1, 2, 3, \dots)$$

当 $f(x)$ 为偶函数时,

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi x}{l} \quad (x \in C)$$

其中

$$a_n = \frac{2}{l} \int_0^l f(x) \cos \frac{n\pi x}{l} dx \quad (n = 0, 1, 2, \dots)$$

Definition 12.2.4 — 傅里叶级数的复数形式.

$$\sum_{n=-\infty}^{\infty} c_n e^{\frac{n\pi x}{l} j}$$

$$\text{where } c_n = \frac{1}{2l} \int_{-l}^l f(x) e^{-\frac{n\pi x}{l} j} dx \quad (n = 0, \pm 1, \pm 2, \dots).$$

Proof. 设周期为 $2l$ 的周期函数 $f(x)$ 的傅里叶级数为

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{l} + b_n \sin \frac{n\pi x}{l} \right)$$

其中系数 a_n 与 b_n 为

$$\left. \begin{aligned} a_n &= \frac{1}{l} \int_{-l}^l f(x) \cos \frac{n\pi x}{l} dx \quad (n = 0, 1, 2, \dots), \\ b_n &= \frac{1}{l} \int_{-l}^l f(x) \sin \frac{n\pi x}{l} dx \quad (n = 1, 2, 3, \dots). \end{aligned} \right\}$$

利用欧拉公式

$$\cos t = \frac{e^{ti} + e^{-ti}}{2}, \quad \sin t = \frac{e^{ti} - e^{-ti}}{2i}$$

记

$$\frac{a_0}{2} = c_0, \quad \frac{a_n - b_n i}{2} = c_n, \quad \frac{a_n + b_n i}{2} = c_{-n} \quad (n = 1, 2, 3, \dots)$$

则表示为

$$c_0 + \sum_{n=1}^{\infty} (c_n e^{\frac{n\pi x}{l} i} + c_{-n} e^{-\frac{n\pi x}{l} i}) = (c_n e^{\frac{n\pi x}{l} i})_{n=0} + \sum_{n=1}^{\infty} (c_n e^{\frac{n\pi x}{l} i} + c_{-n} e^{-\frac{n\pi x}{l} i})$$

$$c_0 = \frac{a_0}{2} = \frac{1}{2l} \int_{-l}^l f(x) dx$$

$$\begin{aligned} c_n &= \frac{a_n - b_n i}{2} \\ &= \frac{1}{2} \left[\frac{1}{l} \int_{-l}^l f(x) \cos \frac{n\pi x}{l} dx - \frac{i}{l} \int_{-l}^l f(x) \sin \frac{n\pi x}{l} dx \right] \\ &= \frac{1}{2l} \int_{-l}^l f(x) \left(\cos \frac{n\pi x}{l} - i \sin \frac{n\pi x}{l} \right) dx \\ &= \frac{1}{2l} \int_{-l}^l f(x) e^{-\frac{n\pi x}{l} i} dx \quad (n = 1, 2, 3, \dots); \end{aligned}$$

$$c_{-n} = \frac{a_n + b_n i}{2} = \frac{1}{2l} \int_{-l}^l f(x) e^{\frac{n\pi x}{l} i} dx \quad (n = 1, 2, 3, \dots)$$

将已得的结果合并写为

$$c_n = \frac{1}{2l} \int_{-l}^l f(x) e^{-\frac{n\pi x}{l} i} dx \quad (n = 0, \pm 1, \pm 2, \dots)$$

■

12.3 Fourier transform

Definition 12.3.1 — The Fourier transform of a continuous function $f(t)$ of a continuous variable t .

$$F(\mu) = \mathfrak{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t) e^{-j2\pi\mu t} dt$$

Using Euler's formula, we can write as

$$F(\mu) = \int_{-\infty}^{\infty} f(t) [\cos(2\pi\mu t) - j \sin(2\pi\mu t)] dt$$

Definition 12.3.2 — inverse Fourier transform.

