利润最大化

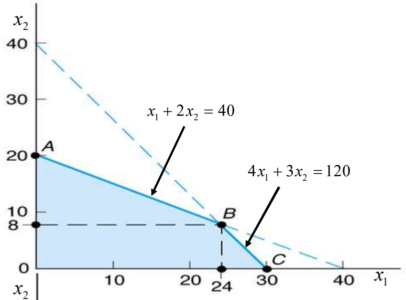
$$\max_{x_1,x_2} 40x_1 + 50x_2$$

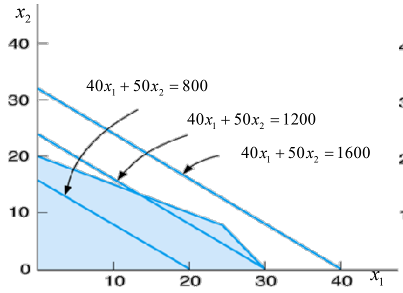
$$s.t. x_1 + 2x_2 \le 40$$

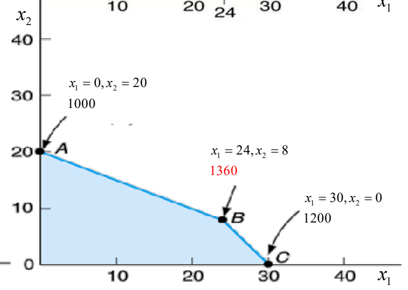
$$4x_1 + 3x_2 \le 120$$

$$x_1 \ge 0$$

$$x_2 \ge 0$$







1.1.4 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。
- nnz(x)是指向量x中非零的项数(number of non-zeros)，有时用 ℓ_0表示 。
- 例子：零向量0_3，单位向量e_2。

$$0_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

1.2 数域

$\mathbb{Q}$	$\mathbb{R}$	$\mathbb{C}$
有理数、	**实数、**	**复数**
$\frac{1}{3}$	π	$1+i$

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。**

例如: $x, y \in \mathbb{R}, x = 1, y = 2.$

$$x + y \in \mathbb{R} \qquad x \times y \in \mathbb{R}$$

实数x 加（乘）实数y：结果$x+y$ ($x\times y$)是实数

1.2 向量空间

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$$

设V是非空子集,P是一数域,向量空间V满足：

1. 向量加法：$V + V \to V$，记作$\forall x, y \in V$，则$x + y \in V$(加法封闭)
2. 标量乘法：$F \times V \to V$，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. $\lambda(x + y) = \lambda x + \lambda y$

向量空间也称为线性空间。 $\mathbb{R}^n$

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

1.4 内积

- 定义：在数域$\mathbb{R}$上的向量空间V定义函数$\langle\cdot,\cdot\rangle$:V× V → $\mathbb{R}$, 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 $\langle a, b\rangle = \langle b, a\rangle, \forall a,b \in V$;

则函数$\langle\cdot,\cdot\rangle$称为内积。

- 例子：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：

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

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

1.4.1 内积的性质

- 交换律：$a^T b = b^T a$
- 结合律：$(\gamma a)^T b = \gamma(a^T b)$
- 分配律：$(a + b)^T c = a^T c + b^T c$
- 例如：

$$(a+b)^T (c+d) = a^T c + a^T d + b^T c + b^T d$$

1.4.2 常用的内积等式

- 选出第i项：$e_i^T a = a_i$
- 向量每一项之和：$\mathbf{1}^T a = a_1 + \dots + a_n$
- 向量每一项的平方和：$a^T a = a_1^2 + \dots + a_n^2$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n$$

第i项为1，其它为0

1.4.3柯西—施瓦茨Cauchy-Schwartzn不等式

设$\langle\cdot,\cdot\rangle$是向量空间$V$上的内积，$\forall x, y \in V$，则有

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

当$x = \beta y$，$\beta \in \mathbb{R}$等式成立.

证明：令$\lambda \in \mathbb{R}$，则有

$$\begin{aligned} 0 &\leq \langle x+\lambda y, x+\lambda y\rangle \\ &= \langle x,x\rangle + \lambda\langle y,x\rangle + \lambda\langle x,y\rangle + \lambda^2\langle y,y\rangle \\ &= \langle x,x\rangle + 2\lambda\langle y,x\rangle + \lambda^2\langle y,y\rangle \end{aligned}$$

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

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

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

当$|\langle x,y\rangle|^2 = \langle x,x\rangle\langle y,y\rangle$时，即有$\langle x,x\rangle + 2\lambda\langle y,x\rangle + \lambda^2\langle y,y\rangle=0$，

也即$\langle x+\lambda y, x+\lambda y\rangle=0$，因此$x+\lambda y=0$，即$x=-\lambda y$.

1.4.3 利润最大化：向量化表示

令 $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, a = \begin{bmatrix} 40 \\ 50 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, c = \begin{bmatrix} 4 \\ 3 \end{bmatrix} x \in \mathbb{R}^2.$

$$
\begin{aligned}
\max_{x_1, x_2} \ & 40x_1 + 50x_2 \\
s.t. \ & x_1 + 2x_2 \le 40 \\
& 4x_1 + 3x_2 \le 120 \\
& x_1 \ge 0 \\
& x_2 \ge 0
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
\max_{x} \ & a^T x \\
s.t. \ & b^T x \le 40 \\
& c^T x \le 120 \\
& x_1 \ge 0 \\
& x_2 \ge 0
\end{aligned}
$$

2.1 线性函数

- **线性函数**(linear function)：f： $\mathbb{R}^n \rightarrow \mathbb{R}$. f是一个将n维向量映射成数的函数。
- 线性函数f满足以下两个性质(□ ∈ $\mathbb{R}$, □ ,□ ∈ $\mathbb{R}^n$)：
  - **齐次性(homogeneity)**： □ (k)=k (□),
  - **叠加性(Additivity)**：□ (□ +□)=□ (□)+□ (□)
- 一个函数如果满足这两个性质，就称其为**线性**函数。
- 例子：
  - 求平均值：$f(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$为线性函数。
  - 求最大值：$f(x) = max\{x_1, x_2, ..., x_n\}$并**不是**线性函数。

    因为：令$x = (1, -1), y = (-1, 1), \alpha = 0.5, \beta = 0.5$,
    有$f(\alpha x + \beta y) = 0 \neq \alpha f(x) + \beta f(y) = 1$

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

2.1.1 内积函数

- 对于n维向量a,满足以下形式的函数被称为**内积函数(inner product function)**：

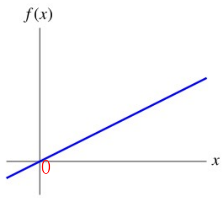
$$f(x) = a^T x = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$$

- 上述$f(x)$可以看作是每项x_i的加权之和。
- 内积函数都是线性的:

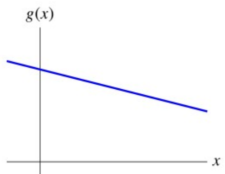
$$
\begin{aligned}
f(\alpha x + \beta y) &= a^T(\alpha x + \beta y) \\
&= a^T(\alpha x) + a^T(\beta y) \\
&= \alpha(a^T x) + \beta(a^T y) \\
&= \alpha f(x) + \beta f(y)
\end{aligned}
$$

2.1.3 仿射函数

- 定义：一个线性函数加上一个常数称为**仿射函数**(affine function)。
- 其一般形式为 $f(x) = a^T x + \mathrm{b}$，其中$a \in \mathbb{R}^n$，b∈ $\mathbb{R}$为标量。
- 函数f： $\mathbb{R}^n \rightarrow \mathbb{R}$为**仿射函数**满足以下条件：
  - $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$， $\alpha + \beta = 1$, $\alpha, \beta \in \mathbb{R}, x, y \in \mathbb{R}^n$



f为线性函数



g为仿射函数，非线性函数

2.2 梯度与偏导

- 假设f：$\mathbb{R}^n \to \mathbb{R}$，函数$f$在$z$点可微,其第$i$个分量的一阶偏导数为：

$$\frac{\partial f}{\partial z_i}(z) = \lim_{t\to 0}\frac{f(z_1,\cdots,z_{i-1},z_i+t,z_{i+1},\cdots,z_n)-f(z)}{t}$$
$$=\lim_{t\to 0}\frac{f(z+te_i)-f(z)}{t}$$

- f在点z的梯度为：

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

2.2 泰勒公式

- **泰勒(近似)公式**：一个用函数在某点的信息描述其附近取值的公式。
- 假设f：$\mathbb{R}^n \to \mathbb{R}$，函数$f$在$z$点充分光滑, 即处处可导。
- 函数$f(x)$在z处作泰勒$(Taylor)$展开：

$$f(x)=f(z)+\frac{\partial f}{\partial x_1}(z)(x_1-z_1)+\frac{\partial f}{\partial x_2}(z)(x_2-z_2)+\cdots+\frac{\partial f}{\partial x_n}(z)(x_n-z_n)$$
$$+\ \underbrace{\frac{1}{2!}\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{\partial f^2}{\partial x_i\partial x_j}}_{\text{二阶偏导}}(z)(x_i-z_i)(x_j-z_j)+\cdots$$

2.2 一阶泰勒近似

- 假设f：$\mathbb{R}^n \to \mathbb{R}$，函数$f$在$z$点可导，其附近的一阶泰勒公式为：

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

- 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。
- 写成内积形式：

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

2.3 回归模型

- **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)。

问题：如何建立模型求解权重系数β和偏移量v？

- 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)

3.1 范数

- **向量范数**(vector norm)：在向量空间中存在一个函数$\|\cdot\|:\mathbb{R}^n\to\mathbb{R}$，且满足以下条件：
  - **齐次性**： $\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$，$\alpha\in\mathbb{R}$且$\boldsymbol{x}\in\mathbb{R}^n$；
  - **三角不等式**：$\|\boldsymbol{x}+\boldsymbol{y}\| \le \|\boldsymbol{x}\|+\|\boldsymbol{y}\|$，$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$；
  - **非负性**：$\|\boldsymbol{x}\| \ge \boldsymbol{0}$，$\boldsymbol{x}\in\mathbb{R}^n$且$\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = 0$；

  则称$\|\cdot\|$为向量范数。
- 例子：向量空间$\mathbb{R}^n$有许多范数：
  - ℓ_1-范数（曼哈顿范数， Manhattan norm）：

$$\|\boldsymbol{x}\|_1 = |\boldsymbol{x}_1| + |\boldsymbol{x}_2|+\ldots+|\boldsymbol{x}_n|,\ x,y\in\mathbb{R}^n,\alpha\in\mathbb{R}.$$

$$\|\alpha x\|_1 = |\alpha x_1| + |\alpha x_2| + \ldots + |\alpha x_n| = |\alpha|\|x\|_1 \ge 0$$

$$\|x+y\|_1 = |x_1+y_1| + \ldots + |x_n+y_n| \le |x_1| + |y_1| + \ldots + |x_n| + |y_n| = \|x\|_1 + \|y\|_1$$

3.1 范数

- 向量空间$\mathbb{R}^n$常用的范数：
  - ℓ_2-范数（欧几里得范数，Euclidean norm）：

$$\|\boldsymbol{x}\|_2 = \sqrt{(\boldsymbol{x}_1^2+\boldsymbol{x}_2^2+\ldots+\boldsymbol{x}_n^2)} = \sqrt{\boldsymbol{x}^T\boldsymbol{x}} = (\langle x,x\rangle)^{\frac{1}{2}}$$

柯西—施瓦茨不等式

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

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

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

$$= \|x\|_2^2 + 2\langle x,y\rangle + \|y\|_2^2 \le \|x\|_2^2 + 2\|x\|_2\,\|y\|_2 + \|y\|_2^2 = (\|x\|_2 + \|y\|_2)^2$$

$$\|x+y\|_2 \le \|x\|_2 + \|y\|_2$$

3.1 范数

- 向量空间$\mathbb{R}^n$常用的范数：
  - ℓ_p-范数，p≥ 1：

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

  - ℓ_1-范数$\|\boldsymbol{x}\|_1$，ℓ_2-范数$\|\boldsymbol{x}\|_2$是ℓ_p-范数的特例。
- **Minkowshi**不等式：