$$f(t) = \int_{-\infty}^{\infty} F(\mu) e^{j2\pi\mu t} d\mu$$

12.4 Discrete Fourier Transform

Definition 12.4.1 — Discrete Fourier Transform.

$$F_m = \sum_{n=0}^{M-1} f_n e^{-j2\pi mn/M} \quad m = 0, 1, 2, \dots, M-1$$

Definition 12.4.2 — inverse discrete Fourier transform (IDFT).

$$f_n = \frac{1}{M} \sum_{m=0}^{M-1} F_m e^{j2\pi mn/M} \quad n = 0, 1, 2, \dots, M-1$$

Definition 12.4.3 — 离散傅里叶变换矩阵 W .

$$W = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \dots & \omega^{-(n-1)} \\ 1 & \omega^{-2} & \omega^{-4} & \dots & \omega^{-2(n-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \omega^{-(n-1)} & \omega^{-2(n-1)} & \dots & \omega^{-(n-1)(n-1)} \end{bmatrix}$$

where $\omega = e^{2\pi j/n}$, $j = \sqrt{-1}$

Corollary 12.4.1 矩阵 $(1/\sqrt{n})W$ 是酉矩阵:

$$\frac{1}{n} W^H W = \frac{1}{n} W W^H = I$$

Corollary 12.4.2 W 的逆 $W^{-1} = (1/n)W^H$ 。

Corollary 12.4.3 W 的共轭转置为

$$W^H = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega^1 & \omega^2 & \dots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \dots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

Corollary 12.4.4 Gram 矩阵的第 i, j 个元素为

$$(W^H W)_{ij} = 1 + \omega^{i-j} + \omega^{2(i-j)} + \dots + \omega^{(n-1)(i-j)}$$

$$(W^H W)_{ii} = n, (W^H W)_{ij} = \frac{\omega^{n(i-j)} - 1}{\omega^{i-j} - 1} = 0 \text{ 如果 } i \neq j$$

最后一步因为 $\omega^n = 1$ 。

Definition 12.4.4 — 离散傅里叶反变换. n 维向量 x 的离散傅里叶反变换是

$$W^{-1}x = (1/n)W^H x$$

13. Factorization of Matrices

13.1 主要的矩阵分解

$$A = LU$$

$$A = LPU$$

$$A = QR$$

$$A = X\Lambda X^{-1}$$

$$S = Q\Lambda Q^T$$

$$A = U\Sigma V^T$$

$$A = CMR$$

13.2 $A = LU$

$$(E_{32}E_{31}E_{21})A = U \text{ becomes } A = \begin{pmatrix} E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} \end{pmatrix} U \quad \text{which is} \quad A = LU$$

高斯消元法将 A 变成上三角矩阵，等价于左乘以一系列的初等矩阵。

List of Figures

1.1	The translation from q to p	10
4.1	Projection onto a line	27
4.2	Projection onto a line from the perspective of linear algebra.	28
6.1	Row Picture and Column Picture for 6.1	40
6.2	An example of convolution	45
9.1	Four Subspace of Matrix A	57
9.2	$A\mathbf{x}^\dagger$ in the column space goes back to $A^\dagger A\mathbf{x}^\dagger = \mathbf{x}^\dagger$ in the row space	62
9.3	Projection from row space to column space	63
10.1	An example of rotation	67
10.2	Elementary Reflection	68
10.3	Projection onto the column space of A , A has orthonormal columns	70
10.4	Projection of \mathbf{b} into the column space of A , A is any matrix. Sourced from [Strang1993IntroductionTL	71
11.1	Projecting onto the column space of A is also projecting onto the column space of Q	79
11.2	$A\mathbf{x}^\dagger$ in the column space goes back to $A^\dagger A\mathbf{x}^\dagger = \mathbf{x}^\dagger$ in the row space	79
11.3	$\max_{1 \leq i < k} q_i^T q_k $	83
11.4	在同一组数据中使用修正 Gram-Schmidt 算法求得的误差, $e_k = \max_{1 \leq i < k} q_i^T q_k $, $k = 2, \dots, n$	84
11.5	Reflection of x	85
11.6	构造的 Householder 矩阵的几何图示	87
11.7	The structure of $H_k H_{k-1} \dots H_1 A$	88
11.8	迭代计算过程中 A 的结构	89

14. List Of Definitions

Chapter 1 对向量的介绍	8
1.1.1 (Definition) Vector	8
1.1.2 (Definition) n 维向量 a 的第 i 元素	8
1.1.3 (Definition) $a = b$	8
1.1.4 (Definition) stacked vector	8
1.1.5 (Definition) 零向量	8
1.1.6 (Definition) 全一向量	9
1.1.7 (Definition) 单位向量	9
1.1.8 (Definition) 稀疏向量	9
1.2.1 (Definition) 向量空间 V	9
1.2.1 (Corollary)	9
1.2.2 (Corollary)	9
1.2.2 (Definition) 数域	9
1.3.1 (Definition) 向量加法	9
1.3.1 (Corollary) 向量位移相加	10
1.3.2 (Definition) 标量与向量的乘法	10
1.3.3 (Definition) 线性组合	10
1.4.1 (Definition) 内积	10
1.4.1 (Corollary) 选出第 i 项	11
1.4.2 (Corollary) 向量每一项之和	11
1.4.3 (Corollary) 向量每一项的平方和	11
1.4.4 (Corollary) 向量元素的平均值	11
1.4.5 (Corollary) Selective sum	11

1.4.2	(Definition) The sum of block vectors	11
1.5.1	(Theorem) Cauchy-Schwartz Inequality	12
1.5.2	(Theorem)	13
1.6.1	(Corollary)	13
Chapter 2 Linear Function		14
2.1.1	(Definition) Linear Function	14
2.1.1	(Theorem)	14
2.1.2	(Definition) 内积函数 (inner product function)	15
2.1.2	(Corollary)	15
2.1.3	(Corollary)	15
2.1.3	(Definition) 仿射函数 (affine function)	15
2.1.4	(Theorem)	15
2.2.1	(Definition) 函数 f 第 i 个分量的一阶偏导数	15
2.2.2	(Definition) f 在点 z 的梯度	16
2.2.3	(Definition) Taylor's Approximation	16
2.2.1	(Corollary) n 阶泰勒多项式	16
2.2.2	(Corollary) 对于高阶余项的公式	17
2.2.3	(Corollary) 麦克劳林 (Maclaurin) 公式	17
2.2.4	(Definition) 一阶泰勒公式	17
2.2.4	(Corollary) 一阶泰勒公式的内积形式	17
2.3.1	(Definition) Regression Model	18
Chapter 3 Norm and Distance		19
3.1.1	(Definition) Vector Norm	19
3.1.1	(Corollary) 柯西—施瓦茨不等式	20
3.1.2	(Definition) ℓ_∞ -范数	20
3.1.3	(Definition) ℓ_p -范数	20
3.1.2	(Theorem) Minkowshi Inequality	20
3.1.3	(Theorem) Hölder Inequality	20
3.2.1	(Definition) 向量 x 的均方值 (mean-square value)	21
3.2.2	(Definition) n 维向量 x 的均方根 (root-mean-square value, RMS)	21
3.3.1	(Theorem) Chebyshev's Inequality	21
3.3.2	(Theorem) Chebyshev's Inequality	21
3.3.3	(Corollary) Chebyshev's Inequality Using RMS	21
3.4.1	(Definition) Euclidean distance	21
3.4.2	(Definition) RMS deviation	21
3.4.1	(Theorem) Trianglar Inequality	22
3.4.3	(Definition) Feature Distance	22
3.4.4	(Definition) Nearest Neighbor	22
3.5.1	(Definition) 算术平均值	22