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

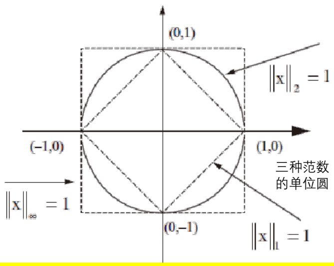
- **Hölder**不等式：

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

3.1 范数

- 向量空间$\mathbb{R}^n$常用的范数：
  - ℓ_∞-范数：

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



$$\max_{1\le i\le n}|x_i| \le \left(|x_1|^p + \cdots + |x_i|^p + \cdots + |x_n|^p\right)^{1/p} \le \left(n\max_{1\le i\le n}|x_i|^p\right)^{1/p}$$

$$= n^{1/p}\max_{1\le i\le n}|x_i| \ \to\ \max_{1\le i\le n}|x_i|\quad (p\to\infty)$$

- ℓ_∞-范数是ℓ_p-范数的特例，p⟶ ∞, $\|\boldsymbol{x}\|_p \longrightarrow \|\boldsymbol{x}\|_\infty$

3.1.1 均方根

- 向量$\boldsymbol{x} \in \mathbb{R}^n$的均方值(mean-square value)为:

$$\frac{\boldsymbol{x}_1^2 + \boldsymbol{x}_2^2 + \dots + \boldsymbol{x}_n^2}{n} = \frac{\|\boldsymbol{x}\|_2^2}{n}$$

- n维向量$\boldsymbol{x}$的均方根(root-mean-square value，RMS)为:

$$\boldsymbol{rms}(x) = \sqrt{\frac{\boldsymbol{x}_1^2 + \boldsymbol{x}_2^2 + \dots + \boldsymbol{x}_n^2}{n}} = \frac{\|\boldsymbol{x}\|_2}{\sqrt{n}}$$

- rms(x)给出了$|\boldsymbol{x}_i|$的“典型”(typical)值。
- 例如，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。

3.1.2 切比雪夫不等式

- 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：$\|\boldsymbol{x}\|_2^2 = x_1^2 + x_2^2 + \dots + x_n^2 \geq ka^2$
- 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$\frac{\|x\|_2^2}{a^2}$
- 以上就是**切比雪夫不等式**。
- 使用均方根(RMS)来描述： $\boldsymbol{rms}(x) = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} = \frac{\|x\|_2}{\sqrt{n}}$

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

- 例子：不会超过4%的项能满足$|x_i| \geq 5 * \boldsymbol{rms}(x)$。

3.3 标准差

- 定义：**标准差**是算术平均值的算术平方根，通常希腊字σ表示。
- 对于n维向量x，其平均值为$\boldsymbol{avg}(x) = \frac{\mathbf{1}^T x}{n}$。
- 去均值(de-meaned)向量为$\tilde{x} = x - \boldsymbol{avg}(x)\mathbf{1}$，因此$\boldsymbol{avg}(\tilde{x}) = 0$。
- x的标准差可以表示为:

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

- $\boldsymbol{std}(x)$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，$\boldsymbol{std}(x) = 0$。
- 一个基本公式:

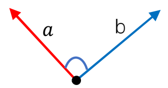
$$rms(x)^2 = avg(x)^2 + std(x)^2$$

课后作业

3.4 角

- 两个非零向量a和b之间的**角**(angle)定义为如下形式:

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



- $\angle(a, b)$的取值范围为$[0, \pi]$，且满足:

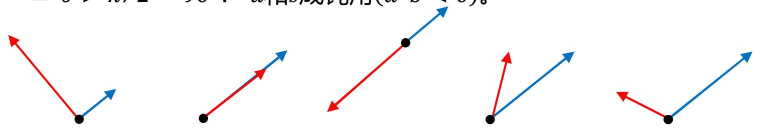
$$a^T b = \|a\|_2 \|b\|_2 \cos\big(\angle(a, b)\big)$$

- 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。

3.4.1 角的分类

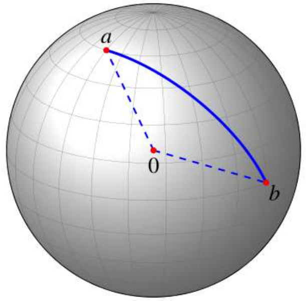
$\theta = \angle(a, b)$

- $\theta = \pi/2 = 90°$： 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$(a^T b = -\|a\|\|b\|)$。
- $\theta < \pi/2 = 90°$： a和b成锐角$(a^T b > 0)$。
- $\theta > \pi/2 = 90°$： a和b成钝角$(a^T b < 0)$。



3.4.1 球面距离

- 如果a，b都在半径为R的球面上，沿球面的距离为R$\angle(a, b)$。



3.4.2 相关系数

- 给定向量a和b，其去均值向量为：

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

- 在$\tilde{a} \neq 0$和 $\tilde{b} \neq 0$下，a和b的相关系数为：

$$\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\|_2 \|\tilde{b}\|_2}$$

- $\rho = \cos\angle(\tilde{a}, \tilde{b})$：
  - $\rho = 0$： a和b是不相关的。
  - $\rho > 0.8$(左右)： a和b高度正相关。
  - $\rho < -0.8$(左右)： a和b高度负相关。
- 高度相关指的是a_i和b_i通常都在它们两者的平均值之上(之下)。

4.1 渐近记号 — o记号

高阶无穷小记号o (小o 记号)

设x，y是同一变化过程中的无穷小，即$x \to 0$，$y \to 0$，如果它们极限

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

则称y是x的高阶无穷小，记作$y = o(x)$。同时存在常数C，满足

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

也即则称y是Cx的高阶无穷小，记作$y = o(Cx)$。

4.1 必要条件

■ 定理1.1　假设函数f在$\hat{x}$可微，则有

$$\hat{x} = \underset{x\in\mathbb{R}^n}{argmin} f(x) \Rightarrow \nabla f(\hat{x}) = 0.$$

证明：函数f在$\hat{x}$一阶泰勒展开，则有

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

假设$\nabla f(\hat{x}) \neq 0$,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$t \longrightarrow 0$,则 $t\|\nabla f(\hat{x})\|_2 \longrightarrow 0$,高阶无穷小$o(t\|\nabla f(\hat{x})\|_2) \longrightarrow 0$

当t足够小时，存在$t\|\nabla f(\hat{x})\|_2 > o(t\|\nabla f(\hat{x})\|_2)$，即

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

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

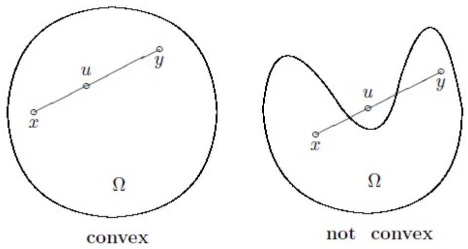
与 $\hat{x} = \underset{\mathbb{R}^n}{\mathrm{argmin}} f(x)$ 矛盾。

4.2 凸集

凸集的定义

定义域$\Omega \in \mathbb{R}^n$称为凸的$(Convex)$集合，则$\forall x, y \in \Omega, \alpha \in \mathbb{R}, 0 \leq \alpha \leq 1$,有

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



$$u = \alpha x + (1-\alpha)y = y - \alpha(y - x)$$

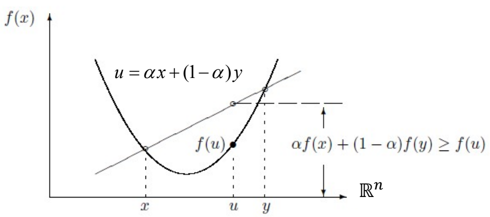
4.2 凸函数

凸函数

设函数$f(x)$定义于称为凸的定义域$\Omega \in \mathbb{R}^n$，当称其为凸函数时，

对于$\forall x, y \in \Omega, \alpha \in \mathbb{R}, 0 \leq \alpha \leq 1$，满足：

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y),$$



4.2 凸函数

例子1：$f(x) = x^2, x \in \mathbb{R}.$　　　$\forall x, y \in \Omega, \alpha \in \mathbb{R}, 0 \leq \alpha \leq 1$

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

例子2：$f(x) = \|x\|$，其中$\|\bullet\|$表示$\mathbb{R}^n$上的向量范数，$x \in \mathbb{R}^n$.

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

例子3：$f(x) = \|x\|_2^2, x \in \mathbb{R}^n$　？作业

4.2 凸函数

引理：可微函数f是凸函数的充要条件：

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

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

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

$\le \alpha f(x) + (1-\alpha) f(y)$，则有

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

令$\alpha \to 1-$，则有$(1-\alpha)(y-x) \to 0$，即$f(y) \ge f(x) + f'(x)(y-x)$.

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

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

可得$\alpha f(x) + (1-\alpha)f(y) \ge f(z) + \alpha f'(z)(x-z) + (1-\alpha)f'(z)(y-z)$

$= f(z) + f'(z)(\alpha x + (1-\alpha)y - z) = f(z)$

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

"$\Rightarrow$"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$g(0) \ge g(1) + g'(1)(-1)$，得证

"$\Leftarrow$"必要条件：与一维类似。 课后作业

4.2 凸函数

定理：如果可微函数f是凸函数，则有

$$\hat{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x) \Leftrightarrow \nabla f(\hat{x}) = 0.$$

证明：已证$\hat{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x) \Rightarrow$ 可得$\nabla f(\hat{x}) = 0$.

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

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

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

可得$f(x) \ge f(\hat{x})$，$\hat{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x)$.

4.3 优化问题：向量偏导

- 假设向量$x, z \in \mathbb{R}^n$：

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

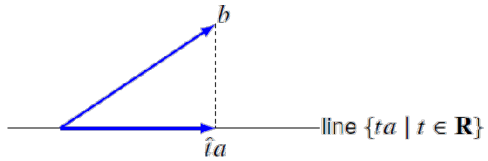
$$f(z) = x^T z + z^T z = \sum_{i=1}^{n} \{x_i z_i + z_i^2\}$$

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

4.3 优化问题：标量

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

$$\min_t \|ta - b\|_2^2$$



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

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

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

4.4 k-means算法

- **K-means算法**是将N个向量$x_i \in \mathbb{R}^n$划分成k类的迭代聚类算法。
- **聚类目标**找到向量x_i的“标签c_i”和“聚类中心z_j”。

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

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

- 步骤：
  - 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个聚类中心的距离，并将其分配到最近的聚类中心所在的聚类中；
  - 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；
  - 4、重复步骤2、3，直到所有聚类中心不再变化。

4.4 k-means算法

- **聚类目标**找到N个向量$x_i \in \mathbb{R}^n$的“标签c_i”和“聚类中心z_j”。
- **K-means算法**划分成k类步骤：
  - 1、在N个点中随机选取k个点作为聚类中心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$c_i = j$；

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

$$c_i = j$$

  - 3、**更新聚类中心**z_j：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；

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

  - 4、重复步骤2、3，直到所有聚类中心不再变化。

5.1 线性相关

- 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

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

则称向量$a_1, \dots a_m$是**线性相关**(linearly dependent)。

- 等价于：至少有一个向量a_i是其它向量的线性组合。
- 向量集$\{a_1\}$是线性相关的，当且仅当$a_1 = 0$。
- 向量集$\{a_1, a_2\}$是线性相关的，当且仅当其中一个$a_1 = \beta a_2, \beta \neq 0$。
- 对于两个以上的向量，则需要用定义去描述。

5.2 线性无关

- 定义：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 dependent)，也称**线性无关**，即：

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

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

- 等价于：不存在一个向量a_i是其它向量的线性组合。
- 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，P.94。
- 例子：n维单位向量$e_1, \dots, e_n$是线性独立的。

$$e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n \text{ 第}i\text{项为1，其它为0。}$$

5.2.1 线性无关向量的线性组合

- 假设x是线性无关向量$a_1, \ldots, a_k$的线性组合：

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

则其系数$\beta_1, \ldots \beta_k$是唯一的，即如果有：

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

则对于$i = 1, \ldots k$，有$\beta_i = \gamma_i$。

- 系数是唯一的原因：

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

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

5.3 基

- 定义：n个线性独立的n维向量$a_1, \ldots, a_n$的集合称为**基**(basis)
- 任何一个n维向量b 都可以用它们的线性组合来表示：

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

- 同一向量的系数$\beta_1, \ldots, \beta_n$是唯一的
- 上述等式称为向量b在基底$a_1, \ldots, a_n$下的分解
- 例子：$e_1, \ldots, e_n$是一组基，那么b在此基底下的分解为：

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

5.4 标准正交向量

- 定义：在**n维**向量集$a_1, \ldots, a_k$中，如果对于$i \neq j$，都有$a_i \perp a_j$，则称它们相互**正交**(orthogonal)。
- 如果n维向量集$a_1, \ldots, a_k$相互正交，且每个向量的模长都为单位长度1，即对于$i = 1, \ldots k$，有$\|a_i\|_2^2 = 1$，则称它们是**标准正交**(orthonormal)的。
- 标准正交向量用内积表示为：

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

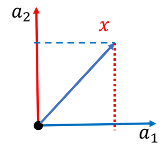
- 标准正交的向量集是线性无关的。
- 根据线性无关的性质，必有向量集向量个数$k \leq n$，P.94。
- 当$k = n$时，$a_1, \ldots, a_n$是n维向量的一个标准正交基。

5.4.2 标准正交分解

- 如果$a_1, \ldots, a_n$是一个标准正交基，对于任意n维向量x：

$$x = (a_1^T x) a_1 + \cdots + (a_n^T x) a_n$$

则称其为x在标准正交基下的标准正交分解。



- 为了验证以上公式，可以让等式两边同时乘以任意a_i：

$$a_i^T x = (a_1^T x) a_i^T a_1 + \cdots + (a_i^T x) a_i^T a_i + \cdots + (a_n^T x) a_i^T a_n = a_i^T x$$

5.5 Gram-Schmidt(正交化)算法

- 给定n维向量$a_1, ..., a_k$，如何将其标准正交化?
- **Gram-Schmidt(正交化)算法**常用于解决上述问题，且能检验$a_1, ..., a_k$是否是线性相关。
- 步骤（$i = 2, ..., k$）： $q_1 = a_1/\|a_1\|_2$
  - 1、**正交化**：$\tilde{q}_i = a_i - (q_1^T a_i)q_1 - ... - (q_{i-1}^T a_i)q_{i-1}$;
  - 2、**检验线性相关**：如果$\tilde{q}_i = 0$，提前退出迭代;
  - 3、**单位化**：$q_i = \tilde{q}_i/\|\tilde{q}_i\|_2$。
- 如果步骤2中未提前结束迭代，那么$a_1, ..., a_k$是线性独立的，而且$q_1, ..., q_k$是标准正交基。
- 如果在第j次迭代中提前结束，说明a_j是$a_1, ..., a_{j-1}$的线性组合，因此$a_1, ..., a_k$是线性相关的。

5.5.1 算法分析

归纳法来证明$q_1, ..., q_{i-1}, q_i$是标准正交的：

- 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：
  - $\tilde{q}_i = a_i - (q_1^T a_i)q_1 - ... - (q_{i-1}^T a_i)q_{i-1}$;
- 等式两边同时乘以q_j^T， $j = 1, ..., i-1$：

$$q_j^T\tilde{q}_i = q_j^T a_i - (q_1^T a_i)(q_j^T q_1) - \cdots - (q_{i-1}^T a_i)(q_j^T q_{i-1})$$
$$= q_j^T a_i - q_j^T a_i = 0 \quad \because q_j^T q_r = 0, j \neq r, q_j^T q_j = 1.$$

- 因此 $\tilde{q}_i \perp q_1, ..., \tilde{q}_i \perp q_{i-1}$。
- 单位化步骤保证了$q_i = \tilde{q}_i/\|\tilde{q}_i\|_2$,即$q_1, ..., q_i$是标准正交。

5.5.1 算法分析

假设Schmidt正交法未在第i次迭代提前终止：

- $\tilde{q}_i = a_i - (q_1^T a_i)q_1 - ... - (q_{i-1}^T a_i)q_{i-1}$;
- a_i是$q_1, ..., q_i$的一个线性组合： $q_i = \tilde{q}_i/\|\tilde{q}_i\|_2$

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

- 则有 $q_i = (1/\|\tilde{q}_i\|_2)(a_i - (q_1^T a_i)q_1 - \cdots - (q_{i-1}^T a_i)q_{i-1})$
- 归纳假设，每个q_{i-1}都是$a_1, ..., a_{i-1}$的线性组合：

$$q_2 = (1/\|\tilde{q}_2\|_2)(a_2 - (q_1^T a_2)q_1)$$
$$= (1/\|\tilde{q}_2\|_2)(a_2 - (q_1^T a_2)\, a_1/\|a_1\|_2)$$
$$q_3 = (1/\|\tilde{q}_3\|)(a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2)$$

课后作业

- 通过对i的归纳证明，可得q_i是$a_1, ..., a_i$的一个线性组合。

5.5.1 算法分析

假设Schmidt正交法在第j次迭代提前终止：

- a_j是$q_1, ..., q_{j-1}$的一个线性组合：

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

- 每一个$q_1, ..., q_{j-1}$都是$a_1, ..., a_{j-1}$的线性组合。
- 因此a_j是$a_1, ..., a_{j-1}$的线性组合。

5.5.2 例子

- 给定三个向量，使用Schmidt正交法来正交化：

$$a_1=(-1,1,-1,1),\quad a_2=(-1,3,-1,3),\quad a_3=(1,3,5,7)$$

- $i=1$：$\tilde{q}_1=a_1, \|\tilde{q}_1\|_2=2$，单位化：

$$q_1=\frac{1}{\|\tilde{q}_1\|_2}\tilde{q}_1=\left(-\frac{1}{2},\frac{1}{2},-\frac{1}{2},\frac{1}{2}\right)$$

- $i=2$：$q_1^T a_2=4$，正交化：

$$\tilde{q}_2=a_2-(q_1^T a_2)q_1=\begin{bmatrix}-1\\3\\-1\\3\end{bmatrix}-4\begin{bmatrix}-1/2\\1/2\\-1/2\\1/2\end{bmatrix}=\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

5.5.2 例子

- $i=2$：$\|\tilde{q}_2\|=2$，单位化：

$$q_2=\frac{1}{\|\tilde{q}_2\|_2}\tilde{q}_2=\left(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)$$

- $i=3$：$q_1^T a_3=2$和$q_2^T a_3=8$，正交化：

$$\tilde{q}_3=a_3-(q_1^T a_3)q_1-(q_2^T a_3)q_2=\begin{bmatrix}1\\3\\5\\7\end{bmatrix}-2\begin{bmatrix}-1/2\\1/2\\-1/2\\1/2\end{bmatrix}-8\begin{bmatrix}1/2\\1/2\\1/2\\1/2\end{bmatrix}=\begin{bmatrix}-2\\-2\\2\\2\end{bmatrix}$$

- $i=3$：$\|\tilde{q}_3\|_2=4$，单位化：

$$q_3=\frac{1}{\|\tilde{q}_3\|_2}\tilde{q}_3=\left(-\frac{1}{2},-\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)$$

5.5.3 时间复杂度

- 第i次迭代的步骤1需要进行$i-1$次内积：

$$q_1^T a_i, \ldots, q_{i-1}^T a_i$$

  需要进行 $(i-1)$(2n-1)次flop。

- 计算$\tilde{q}_i$需要进行 2n$(i-1)$ 次flop。　$\tilde{q}_i=a_i-(q_1^T a_i)q_1-\ldots-(q_{i-1}^T a_i)q_{i-1}$;

- 计算$\|\tilde{q}_i\|_2$和q_i需要进行 3n 次flop。

- 总共需要：

$$\sum_{i=1}^{k}\left((4n-1)(i-1)+3n\right)=(4n-1)\frac{k(k-1)}{2}+3nk\approx 2nk^2$$

- 因此Schmidt正交法的时间复杂度为$2nk^2$(其中n为向量维数，k为向量个数)。

6.2 矩阵形状

- 标量：不区分一个1×1矩阵和一个标量
- 向量：不区分一个n×1矩阵和一个向量
- 行向量和列向量
  - 一个1×n矩阵被称为一个行向量
  - 一个n×1矩阵杯称为一个列向量
- 高形，宽形和方形矩阵，一个大小为m×n的矩阵为：
  - 高的，如果m>n
  - 宽的，如果m<n
  - 方的，如果m=n

$$\begin{bmatrix}1&0\\1&1\\1&0\end{bmatrix}\quad\begin{bmatrix}1&0&0\\1&1&1\end{bmatrix}\quad\begin{bmatrix}1&0&0\\1&1&1\\3&2&1\end{bmatrix}$$

6.4 矩阵的行表示和列表示

- 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$a_1, ..., a_n \in \mathbb{R}^m$，则有

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

- 或者通过其行向量$b_1, ..., b_m$进行表示：

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

6.6 对称和Hermitian矩阵

- 对称矩阵：$A_{ij} = A_{ji}$

$$A = \begin{bmatrix} 4 & 3 & -2 \\ 3 & -1 & 5 \\ -2 & 5 & 0 \end{bmatrix}, A^T = A.$$

- Hermitian矩阵：$A_{ij} = \overline{A}_{ji}$(共轭复数)

$$\begin{bmatrix} 4 & 3-2i & -1+i \\ 3+2i & -1 & 2i \\ -1-i & -2i & 0 \end{bmatrix}$$

6.7 对角矩阵，三角矩阵

- 对角矩阵：对角线上元素不全为0，其余元素全为0

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

- 下三角矩阵：方形矩阵且当$i < j$时$A_{ij} = 0$

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & -2 & 0 \\ 1 & 1 & 5 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 5 \end{bmatrix}$$

- 上三角矩阵：方形矩阵且当$i > j$时$A_{ij} = 0$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & -2 & 1 \\ 0 & 0 & 5 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 5 \end{bmatrix}$$

6.10 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：$(A^H)_{ij} = \overline{A}_{ji}$，$i = 1, ..., n; j = 1, ..., m;$
- 例如矩阵A∈ $\mathbb{C}^{m\times n}$，则其共轭转置为一个n × m矩阵

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

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

6.11 矩阵乘法

- 定义：设矩阵A∈ $\mathbb{R}^{m\times p}$,$B \in \mathbb{R}^{p\times n}$,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$C_{ij}$

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

- 例子：

$$\begin{bmatrix} -1.5 & 3 & 2 \\ 1 & -1 & 0 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ 0 & -2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3.5 & -4.5 \\ -1 & 1 \end{bmatrix}$$

- 矩阵A的列大小必须等于B的行大小

6.12 矩阵乘法性质

- 结合律：$(AB)C = A(BC)$
- 分配律：$A(B + C) = AB + AC$
- $(AB)^T = B^T A^T$，$(AB)^{\mathrm{H}} = B^H A^H$
- 一般情况下：$AB \neq BA$
- 对于方阵A有，$IA = AI = A$

6.13 分块矩阵乘法

- 例子：

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

6.14 矩阵-向量乘积

- 矩阵$A \in \mathbb{R}^{m\times n}$和一个向量$x \in \mathbb{R}^n$的积为

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

- Ax是矩阵A列向量的线性组合

$$Ax = [a_1 \quad a_2 \quad \cdots \quad a_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1a_1 + \cdots + x_na_n$$

6.15 矩阵-向量乘积函数

- 给定矩阵$A \in \mathbb{R}^{m\times n}$，定义函数$f: \mathbb{R}^n \to \mathbb{R}^m$，$f(x) = Ax$
- 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

$$\begin{aligned} f(x) &= f(x_1e_1 + x_2e_2 + \cdots + x_ne_n) \\ &= x_1f(e_1) + x_2f(e_2) + \cdots + x_nf(e_n) \\ &= [f(e_1) \quad \cdots \quad f(e_n)]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

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

6.15 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^2flops
  - A为稀疏矩阵时：flops << 2mn

6.16 例子($f: \mathbb{R}^3 \to \mathbb{R}^3$)