3.5.2	(Definition) De-meaned Vector	22
3.5.3	(Definition) x 的标准差	22
3.5.1	(Theorem)	22
3.6.1	(Definition) 两个非零向量 a 和 b 之间的角 (angle)	22
3.6.2	(Definition) 球面的距离	23
3.6.3	(Definition) a 和 b 的相关系数	23
4.1	(Problem)	24
Chapter 4 优化问题初步		24
4.1.1	(Definition) 高阶无穷小记号 o	24
4.1.1	(Corollary)	24
4.2.1	(Definition) 凸集	25
4.2.2	(Definition) 凸函数	25
4.2.1	(Theorem) 可微函数 f 是凸函数的充要条件	26
4.2.2	(Theorem)	26
4.3.1	(Definition) 向量对向量的导数	26
4.2	(Problem)	27
4.3	(Problem)	28
Chapter 5 Linear Independence		30
5.1.1	(Definition) 线性相关 (linearly dependent)	30
5.1.1	(Corollary)	30
5.1.2	(Corollary)	30
5.1.2	(Definition) 线性独立 (linearly independent)	30
5.1.3	(Corollary)	30
5.1.4	(Theorem)	31
5.2.1	(Definition) 基 (Basis)	31
5.2.2	(Definition) 向量 b 在基底 a_1, \dots, a_n 下的分解	31
5.3.1	(Definition) Orthogonal Vectors	31
5.3.2	(Definition) Orthonormal Vectors	32
5.3.1	(Corollary)	32
5.3.2	(Corollary)	32
5.3.3	(Definition) n 维向量的一个标准正交基	33
5.3.4	(Definition) x 在标准正交基下的标准正交分解	33
5.4.1	(Theorem)	33
5.4.2	(Corollary)	34
5.4.3	(Corollary)	34
5.4.4	(Corollary)	34
5.4.5	(Corollary)	35
5.4.6	(Corollary)	35

Chapter 6	Matrices	37
6.1.1	(Definition) 矩阵	37
6.1.2	(Definition) 标量	37
6.1.3	(Definition) 向量	37
6.1.4	(Definition) 行向量, 列向量	37
6.1.5	(Definition) 高形, 宽形和方形矩阵	37
6.1.6	(Definition) 分块矩阵	37
6.1.7	(Definition) 矩阵的列向量表示	38
6.1.8	(Definition) 矩阵的行向量表示	38
6.2.1	(Definition) 矩阵数乘	38
6.2.2	(Definition) 矩阵加法	38
6.2.3	(Definition) Transpose	38
6.2.1	(Corollary) 转置的性质	38
6.2.4	(Definition) 共轭转置	38
6.2.2	(Corollary) 共轭转置的性质	39
6.2.5	(Definition) 矩阵乘法	39
6.2.3	(Corollary) 矩阵乘法性质	39
6.2.6	(Definition) 分块矩阵乘法	39
6.2.7	(Definition) 矩阵-向量乘积 Ax	39
6.2.4	(Corollary)	39
6.2.8	(Definition) 矩阵-向量乘积函数 $f(x) = Ax$	41
6.2.9	(Definition) Matrix Power	41
6.3.1	(Definition) Zero Matrix	43
6.3.2	(Definition) 单位矩阵	43
6.3.1	(Corollary)	43
6.3.3	(Definition) Symmetric Matrices	43
6.3.4	(Definition) Hermitian Matrices	43
6.3.5	(Definition) Diagonal Matrices	43
6.3.6	(Definition) 下三角矩阵	43
6.3.7	(Definition) 上三角矩阵	43
6.3.8	(Definition) Selector matrices	45
6.3.9	(Definition) Downsampling	45
6.3.10	(Definition) 关联矩阵	45
6.3.11	(Definition) 一维卷积	45
6.3.2	(Corollary)	45
6.3.3	(Corollary) 卷积性质	46
6.3.4	(Corollary)	46
6.3.12	(Definition) 多项式	46
6.3.13	(Definition) Vandermonde Matrices	46
6.3.14	(Definition) Discrete Fourier Transform (DFT)	46

6.3.15 (Definition) Discrete Inverse Fourier Transform	47
6.3.16 (Definition) 半正定矩阵	47
6.3.17 (Definition) 正定矩阵	47
6.3.18 (Definition) 二次型	47
6.3.5 (Corollary)	47
6.3.6 (Theorem) 正定矩阵对角元素性质	48
6.3.7 (Theorem) 半正定矩阵对角元素性质	48
6.4.1 (Definition) 实矩阵 A 的 Gram 矩阵	48
6.4.2 (Definition) 复矩阵的 A 的 Gram 矩阵	48
6.4.1 (Theorem)	48
6.4.2 (Theorem)	48
6.4.3 (Corollary)	48
Chapter 7 Matrices Norms	49
7.1.1 (Definition) Matrix Norm	49
7.1.2 (Definition) 从属于给定向量范数 $\ x\ _v$ 的矩阵范数	50
7.1.3 (Definition) 向量范数和算子范数相容	50
7.1.1 (Theorem) 算子范数服从乘法范数相容性	50
7.1.4 (Definition) A 的列范数	51
7.1.5 (Definition) A 的行范数	51
7.1.6 (Definition) A 的 2-范数	51
Chapter 8 适定问题	54
8.3.1 (Definition) 非奇异矩阵 A 的条件数 (condition number)	54
8.3.1 (Corollary) 非奇异矩阵 A 的条件数 (condition number) 性质	54
Chapter 9 Inverse of Matrices	55
9.1.1 (Definition) A 的左逆	55
9.1.2 (Definition) A 的右逆	55
9.1.1 (Theorem)	56
9.1.2 (Theorem)	56
9.1.3 (Theorem)	56
9.1.3 (Definition) 逆 A^{-1}	56
9.2.1 (Definition) Linear Equation Systems	56
9.2.1 (Theorem)	56
9.2.2 (Theorem)	57
9.2.3 (Theorem)	57
9.4.1 (Theorem)	57
9.4.2 (Theorem)	58
9.4.3 (Corollary)	58
9.4.4 (Theorem)	59

9.4.5 (Theorem)	59
9.4.6 (Corollary)	59
9.4.7 (Theorem)	59
9.5.1 (Theorem) 转置 A^T 和共轭转置 A^H	60
9.5.2 (Corollary)	60
9.6.1 (Corollary) Gram Matrix 可逆等价于 A 列线性无关	60
9.7.1 (Definition) Pseudo-inverse	61
9.7.1 (Theorem)	61
9.7.2 (Theorem)	61
9.7.3 (Theorem)	61
9.7.4 (Corollary)	61
9.7.5 (Corollary)	61
10.1.1 (Theorem)	64
10.1.2 (Theorem)	65
10.1.3 (Theorem)	65
10.1.4 (Theorem)	65
10.1.5 (Theorem)	65
10.1.6 (Theorem)	65
10.1.7 (Theorem)	66
10.1.8 (Theorem)	66
Chapter 10 Orthogonal Matrices	66
10.2.1 (Definition) 正交矩阵	66
10.2.1 (Theorem) 正交矩阵满足非奇异性	66
10.2.2 (Corollary)	66
10.2.3 (Corollary)	66
10.3.1 (Definition) 置换	66
10.3.1 (Theorem)	66
10.3.2 (Corollary)	66
10.3.3 (Corollary)	67
10.5.1 (Definition) Householder matrix, reflector matrix	68
10.5.1 (Theorem)	68
10.5.2 (Theorem)	68
10.5.3 (Theorem)	68
10.5.4 (Corollary)	68
10.5.5 (Corollary)	69
10.6.1 (Corollary) 正交矩阵乘积的正交性	69
10.8.1 (Theorem)	69
10.8.2 (Theorem)	70
10.8.3 (Theorem)	70
10.9.1 (Definition) 向量集合张成的空间	70