- f颠倒向量x中的元素的顺序，一个线性函数$f(x) = Ax$,其中

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

- f对向量x中的元素进行升序排序，非线性;
- f将向量x中的元素替换成相应的绝对值，非线性;

6.17 反转和循环移位

- 反转矩阵

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & ⋰ & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad Ax = \begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_2 \\ x_1 \end{bmatrix}$$

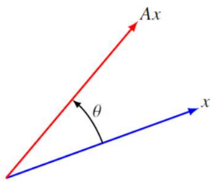
- 循环移位矩阵

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad Ax = \begin{bmatrix} x_n \\ x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{bmatrix}$$

6.18 平面旋转

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

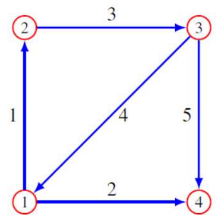
- Ax将向量x进行旋转，角度为θ



6.19 节点弧关联矩阵

- 假设有向图G有m个顶点，n条弧
- 则关联矩阵A大小为m×n，其中

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



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

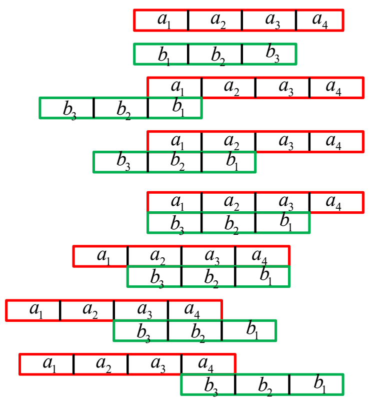
6.20 向量卷积

- 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$c\in\mathbb{R}^{m+n-1}$

$$c_k=\sum_{i+j=k+1}a_ib_j,\quad k=1,\ldots n+m-1$$

- 记为c= a *b
- 例如：n=4，m=3

$$c_1=a_1b_1$$
$$c_2=a_1b_2+a_2b_1$$
$$c_3=a_1b_3+a_2b_2+a_3b_1$$
$$c_4=a_2b_3+a_3b_2+a_4b_1$$
$$c_5=a_3b_3+a_4b_2$$
$$c_6=a_4b_3$$



6.21 向量卷积性质

- 假设向量a和b分别是多项式$p(x)$，$q(x)$的系数

$$p(x)q(x)=c_1+c_2x+\cdots+c_{m+n-1}x^{m+n-2}$$

- 则c= a *b是多项式$p(x)q(x)$的系数
- 性质：$p(x)=a_1+a_2x+\cdots+a_nx^{n-1}$，$q(x)=b_1+b_2x+\cdots+b_mx^{m-1}$
  - 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

6.22 向量和Toeplitz矩阵

- 如果固定a,则$c = a * b$是一个线性函数
- 如果固定b,则$c = a * b$是一个线性函数
- 例子：4维向量a和3维向量b，则c= a *b

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

- 上述的矩阵向量乘积中的矩阵被称为Toeplitz矩阵

6.23 Vandermonde矩阵

- 多项式$p(t)$其度为n-1,系数为$x_1, x_2, \dots, x_n$

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

- $p(t)$在m个点中$t_1, t_2, \dots, t_m$的值为

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

- 矩阵A被称为**Vandermonde矩阵**

6.24 离散傅里叶变换(DFT)

- DFT将n维复向量x映射为n维复向量y($\mathbb{C}^n \to \mathbb{C}^n$)

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

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

- 其中$\omega = e^{2\pi i/n}$
- DFT矩阵W的第k行第l列的元素为$W_{kl} = \omega^{-(k-1)(l-1)}$

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

6.25 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：

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

- 对称矩阵$A \in \mathbb{R}^{n\times n}$称为**正定矩阵**，满足以下条件：

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

这是半正定矩阵的一个子集。

- 注：如果对称矩阵$A \in \mathbb{R}^{n\times n}$，则$x^T Ax$是函数：

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

这叫做**二次型**。

6.27 半正定矩阵例子

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

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

- 如果$a>4$，矩阵A为正定矩阵：

$$x^TAx>0 \quad \forall x\neq 0$$

- 如果$a=4$，矩阵A为半正定矩阵，但不是正定矩阵：

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

- 如果$a<4$，矩阵A不是半正定矩阵：

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

6.28 正定矩阵性质

- 正定矩阵A都是非奇异的

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

(最后一步由正定性得到的)

- 正定矩阵A有正的对角元素

$$A_{ii}=e_i^TAe_i>0$$

- 每个半正定矩阵A都有非负的对角元素

$$A_{ii}=e_i^TAe_i\geq 0$$

6.29 Gram矩阵

Gram矩阵A的定义：

$$A=B^TB$$

- 每个Gram矩阵都是半正定的：

$$x^TAx=x^TB^TBx=\|Bx\|_2^2\geq 0 \quad \forall x$$

- 如果Gram矩阵是正定的，则要满足：

$$x^TAx=x^TB^TBx=\|Bx\|_2^2>0 \quad \forall x\neq 0$$

也就是说， B的列向量是线性无关的。

6.3 矩阵范数

- **矩阵范数(Matrix norm)**：向量空间中存在一个函数$\|\cdot\|:\mathbb{R}^{m\times n}\to\mathbb{R}$，且满足以下条件：
  - **齐次性**：$\|\alpha A\|=|\alpha|\|A\|$，$\alpha\in\mathbb{R}$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**：$\|A\|\geq 0$，$A\in\mathbb{R}^{m\times n}$且$\|A\|=0\Leftrightarrow A=0$；

  则称$\|\cdot\|$为矩阵范数。
- 例子：向量空间$\mathbb{R}^{m\times n}$矩阵范数：
  - F-范数(Frobenius norm) :$\|A\|_F=\left(\sum_{i=1}^n\sum_{j=1}^n a_{ij}^2\right)^{1/2}$.

$\|A\|_F\geq 0, \|\alpha A\|_F=|\alpha|\|A\|_F, \alpha\in\mathbb{R}.$

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

Minkowshi不等式 $=\|A\|_F+\|B\|_F$

6.3 算子范数

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

$$\|A\|_v=\max_{x\neq 0}\left\{\frac{\|Ax\|_v}{\|x\|_v}\right\}=\max_{x\neq 0}\left\{\left\|A\frac{x}{\|x\|_v}\right\|_v\right\}=\max_{\|y\|_v=1}\left\{\|Ay\|_v\right\}\quad\text{---(1)}$$

可以验证$\|A\|_v$满足矩阵范数定义。

$$\|A\|_v\geq 0, \|\alpha A\|_v=|\alpha|\|A\|_v \quad \alpha\in\mathbb{R}$$

$$\|A+B\|_v=\max_{\|y\|_v=1}\|(A+B)y\|_v\leq\max_{\|y\|_v=1}\left\{\|Ay\|_v+\|By\|_v\right\}$$

$$\leq\max_{\|y\|_v=1}\|Ay\|_v+\max_{\|y\|_v=1}\|By\|_v=\|A\|_v+\|B\|_v$$

由(1)式确定的$\|A\|_v$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

6.3 算子范数

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

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

称向量范数和算子范数相容。

$A\in\mathbb{R}^{m\times n}, B\in\mathbb{R}^{n\times p}$

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

即算子范数服从乘法范数相容性。

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

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

A的每列绝对值之和的最大值,称为A的列范数

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

A的每行绝对值之和的最大值,称为A的行范数

3) $\|A\|_2=\max_{x\neq 0}\left(\|Ax\|_2/\|x\|_2\right)=\sqrt{\lambda_{\max}(A^TA)}$

$\lambda_{\max}(A^TA)$为A^TA的特征值的绝对值的最大值，称为A的2范数

6.4 条件数

- **定义：**非奇异矩阵A的**条件数**(condition number)：

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

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， x的相对误差比b的相对误差大得多。

- 由矩阵A定义的问题，称为适定问题或病态问题。

7.1 矩阵左逆

- 当一个矩阵X满足$XA = I$时，X被称为A的左逆;
- 当左逆存在时，则称A是可左逆的;
- 当左逆存在时，则A至少有一个左逆;

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

- 矩阵A是可左逆的，其左逆矩阵有两个

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

7.2 矩阵右逆

- 若矩阵X满足$AX = I$，则称X为矩阵A的右逆;
- 当右逆存在时，则称A是可右逆的;
- 当右逆存在时，则A至少有一个右逆;

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

- 矩阵B可右逆，以下矩阵都是B的右逆

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

7.3 性质

- 维度：一个大小为m × n的矩阵，其左逆或右逆的维度为n × m
- A的左逆为X当且仅当X^T是A^T的右逆

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

- A的右逆为X当且仅当X^T是A^T的左逆

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

7.4 矩阵的逆

- 如果矩阵A存在左逆和右逆，则左逆和右逆一定相等

$$XA = I, AY = I \quad => \quad X = XI = X(AY) = (XA)Y = Y$$

- 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
- 例子：

$$A = \begin{bmatrix} -1 & 1 & 3 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix}, \quad A^{-1} = \frac{1}{4}\begin{bmatrix} 2 & 4 & 1 \\ 0 & -2 & 1 \\ -2 & -2 & 0 \end{bmatrix}, AA^{-1} = A^{-1}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

7.6 线性方程组求解

- 如果矩阵A**可左逆**，假设X是矩阵A的左逆，则

$$Ax = b \quad => \quad x = XAx = Xb$$

至多一个解，如有解则$x = Xb$。

- 如果矩阵A**可右逆**，假设Y是矩阵A的右逆，则

为什么?

$$x = Yb \quad => \quad Ax = AYb = b$$

至少一个解，即$x = Yb$。

- 如果矩阵A**可逆的**，假设X是矩阵A的逆，则

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

唯一解。

7.7 非奇异矩阵

- 对于方阵A∈ $\mathbb{R}^{n\times n}$，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

- 此时矩阵A为非奇异矩阵，条件1与3，可得A为可逆矩阵。

非奇异矩阵⇔可逆矩阵。

7.8 证明1->2

- 证明：A可左逆，则A列向量线性无关
- 假设A的左逆是B，则

$$Ax = 0 \quad => \quad BAx = 0$$
$$=> \quad Ix = 0$$

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

$$Ax = x_1a_1 + x_2a_2 + \ldots + x_na_n = 0$$

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

如果A ∈ $\mathbb{R}^{m\times n}$，则有$m \geq n$！ 高或方的矩阵 $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

7.8 证明1->2

- 矩阵A有左逆X，则A列向量线性无关

如果A ∈ $\mathbb{R}^{m\times n}$，根据维度定理则有$m \geq n$!

$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ 高或方的矩阵

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

$$Ax = x_1a_1 + x_2a_2 + \ldots + x_na_n = b$$
$$Ay = y_1a_1 + y_2a_2 + \ldots + y_na_n = b$$
$$Ax - Ay = A(x - y) = 0 \Rightarrow x = y$$

当 b∈ $\mathbb{R}^m, b \notin \{y|y = Ax, x \in \mathbb{R}^n\}$ 时，无解！ $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$Ax = b$至多一个解，如有解则$x = Xb$。

7.9 证明2->3

- 证明：若方阵A列向量线性无关，则A可右逆
- 假设A∈ $\mathbb{R}^{n\times n}$为方阵且列向量线性无关 $A=[a_1,a_2,\cdots,a_n]$
- 则对于任意向量b ∈ $\mathbb{R}^n$，则向量组$[a_1,a_2,...,a_n,b]$ 线性相关，存在不全为0，使得以下等式成立

$$x_1a_1+x_2a_2+\cdots+x_na_n+x_{n+1}b=0$$

- 因为A列向量线性无关，则$x_{n+1}\neq 0$，即b是A列向量的线性组合；

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

- 存在向量c_1，...，$c_n\in\mathbb{R}^n$,使得$Ac_1=e_1,Ac_2=e_2$，...，$Ac_n=e_n$；
- 则矩阵$C=[c_1\ c_2\ ...c_n]$ 是矩阵A的右逆，$AC=I$.

7.13 非奇异Gram矩阵

引理：矩阵$A\in\mathbb{R}^{m\times n}$， $G=A^TA$，
矩阵A列向量线性无关“ ⇔ ”Gram矩阵G非奇异

证明：“ ⇒ ”假设矩阵A列向量线性无关，A^TA奇异。
则存在A^TAx=0，$x\neq 0$, 可得x^TA^TAx=$\|Ax\|_2^2$=0，
即Ax=0与列向量线性无关矛盾.

“ ⇐ ”假设A^TA非奇异, 矩阵A列向量线性相关。
则有Ax=0, $x\neq 0$, 可得A^TAx=0,
即A^TA是奇异矩阵。

7.14 伪逆

- 矩阵$A\in\mathbb{R}^{m\times n}$，当m≥n时，列向量线性无关，即A的Gram矩阵可逆；
- A的伪逆A^+定义：

$$A^+=(A^TA)^{-1}A^T$$

则A的伪逆存在。

- A^+为矩阵A的左逆：

$$A^+A=(A^TA)^{-1}A^TA=(A^TA)^{-1}(A^TA)=I$$

- 当A为方阵，伪逆等于矩阵的逆：

$$(A^TA)^{-1}A^T=A^{-1}A^{-T}A^T=A^{-1}$$

7.14 伪逆

- 矩阵 A ∈ $\mathbb{R}^{m\times n}$，当m≤n时，且行向量线性无关，则AA^T可逆。
- A^T列向量线性无关，即$(A^T)^TA^T$=AA^T非奇异；
- 定义伪逆：

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

- 伪逆$A^\dagger$为A的右逆

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

- 当A为方阵时，右逆等于矩阵的逆

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

7.15 伪逆

- 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^TA为非奇异矩阵

作业 11.2 11.3

- 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

8.1 正交单位向量

- 如果一个向量集合$a_1, a_2, \dots, a_n \in \mathbb{R}^m$满足：
  - 向量有单位范数：$\|a_i\| = 1$。
  - 向量之间相互正交：如果$i \neq j$，有$a_i^T a_j = 0$。

  则称这些向量是**标准正交**的。
- 例子：

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

8.2 标准列正交矩阵

- 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：

$$\begin{aligned} A^TA &= [a_1 \quad a_2 \quad \cdots \quad a_n]^T[a_1 \quad a_2 \quad \cdots \quad a_n] \\ &= \begin{bmatrix} {a_1}^Ta_1 & {a_1}^Ta_2 & \cdots & {a_1}^Ta_n \\ {a_2}^Ta_1 & {a_2}^Ta_2 & \cdots & {a_2}^Ta_n \\ \vdots & \vdots & \ddots & \vdots \\ {a_n}^Ta_1 & {a_n}^Ta_2 & \cdots & {a_n}^Ta_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I \end{aligned}$$

8.3 矩阵-向量乘积

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

$$\langle Ax, Ay \rangle = (Ax)^T(Ay) = x^TA^TAy = x^Ty$$

  - 保持原范数：

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

  - 保持原距离：

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

  - 保持原角度：

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

8.4 左可逆性

- 如果矩阵$A \in \mathbb{R}^{m\times n}$有标准正交列，则：
  - A是左可逆的，其左逆为A^T，根据定义：

$$A^T A = I$$

  - A有线性无关的列向量：

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

  - A是高的或者方的，即$m \geq n$。

列向量$a_1, a_2, ..., a_n \in \mathbb{R}^m$，维度定理$n \leq m$

8.5 正交矩阵

- 定义：所有列两两相互正交的方形实矩阵称为**正交矩阵。**
- 正交矩阵满足非奇异性，即如果方形矩阵A是正交的，则：
  - A是可逆的，左逆等于右逆，且它的逆为A^T：

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

  - A^T也是一个正交矩阵。
  - A的行是标准正交的，即范数为1且相互正交。
- 注意：如果$A \in \mathbb{R}^{m\times n}$有标准正交列以及$m > n$，则$AA^T \neq I$。

8.6 置换矩阵

- 让$\pi = (\pi_1, \pi_2, \cdots, \pi_n)$为$(1,2, ..., n)$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：

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

- Ax是x的一个置换：$Ax = (x_{\pi_1}, x_{\pi_2}, \cdots, x_{\pi_n})$。
- A在每一行和每一列中都有一个等于1的元素。
- 置换矩阵满足正交性，即所有置换矩阵都是正交的：
  - $A^T A = I$，因为A的每一行有一个元素等于1：

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

  - $A^T = A^{-1}$是逆置换矩阵。

8.6 置换矩阵例子

- 让$\pi = (\pi_1, \pi_2, \cdots, \pi_n)$为$(1,2, ..., n)$的一个重新排序的排列。
- 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：

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

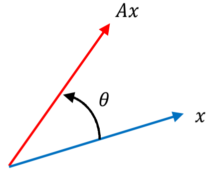
- A^T是与置换相关的置换矩阵：

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

8.7 平面旋转

- 在一个平面的旋转可以用矩阵表示为：

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



- 在$\mathbb{R}^n$的坐标平面上旋转：例如，

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

描述了在$\mathbb{R}^3$中$(x_1,\ x_3)$平面的旋转。

8.8 反射算子

- 反射算子(reflector)：一个矩阵的形式为：

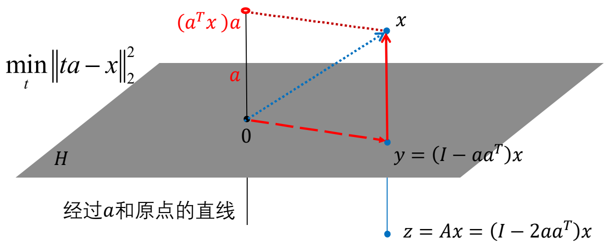
$$A=I-2aa^T$$

其中，向量a 满足$\|a\|_2=1$。

- 性质：
  - 反射矩阵(reflector matrix)是对称的 $A^T=A$
  - 反射矩阵(reflector matrix)是正交的

$$A^TA=\left(I-2aa^T\right)\left(I-2aa^T\right)=I-4aa^T+4aa^Taa^T=I$$

8.8.1 反射算子的几何解释



- $H=\{u|a^Tu=0\}$是与a正交的向量的(超)平面。
- 如果$\|a\|_2$=1, x在H上的投影为：

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

Gram-Schmidt正交算法

- x关于超平面H的对称点z由其与反射算子的乘积给出：

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

8.8.2 练习

- 假设$\|a\|_2=1$；给出x在$H=\{u|a^Tu=0\}$上的投影为：

$$y=x-\left(a^Tx\right)a$$

- 1、证明：$y\in H$。

$$a^Ty=a^T\left(x-a\left(a^Tx\right)\right)=a^Tx-\left(a^Ta\right)\left(a^Tx\right)=a^Tx-a^Tx=0$$

- 2、考虑任意$z\in H(z\neq y)$，证明$\|x-z\|_2^2>\|x-y\|_2^2$。

$$\begin{aligned}\|x-z\|_2^2&=\|x-y+y-z\|_2^2\\&=\|x-y\|_2^2+2(x-y)^T(y-z)+\|y-z\|_2^2\\&=\|x-y\|_2^2+2\left(a^Tx\right)a^T(y-z)+\|y-z\|_2^2\\&=\|x-y\|_2^2+\|y-z\|_2^2\qquad\left(\text{因为}a^Ty=a^Tz=0\right)\\&>\|x-y\|_2^2\end{aligned}$$

8.12 值域范围

- 一个向量集合张成的空间是其所有线性组合的集合：

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

- 矩阵$A \in \mathbb{R}^{m\times n}$的范围为其列向量张成的空间：

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

- 例子：

$$range\left(\begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 0 & -1 \end{bmatrix}\right) = \left\{\begin{bmatrix} x_1 \\ x_1 + 2x_2 \\ -x_2 \end{bmatrix} \middle| x_1, x_2 \in \mathbb{R}\right\}$$

8.13 值域投影

- 假设矩阵$A \in \mathbb{R}^{m\times n}$具有标准正交列,向量$Ax$与b的最短距离：

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

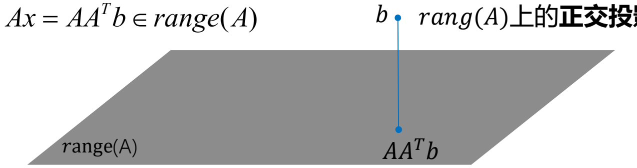
$$f(x) = \|Ax - b\|_2^2 = (Ax - b)^T (Ax - b) = x^T A^T A x - 2x^T A^T b + b^T b$$

$$= x^T x - 2x^T A^T b + b^T b, \quad \because A^T A = I$$

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

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

AA^Tb称为向量$b \in \mathbb{R}^m$在$rang(A)$上的**正交投影。**



- $\hat{x} = A^T b$满足$\|A\hat{x} - b\| < \|Ax - b\|$，对于所有$x \neq \hat{x}$。

8.13.1 验证

- b到$range(A)$内任意点Ax的距离的平方和为：

$$\|Ax - b\|_2^2 = \|A(x - \hat{x}) + A\hat{x} - b\|_2^2 \quad (\text{其中}\hat{x} = A^T b)$$

$$= \|A(x - \hat{x})\|_2^2 + \|A\hat{x} - b\|_2^2 + 2(x - \hat{x})^T A^T (A\hat{x} - b)$$

$$= \|A(x - \hat{x})\|_2^2 + \|A\hat{x} - b\|_2^2$$

$$= \|(x - \hat{x})\|_2^2 + \|A\hat{x} - b\|_2^2$$

$$\geq \|A\hat{x} - b\|_2^2$$

当且仅当$x = \hat{x}$，等号成立。

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

$$\|A(x - \hat{x})\|_2^2 = (A(x - \hat{x}))^T A(x - \hat{x}) = (x - \hat{x})^T A^T A(x - \hat{x}) = (x - \hat{x})^T (x - \hat{x})$$
$$= \|x - \hat{x}\|_2^2$$

8.15 Unitary矩阵

- 定义：列正交的方形复数矩阵称为**酉矩阵。**

- 酉矩阵的逆：

$$\left.\begin{matrix} A^H A = I \\ A\text{是方的} \end{matrix}\right\} \quad \Rightarrow \quad AA^H = I$$

- 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。

8.16 离散傅里叶变换矩阵

- 离散傅里叶变换矩阵W：($\omega = e^{2\pi j/n}$，$j = \sqrt{-1}$)

$$W = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \cdots & \omega^{-(n-1)} \\ 1 & \omega^{-2} & \omega^{-4} & \cdots & \omega^{-2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{-(n-1)} & \omega^{-2(n-1)} & \cdots & \omega^{-(n-1)(n-1)} \end{bmatrix}$$

- 矩阵$(1/\sqrt{n})W$是酉矩阵：

$$\frac{1}{n}W^H W = \frac{1}{n}WW^H = I$$

- W的逆$W^{-1} = (1/n)W^H$。
- n维向量x的离散傅里叶反变换是$W^{-1}x = (1/n)W^H x$。

8.17 DFT矩阵的Gram矩阵

证明$W^H W = nI$。 作业 11.4

- W的共轭转置为：

$$W^H = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{1} & \omega^{2} & \cdots & \omega^{n-1} \\ 1 & \omega^{2} & \omega^{4} & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

- Gram矩阵的第i,j个元素为：

$$(W^H W)_{ij} = 1 + \omega^{i-j} + \omega^{2(i-j)} + \cdots + \omega^{(n-1)(i-j)}$$

$$\left(W^H W\right)_{ii} = n, \quad \left(W^H W\right)_{ij} = \frac{\omega^{n(i-j)} - 1}{\omega^{i-j} - 1} = 0 \text{ 如果} i \neq j$$

最后一步因为$\omega^n = 1$。

9.2 前向回代

- 问题：当A是具有非零对角元素的下三角矩阵时，解$Ax = b$。
- 前向回代(Forward Substitution)算法求解：

$$x_1 = b_1/A_{11}$$

$$x_2 = (b_2 - A_{21}x_1)/A_{22}$$

$$x_3 = (b_3 - A_{31}x_1 - A_{32}x_2)/A_{33}$$

$$\vdots$$

$$x_n = \left(b_n - A_{n1}x_1 - A_{n2}x_2 - \cdots - A_{n,n-1}x_{n-1}\right)/A_{nn}$$