10.9.2 (Definition) A 的范围 (列空间)	70
10.1 (Problem)	70
10.9.1 (Theorem)	71
Chapter 11 QR 分解与 Householder 变换	72
11.1.1 (Definition) Lower Triangular Matrices	72
11.1.2 (Definition) Upper Triangular Matrices	72
11.1.3 (Definition) 单位上三角矩阵, 单位下三角矩阵	72
11.1 (Problem)	72
11.2 (Problem)	72
11.1.1 (Theorem)	73
11.1.2 (Theorem) 高斯消元法	73
11.1.3 (Theorem)	73
11.2.1 (Theorem) QR Factorization	74
11.2.2 (Corollary)	74
11.2.3 (Corollary)	74
11.2.4 (Corollary)	74
11.2.5 (Corollary)	74
11.2.6 (Corollary)	74
11.2.7 (Theorem) 满秩矩阵 A 的 QR 分解的存在唯一性	75
11.2.8 (Theorem)	76
11.3.1 (Corollary)	76
11.3.1 (Definition) 线性无关列向量的矩阵 A 的伪逆	77
11.3.2 (Theorem)	77
11.3.3 (Corollary)	77
11.3.4 (Corollary)	77
11.3.5 (Theorem)	78
11.3.6 (Theorem)	78
11.3.7 (Theorem)	78
11.4.1 (Corollary)	80
11.4.2 (Corollary)	80
11.4.3 (Corollary)	80
11.4.4 (Theorem)	80
11.6.1 (Theorem)	84
11.6.2 (Theorem)	85
11.6.1 (Definition)	86
11.6.3 (Theorem)	86
11.6.2 (Definition) Constructed Householder Matrix	86
11.6.4 (Theorem)	86
11.3 (Problem)	90
11.7.1 (Definition) v_1, \dots, v_n 的完整表示	91

11.7.1 (Theorem)	92
11.7.2 (Definition) 矩阵-向量积 $H_k x$	92
11.8.1 (Definition) Givens Rotation in the 1-2 plane	92
11.8.1 (Theorem)	93
11.8.2 (Theorem) H 的逆是它本身	94
Chapter 12 Fourier Series, Fourier Transform	96
12.1.1 (Definition) Complex Numbers	96
12.1.2 (Definition) Complex Number in Polar Coordinates	96
12.1.1 (Theorem) Euler's Formular	96
12.1.2 (Corollary)	96
12.1.3 (Definition) 正弦函数	96
12.1.4 (Definition) 三角级数	97
12.1.5 (Definition) 三角函数系	97
12.1.6 (Definition) 复数域的 Gram 矩阵	97
12.1.7 (Definition) 酉矩阵	98
12.1.8 (Definition) 酉矩阵的逆	98
12.2.1 (Definition) 傅里叶级数	98
12.2.1 (Theorem) 收敛定理, 狄利克雷 (Dirichlet) 充分条件	99
12.2.2 (Corollary)	99
12.2.2 (Definition) 正弦级数	99
12.2.3 (Definition) 余弦级数	99
12.2.3 (Theorem)	100
12.2.4 (Definition) 傅里叶级数的复数形式	100
12.3.1 (Definition) The Fourier transform of a continuous function $f(t)$ of a continuous variable t	101
12.3.2 (Definition) inverse Fourier transform	102
12.4.1 (Definition) Discrete Fourier Transform	102
12.4.2 (Definition) inverse discrete Fourier transform (IDFT)	102
12.4.3 (Definition) 离散傅里叶变换矩阵 W	102
12.4.1 (Corollary)	102
12.4.2 (Corollary)	102
12.4.3 (Corollary)	102
12.4.4 (Corollary)	102
12.4.4 (Definition) 离散傅里叶反变换	103