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

- 时间复杂度：1+3+5+⋯+(2n-1)=n^2 flops

9.4 三角矩阵的逆矩阵

- 对角元素非零的三角矩阵A是非奇异的，即：

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

- A的逆可以通过逐列解方程AX = I 来计算得到：

$$A[x_1 \quad x_2 \quad \cdots \quad x_n] = [e_1 \quad e_2 \quad \cdots \quad e_n]$$

(x_i是矩阵X的第i列向量)

$$x_1 = b_1/A_{11}$$

$$x_2 = (b_2 - A_{21}x_1)/A_{22}$$

$$x_3 = (b_3 - A_{31}x_1 - A_{32}x_2)/A_{33}$$

$$\vdots$$

$$x_n = (b_n - A_{n1}x_1 - A_{n2}x_2 - \cdots - A_{n,n-1}x_{n-1})/A_{nn}$$

- 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∈ $\mathbb{R}^{n\times n}$逆的复杂度：

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

9.5 QR分解

- 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

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

Gram-Schmidt(正交化): $\tilde{q}_i = a_i - (q_1^T a_i)q_1 - \ldots - (q_{i-1}^T a_i)q_{i-1} \quad R_{ji} = (q_j^T a_i)$

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

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

- 对角元素R_{ii}是非零的。
- 若$R_{ii} < 0$，改变$R_{ii}, \cdots, R_{in}$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。

9.5.1 矩阵的QR分解

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

$$A = QR$$

- ***Q*因子**:
  - $Q \in \mathbb{R}^{m\times n}$具有标准正交列($Q^T Q = I$)
  - 如果 A是方阵(m=n)，则Q是正交的 ($Q^T Q = QQ^T = I$)
- ***R*因子**:
  - $R \in \mathbb{R}^{n\times n}$的上三角矩阵
  - R是非奇异的(对角元素是非零的)

9.5.1 矩阵的QR分解

- QR分解步骤:
  - 1. 设矩阵A的列向量依次为$a_1, a_2, \cdots, a_n$，由于A为非奇异矩阵，则列向量线性无关;
  - 2. 对列向量$a_1, a_2, \ldots, a_n$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$q_1, q_2, \ldots, q_n$，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$A = QR$ 。

9.5.1 矩阵中的QR分解

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

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

9.6 QR分解和(伪)逆

- 具有线性无关列向量的矩阵A的伪逆为：

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

- 将A的伪逆表示为QR因子：

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

- 对于方阵非奇异矩阵A，其逆为：

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

9.8 值域上的投影

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

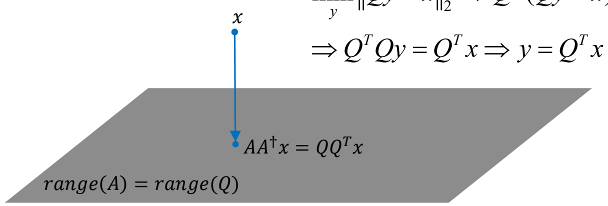
$$AA^{\dagger}=QRR^{-1}Q^{T}=QQ^{T}$$

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

- $QQ^{T}x$是x在Q值域上的投影：

$$\min_{y}\|Qy-x\|_2^2 \Rightarrow Q^{T}(Qy-x)=0$$

$$\Rightarrow Q^{T}Qy=Q^{T}x \Rightarrow y=Q^{T}x$$



9.11.2 时间复杂度

Gram–Schmidt方法第k次循环的复杂度：

- $a_k \in \mathbb{R}^m$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。$\tilde{q}_k = a_k - a_k^T q_1 q_1 - a_k^T q_2 q_2 - \cdots - a_k^T q_{k-1} q_{k-1}$
- 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：

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

9.11.3 数值实验

- GS的MATLAB代码：

```
[m, n] = size(A);
Q = zeros(m,n);
R = zeros(n,n);
for k = 1:n
    R(1:k-1,k) = Q(:,1:k-1)' * A(:,k);
    v = A(:,k) - Q(:,1:k-1) * R(1:k-1,k);
    R(k,k) = norm(v);
    Q(:,k) = v / R(k,k);
end;
```

- 矩阵A 的构造： $A = USV$,其中U和V是正交矩阵，S是对角矩阵，即：

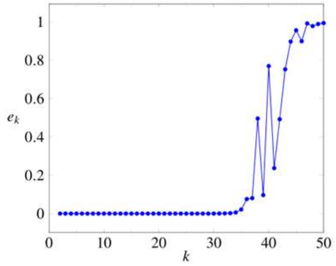
$$S_{ii} = 10^{-10(i-1)/(n-1)},\quad i = 1,\cdots,n.$$

- 把GS应用到一个大小为m = n = 50的方形矩阵A上。

9.11.3 数值实验

- 图中显示了q_k与前面列之间的正交性的偏差：

$$e_k = \max_{1\le i<k} \left| q_i^T q_k \right|, \quad k = 2, \ldots, n$$



- 注：失去正交性是由于浮点数存储的舍入误差。

9.12 Householder算法

- **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；
- 计算一个“完整的”QR因数分解：

$$A = [Q \quad \tilde{Q}] \begin{bmatrix} R \\ 0 \end{bmatrix}, \quad [Q \quad \tilde{Q}] \quad \text{是正交的矩阵}$$

- 完整的Q因子被构造成正交矩阵的乘积：

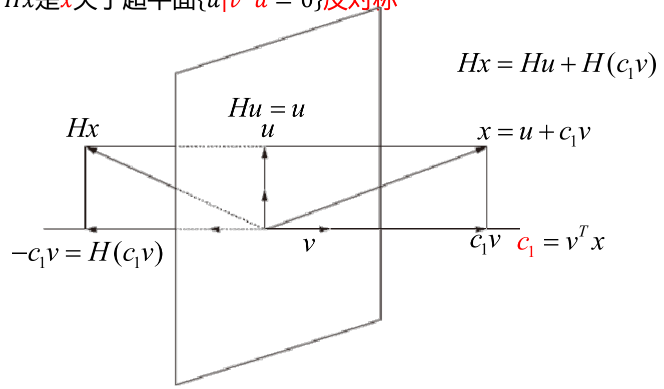
$$[Q \quad \tilde{Q}] = H_1 H_2 \cdots H_n$$

每个$H_i \in \mathbb{R}^{m\times m}$是对称的正交的“反射算子”(reflector)。

9.12.1 反射算子

$$H = I - 2vv^T \qquad \text{其中} \|v\|_2 = 1$$

- Hx是x关于超平面$\{u | v^T u = 0\}$反对称



9.12.1反射算子

$$H = I - 2vv^T \qquad \text{其中} \|v\|_2 = 1$$

- Hx是x关于超平面$\{u | v^T u = 0\}$反对称
- H是对称：$H^T = H$
- H是正交：$H^T H = I$
- 矩阵向量积Hx能化简为：

$$Hx = x - 2(v^T x)v$$

如果v和x的长度是p，复杂度是4p flops。

9.12.2 构造反射算子

给定非零p维向量$y = (y_1, y_2, \dots, y_p)$，定义：

$$w = \begin{bmatrix} y_1 + sign(y_1)\|y\|_2 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}, \quad v = \frac{w}{\|w\|_2}, \quad \begin{matrix} sign(y_1) = 1, y_1 \ge 0; \\ sign(y_1) = -1, y_1 < 0 \end{matrix}$$

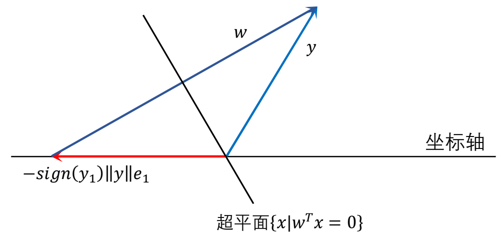
- 向量w满足：

$$\|w\|_2^2 = w^T w = 2\left(\|y\|_2^2 + |y_1|\|y\|_2\right) = 2y^T\left(y + sign(y_1)\|y\|_2 e_1\right) = 2y^T w$$

- 反射算子$H = I - 2vv^T = I - 2\frac{ww^T}{\|w\|_2^2}$将$y$映射为：

$$Hy = y - \frac{2(w^T y)}{\|w\|_2^2} w = y - w = -sign(y_1)\|y\|_2 e_1 = \begin{bmatrix} -sign(y_1)\|y\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

9.12.3 几何意义



- 关于超平面$\{x|w^T x = 0\}$，其法向量：

$$v = \frac{w}{\|w\|_2}, w = y + sign(y_1)\|y\|_2 e_1$$

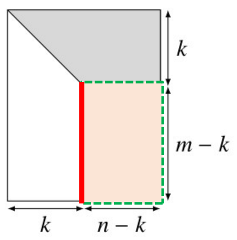
反射算子将y映射到向量$-sign(y_1)\|y\|_2 e_1$。

9.12.5 Householder算法

- 下面的算法用$\begin{bmatrix} R \\ 0 \end{bmatrix}$来代替$A \in \mathbb{R}^{m\times n}$
- Householder算法步骤：(k从1到n)
  - 1. 令$y = A_{k:m,k} \in \mathbb{R}^{m-k+1}$，计算向量$v_k$；

    $w = y + sign(y_1)\|y\|_2 e_1, \quad v_k = \frac{1}{\|w\|_2} w$

  - 2.将$A_{k:m,k:n} \in \mathbb{R}^{(m-k+1)\times(n-k+1)}$与反射算子$I - 2v_k v_k^T$相乘：

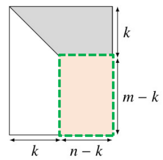


$$A_{k:m,k:n} := A_{k:m,k:n} - 2v_k\left(v_k^T A_{k:m,k:n}\right)$$

9.12.5 Householder算法注解

- 在步骤2中，将$A_{k:m,k:n}$与反射算子$I - 2v_k v_k^T$相乘：

$$\left(I - 2v_k v_k^T\right) A_{k:m,k:n} = A_{k:m,k:n} - 2v_k\left(v_k^T A_{k:m,k:n}\right)$$



- 等价于用$H_k \in \mathbb{R}^{m\times m}$乘以$A$：

$$H_k = \begin{bmatrix} I & 0 \\ 0 & I - 2v_k v_k^T \end{bmatrix} = I - 2\begin{bmatrix} 0 \\ v_k \end{bmatrix}\begin{bmatrix} 0 \\ v_k \end{bmatrix}^T$$

I	0
0	$I - 2v_k v_k^T$

- 算法将下列矩阵来代替A：

$$\begin{bmatrix} R \\ 0 \end{bmatrix}$$

返回(return)向量$v_1, \dots, v_n$, 其中**v_k的长度为m−k+1**。

9.12.6 例子

R的**第一列**:

$$A=\begin{bmatrix}-1&-1&1\\1&3&3\\-1&-1&5\\1&3&7\end{bmatrix}$$

■ 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$I_4-2v_1v_1^T$和A的乘积代替A:

$$A:=\left(I_4-2v_1v_1^T\right)A=\begin{bmatrix}2&4&2\\0&4/3&8/3\\0&2/3&16/3\\0&4/3&20/3\end{bmatrix}$$

9.12.6 例子

R的**第二列**:

$$A=\begin{bmatrix}2&4&2\\0&4/3&8/3\\0&2/3&16/3\\0&4/3&20/3\end{bmatrix}$$

■ 计算将$A_{2:4,2}$映射到e_1乘积的反射器:

$$y=\begin{bmatrix}4/3\\2/3\\4/3\end{bmatrix},\quad w=y+\|y\|_2 e_1=\begin{bmatrix}10/3\\2/3\\4/3\end{bmatrix},\quad v_2=\frac{1}{\|w\|_2}w=\frac{1}{\sqrt{30}}\begin{bmatrix}5\\1\\2\end{bmatrix}$$

■ 用$I_3-2v_2v_2^T$和$A_{2:4,2:3}$的乘积代替$A_{2:4,2:3}$:

$$A:=\begin{bmatrix}1&0\\0&I_3-2v_2v_2^T\end{bmatrix}A=\begin{bmatrix}2&4&2\\0&-2&-8\\0&0&16/5\\0&0&12/5\end{bmatrix}$$

9.12.6 例子

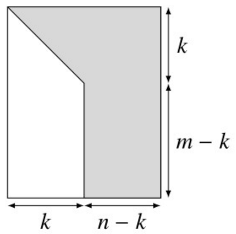
■ 最终结果:

$$\begin{aligned}H_3H_2H_1A&=\begin{bmatrix}I_2&0\\0&I_2-2v_3v_3^T\end{bmatrix}\begin{bmatrix}I_1&0\\0&I_3-2v_2v_2^T\end{bmatrix}\left(I_4-2v_1v_1^T\right)A\\&=\begin{bmatrix}I_2&0\\0&I_2-2v_3v_3^T\end{bmatrix}\begin{bmatrix}I_1&0\\0&I_3-2v_2v_2^T\end{bmatrix}\begin{bmatrix}2&4&2\\0&4/3&8/3\\0&2/3&16/3\\0&4/3&20/3\end{bmatrix}\\&=\begin{bmatrix}I_2&0\\0&I_2-2v_3v_3^T\end{bmatrix}\begin{bmatrix}2&4&2\\0&-2&-8\\0&0&16/5\\0&0&12/5\end{bmatrix}\\&=\begin{bmatrix}2&4&2\\0&-2&-8\\0&0&-4\\0&0&0\end{bmatrix}\end{aligned}$$

9.12.7 时间复杂度

$$\left(I-2v_kv_k^T\right)A_{k:m,k:n}=A_{k:m,k:n}-2v_k\left(v_k^TA_{k:m,k:n}\right)$$

Householder方法第k次循环的复杂度:



■ $v_k^TA_{k:m,k:n}$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$v_1,\ldots,v_n$的总复杂度:

$$\sum_{k=1}^{n}4(m-k+1)(n-k+2)\approx\int_0^n 4(m-t)(n-t+1)\,dt$$

$$\approx 2mn^2-\frac{2}{3}n^3\ \text{flops}$$

10.2 QR分解求解线性方程组

- 使用QR分解求解线性方程组$Ax=b$，矩阵$A\in\mathbb{R}^{n\times n}$为非奇异矩阵
  - 1.首先对A进行QR分解，得到 $A=QR$
  - 2.计算$y=Q^Tb$
  - 3.通过回代法求解$Rx=y$
- 复杂度：$2n^3+3n^2\approx 2n^3$ flops
  - QR分解复杂度:$2n^3$
  - 矩阵向量乘法： $2n^2$
  - 回代法： n^2

10.2应用QR分解

- 计算非奇异矩阵$A\in\mathbb{R}^{n\times n}$的逆$A^{-1}$，通过$AX=I$，即$QRX=I$.

$$X=[x_1,x_2,\cdots,x_n],x_i\in\mathbb{R}^n,i=1,\cdots,n.$$

$$I=[e_1,e_2,\cdots,e_n],e_i\in\mathbb{R}^n,i=1,\cdots,n.$$

$$QRX=I\Rightarrow RX=Q^TI$$

$$Rx_1=Q^Te_1,Rx_2=Q^Te_2,\cdots,Rx_n=Q^Te_n$$

- 复杂度：$2n^3+n^3\approx 3n^3$ flops
  - QR分解复杂度:$2n^3$
  - 回代法：一次回代 n^2，则n次回代n^3

10.3 LU分解

- $A=LU$分解
  - L为下三角矩阵并且对角线元素全为1，U为上三角矩阵。

$$A=\begin{pmatrix} a_{11} & \cdots & a_{1k} & \cdots & a_{1n}\\ \vdots & \ddots & \vdots & & \vdots\\ a_{k1} & \cdots & a_{kk} & \cdots & a_{kn}\\ \vdots & & \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{nk} & \cdots & a_{nn}\end{pmatrix}=LU$$

$$L=\begin{pmatrix} 1 & 0 & \cdots & 0\\ l_{21} & 1 & \ddots & 0\\ \vdots & \vdots & \ddots & 0\\ l_{n1} & l_{n2} & \cdots & 1\end{pmatrix}\qquad U=\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n}\\ 0 & u_{22} & \cdots & u_{2n}\\ \vdots & 0 & \ddots & \vdots\\ 0 & 0 & 0 & u_{nn}\end{pmatrix}$$

10.3 LU分解

$$A=\begin{pmatrix} a_{11} & \cdots & a_{1r} & \cdots & a_{1n}\\ \vdots & \ddots & \vdots & & \vdots\\ a_{r1} & \cdots & a_{rr} & \cdots & a_{rn}\\ \vdots & & \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{nr} & \cdots & a_{nn}\end{pmatrix}=\begin{pmatrix} 1 & 0 & 0 & \cdots & 0\\ \vdots & \ddots & 0 & \ddots & 0\\ l_{r1} & \cdots & 1 & \ddots & \vdots\\ \vdots & & \vdots & \ddots & 0\\ l_{n1} & \cdots & l_{nr} & \cdots & 1\end{pmatrix}\cdot\begin{pmatrix} u_{11} & \cdots & u_{1r} & \cdots & u_{1n}\\ 0 & \ddots & \vdots & & \vdots\\ 0 & 0 & u_{rr} & \cdots & u_{rn}\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & u_{nn}\end{pmatrix}$$

U的第一行元素u_{1j}为

$$a_{1j}=u_{1j}\Rightarrow u_{1j}=a_{1j},j=1,\cdots,n.$$

L的第一列元素l_{i1}为

$$l_{i1}=\frac{a_{i1}}{u_{11}}\quad i=2,3,\cdots,n.$$

U第r行主对角线以右元素u_{rj}

$$u_{rj}=a_{rj}-\sum_{k=1}^{r-1}l_{rk}u_{kj}$$

$$j=r,\cdots,n.$$

L第r列主对角线以下元素l_{ir}

$$l_{ir}=\frac{a_{ir}-\sum_{k=1}^{r-1}l_{ik}u_{kr}}{u_{rr}}$$

$$i=r+1,\cdots,n.$$

10.4 LU复杂度

- 求解$Ax = b$，A为非奇异矩阵
  - 1. 对矩阵A进行LU分解$\left(\frac{2}{3}n^3\ flops\right)$
  - 2.回代法：求解$Ly\ = b(n^2\ flops)$
  - 3.回代法：求解$Ux\ = y\ (n^2\ flops)$
- 复杂度：$\frac{2}{3}n^3 + 2n^2 \approx \frac{2}{3}n^3$ flops
- LU算法为求解方程组$Ax = b$的标准解法

10.7 例子

$$A = \begin{bmatrix} 8 & 2 & 9 \\ 4 & 9 & 4 \\ 6 & 7 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & L_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

- 计算U的第一行和L的第一列

$$(u_{11}, u_{12}, u_{13}) = (8, 2, 9) \qquad (l_{21}, l_{31}) = (\frac{1}{2}, \frac{3}{4})$$

- 然后计算U的第二行和L的第二列

$$u_{rj} = a_{rj} - \sum_{k=1}^{r-1} l_{rk} u_{kj}$$

$$u_{22} = a_{22} - l_{21}u_{12} = 8 \qquad l_{32} = \frac{a_{32} - l_{31}u_{12}}{u_{22}} = \frac{11}{16}$$

$$u_{23} = a_{23} - l_{21}u_{13} = -\frac{1}{2}$$

$$l_{ir} = \frac{a_{ir} - \sum_{k=1}^{r-1} l_{ik} u_{kr}}{u_{rr}}$$

- 最后计算U的第三行

$$u_{33} = a_{33} - l_{31}u_{13} - l_{32}u_{23} = -\frac{83}{32}$$

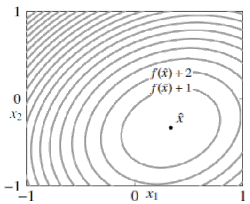
11.2 最小二乘法

给定A ∈ $\mathbb{R}^{m\times n}$, b ∈ $\mathbb{R}^m$,求解$x \in \mathbb{R}^n$让目标函数最小：

$$\min_x \|Ax - b\|_2^2 = \min_x \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right)^2$$

例如：$f(x) = \|Ax - b\|_2^2$,

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



$$f(x) = \|Ax - b\|_2^2 = (2x_1 - 1)^2 + (-x_1 + x_2)^2 + (\ 2x_2 + 1)^2$$

$$\frac{\partial f}{\partial x_1} = 10x_1 - 2x_2 - 4 \qquad \frac{\partial f}{\partial x_2} = -2x_1 + 10x_2 + 4$$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = 0 \Rightarrow \quad \hat{x} = \left(\frac{1}{3}, -\frac{1}{3}\right)^T$$

11.7 正规方程

$$A^T Ax = A^T b$$

- 最小二乘法问题的正规方程
- 系数矩阵A^TA是A的Gram矩阵
- 等价于“$\nabla f(x) = 0, f(x) = \|Ax - b\|_2^2$”
- 最小二乘法问题所有的解都满足正规方程

- 如果A的列线性无关，则
  - A^TA为非奇异矩阵
  - 正规方程此时有唯一解

11.8 QR分解求解

- 若$A \in \mathbb{R}^{m\times n}$的列向量线性无关，则存在A=QR分解，$Q \in \mathbb{R}^{m\times n}$，$R \in \mathbb{R}^{n\times n}$最小二乘法问题的解：

$$\begin{aligned}\hat{x} &= (A^T A)^{-1} A^T b = ((QR)^T (QR))^{-1} (QR)^T b \\ &= (R^T Q^T QR)^{-1} R^T Q^T b \\ &= (R^T R)^{-1} R^T Q^T b \\ &= R^{-1} Q^T b\end{aligned}$$

- 算法复杂度：
  - 首先对A进行QR分解$A = QR(2mn^2\ flops)$
  - 计算矩阵向量乘积$d = Q^T b(2mn\ flops)$
  - 通过回代求解$Rx = d(n^2\ flops)$
  - 复杂度：$2mn^2\ flops$

11.9 例子

$$A = \begin{bmatrix} 3 & -6 \\ 4 & -8 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} -1 \\ 7 \\ 0 \end{bmatrix}$$

- 1. 首先对A进行QR分解

$$Q = \begin{bmatrix} 3/5 & 0 \\ 4/5 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 5 & -10 \\ 0 & 1 \end{bmatrix}$$

- 2.计算$d = Q^T b$=（5，0）
- 3.求解$Rx = d$

$$\begin{bmatrix} 5 & -10 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

- 4.解的$x_1 = 1, x_2 = 0$

11.10 梯度下降法

- 设函数$f(x)$可微，根据泰勒公式，在$x^{(k)}$的一阶公式为

$$f(x^{(k+1)}) = f(x^{(k)}) + \left\langle \nabla f(x^{(k)}), x^{(k+1)} - x^{(k)} \right\rangle + o\left(\left\| x^{(k+1)} - x^{(k)} \right\|\right).$$

- 如果$\left\|x^{(k+1)} - x^{(k)}\right\|_2$足够小，则有

$$f(x^{(k+1)}) - f(x^{(k)}) \approx \left\langle \nabla f(x^{(k)}), x^{(k+1)} - x^{(k)} \right\rangle.$$

- 根据柯西不等式，$\left|\left\langle \nabla f(x^{(k)}), x^{(k+1)} - x^{(k)} \right\rangle\right| \le \left\|\nabla f(x^{(k)})\right\|_2 \left\|x^{(k+1)} - x^{(k)}\right\|_2$，

$$\left\langle \nabla f(x^{(k)}), x^{(k+1)} - x^{(k)} \right\rangle \ge -\left\|\nabla f(x^{(k)})\right\|_2 \left\|x^{(k+1)} - x^{(k)}\right\|_2,$$

当 $x^{(k+1)} - x^{(k)} = -\alpha_k \nabla f(x^{(k)})$，$\alpha_k > 0$ 时，等式成立。

迭代公式：$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f(x^{(k)})$.　　$f(x^{(k+1)}) < f(x^{(k)})$.

11.10 梯度下降法

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2,\quad A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$$

- 令$f(x) = \frac{1}{2}\|Ax - b\|_2^2$，则$f$为凸函数，并有$\nabla f(x) = A^T(Ax - b)$
- 则$A^T A \in \mathbb{R}^{n\times n}$，“列向量线性相关”导致其不可逆或n非常大！
- 通过梯度下降法迭代求解，$x^{(k+1)} = x^{(k)} - \alpha^{(k)} A^T\left(Ax^{(k)} - b\right)$
- 为了估计$\alpha^{(k)}$，通过线性搜索估计：

$$\alpha^{(k)} = \underset{\alpha\in\mathbb{R}}{\operatorname{argmin}} f(x^{(k)} - \alpha A^T(Ax^{(k)} - b)).$$

即$\alpha^{(k)}$**是最优步长。令**$g(\alpha) = f(x^{(k)} - \alpha A^T(Ax^{(k)} - b))$**是关于**$\alpha$**的凸函数，则有**

$$\min_{\alpha} g(\alpha) \Rightarrow g'(\alpha) = 0 \Rightarrow \alpha^{(k)} = \frac{\left\|A^T(Ax^{(k)} - b)\right\|_2^2}{\left\|AA^T(Ax^{(k)} - b)\right\|_2^2}.$$

11.10 梯度下降法

$$\min_{x\in\mathbb{R}^n}\frac{1}{2}\|Ax-b\|_2^2,\quad A\in\mathbb{R}^{m\times n}, b\in\mathbb{R}^m$$

梯度下降法求解最小二乘法

作业 12.3 12.8

1. 初始$x^{(0)}$
2. $for\quad k=0,1,\cdots,do$

$$p^{(k)}=A^T(Ax^{(k)}-b)$$

$$\alpha^{(k)}=\frac{\|p^{(k)}\|_2^2}{\|Ap^{(k)}\|_2^2}$$

$$x^{(k+1)}=x^{(k)}-\alpha^{(k)}p^{(k)}$$

当满足迭代条件，则退出.

13.1 KKT条件：例子

$$\max_x\{f(x)=20x_1+10x_2\}$$

$$s.t.\quad g_1(x)=x_1^2+x_2^2\le 1, g_2(x)=x_1+2x_2\le 2,$$

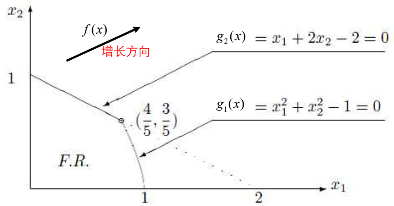
$$g_3(x)=-x_1\le 0, g_4(x)=-x_2\le 0$$

$$\nabla_x L(x,u)=\nabla_x f(x)-u_1\nabla_x g_1(x)-u_2\nabla_x g_2(x)-u_3\nabla_x g_3(x)-u_4\nabla_x g_4(x)=0, u_j\ge 0$$

$$\nabla_x f(x)=\begin{bmatrix}20\\10\end{bmatrix},\nabla_x g_1(x)=\begin{bmatrix}2x_1\\2x_2\end{bmatrix},\nabla_x g_2(x)=\begin{bmatrix}1\\2\end{bmatrix},\nabla_x g_3(x)=\begin{bmatrix}-1\\0\end{bmatrix},\nabla_x g_4(x)=\begin{bmatrix}0\\-1\end{bmatrix}.$$

检测$g_1(x)$边界　　检测$g_2(x)$边界　　检测$g_3(x)$边界　　检测$g_4(x)$边界

$\nabla_x f(x)=u_1\nabla_x g_1(x)$, $\nabla_x f(x)\ne u_2\nabla_x g_2(x)$, $\nabla_x f(x)\ne u_3\nabla_x g_3(x)$, $\nabla_x f(x)\ne u_4\nabla_x g_4(x)$.



$u_2=u_3=u_4=0$, $u_1\ne 0$.

13.1 KKT条件：例子

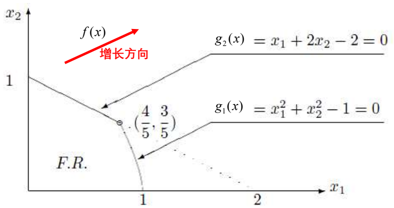
$$\max_x\{f(x)=20x_1+10x_2\}$$

$$s.t.\quad g_1(x)=x_1^2+x_2^2\le 1, g_2(x)=x_1+2x_2\le 2,$$

$$g_3(x)=-x_1\le 0, g_4(x)=-x_2\le 0$$

即$\nabla_x f(x)=u_1\nabla_x g_1(x)$, $\begin{bmatrix}20\\10\end{bmatrix}=u_1\begin{bmatrix}2x_1\\2x_2\end{bmatrix}\Rightarrow x_1=2x_2$，代入$g_1(x)=x_1^2+x_2^2-1=5x_2^2-1=0$.

由于$x_2\ge 0$, 可得$\begin{bmatrix}x_1\\x_2\end{bmatrix}=\frac{\sqrt{5}}{5}\begin{bmatrix}2\\1\end{bmatrix}$，$u_1=5\sqrt{5}$.



13.2 最小范数优化问题

$$\min_x\frac{1}{2}\|x\|_2^2$$

$$s.t.\quad Cx=d,$$

引入拉格朗日函数：

$$L(x,\lambda)=\frac{1}{2}\|x\|_2^2-\lambda^T(Cx-d)$$

对拉格朗日函数求导：

$$\nabla_x L(x,\lambda)=x-C^T\lambda=0\Rightarrow x=C^T\lambda$$

"矩阵C行线性无关" 可得：

$$Cx=CC^T\lambda=d\Rightarrow\lambda=(CC^T)^{-1}d$$

则有：

$$\hat{x}=C^T\lambda=C^T(CC^T)^{-1}d=C^\dagger d.$$

13.2 验证最优解

- 1.首先证明解$\hat{x}$满足等式 $\hat{x}=C^T\lambda=C^T(CC^T)^{-1}d=C^\dagger d.$

$$C\hat{x}=CC^T(CC^T)^{-1}d=d$$

- 2.在$Cx=d$，$x\neq\hat{x}$的情况下，

$$\begin{aligned}\hat{x}^T(x-\hat{x})&=d^T(CC^T)^{-1}C(x-\hat{x})\\&=d^T(CC^T)^{-1}(Cx-C\hat{x})\\&=d^T(CC^T)^{-1}(d-d)\\&=0\end{aligned}$$

- 3.在$Cx=d$，$x\neq\hat{x}$的情况下，证明$\|x\|_2^2>\|\hat{x}\|_2^2$

$$\begin{aligned}\|x\|_2^2&=\|x-\hat{x}+\hat{x}\|_2^2\\&=\|\hat{x}\|_2^2+2\hat{x}^T(x-\hat{x})+\|x-\hat{x}\|_2^2\\&=\|\hat{x}\|_2^2+\|x-\hat{x}\|_2^2\\&>\|\hat{x}\|_2^2\end{aligned}$$

13.2 QR分解求解

- 对矩阵$C^T\in\mathbb{R}^{n\times p}$进行QR分解，$C^T=QR$

$$\begin{aligned}\hat{x}&=C^T(CC^T)^{-1}d\\&=QR(R^TQ^TQR)^{-1}d\\&=QR(R^TR)^{-1}d\\&=QR^{-T}d\end{aligned}$$

- 复杂度：
  - 1.矩阵$C^T=QR$分解，$C^T=QR(2np^2\ flops)$
  - 2.回代法求解$R^Tz=d$ $(p^2\ flops)$
  - 3.计算$\hat{x}=Qz(2np\ flops)$
  - 总复杂度：$(\approx 2np^2\ flops)$

13.2 例子

$$C=\begin{bmatrix}1&-1&1&1\\1&0&1/2&1/2\end{bmatrix},\quad d=\begin{bmatrix}0\\1\end{bmatrix}$$

- 对矩阵C^T进行QR分解，$C^T=QR$

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

- 通过回代法求解$R^Tz=d$

$$\begin{bmatrix}2&0\\1&1/\sqrt{2}\end{bmatrix}\begin{bmatrix}z_1\\z_2\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix}$$

- 解$z_1=0,\ z_2=\sqrt{2}$
- 得$\hat{x}=Qz=(1,1,0,0)$

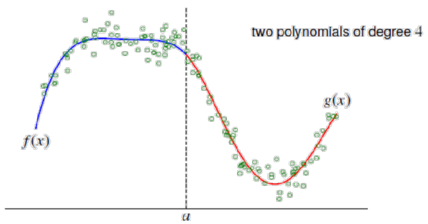
13.3 分段多项式拟合问题

- 两个多项式$f(x),g(x)$对于样本点$(x_1,y_1),\ldots,(x_N,y_N)$进行拟合

$$f(x_i)\approx y_i,x_i\le a$$
$$g(x_i)\approx y_i,x_i>a$$

- 拟合要求：函数值和导数值必须在分段位置a连续

$$f(a)=g(a),f'(a)=g'(a)$$



13.3分段多项式拟合问题

- 假设样本点$(x_1,y_1),...,(x_N,y_N)\in\mathbb{R}^2$，有$x_1,...,x_M\le a$, $x_{M+1},...,x_N>a$;
- d阶多项式$f(x)=\theta_1+\theta_2x+...+\theta_dx^{d-1}$,
  $g(x)=\theta_{d+1}+\theta_{d+2}x...+\theta_{2d}x^{d-1}$:

$$A=\left[\begin{array}{cccccccc}1 & x_1 & \cdots & x_1^{d-1} & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_M & \cdots & x_M^{d-1} & 0 & 0 & \cdots & 0\\ \hline 0 & 0 & \cdots & 0 & 1 & x_{M+1} & \cdots & x_{M+1}^{d-1}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & 1 & x_N & \cdots & x_N^{d-1}\end{array}\right],\theta=\begin{bmatrix}\theta_1\\ \vdots\\ \theta_d\\ \theta_{d+1}\\ \vdots\\ \theta_{2d}\end{bmatrix},b=\begin{bmatrix}y_1\\ \vdots\\ y_M\\ y_{M+1}\\ \vdots\\ y_N\end{bmatrix},$$

参数θ需要确定?

$$A\theta\approx b$$

$$\min_{\theta}\|A\theta-b\|_2^2$$

13.3分段多项式拟合问题

- 假设样本点$(x_1,y_1),...,(x_N,y_N)\in\mathbb{R}^2$，有$x_1,...,x_M\le a$, $x_{M+1},...,x_N>a$;
- d阶多项式$f(x)=\theta_1+\theta_2x+...+\theta_dx^{d-1}$,
  $g(x)=\theta_{d+1}+\theta_{d+2}x...+\theta_{2d}x^{d-1}$:
- 约束条件: $f(a)=g(a),f'(a)=g'(a)$

$$\theta=\begin{bmatrix}\theta_1\\ \vdots\\ \theta_d\\ \theta_{d+1}\\ \vdots\\ \theta_{2d}\end{bmatrix}$$

$$C=\left[\begin{array}{cccc|cccc}1 & a & \cdots & a^{d-1} & -1 & -a & \cdots & -a^{d-1}\\ \hline 0 & 1 & \cdots & (d-1)a^{d-2} & 0 & -1 & \cdots & -(d-1)a^{d-2}\end{array}\right],d=\begin{bmatrix}0\\0\end{bmatrix}$$

$$C\theta=d$$

13.3分段多项式拟合问题

- 假设样本点$(x_1,y_1),...,(x_N,y_N)\in\mathbb{R}^2$，有$x_1,...,x_M\le a$, $x_{M+1},...,x_N>a$;
- d阶多项式$f(x)=\theta_1+\theta_2x+...+\theta_dx^{d-1}$,
  $g(x)=\theta_{d+1}+\theta_{d+2}x...+\theta_{2d}x^{d-1}$:
- 约束条件: $f(a)=g(a),f'(a)=g'(a)$

$$\min_{\theta}\quad \sum_{i=1}^{M}(f(x_i)-y_i)^2+\sum_{i=M+1}^{N}(g(x_i)-y_i)^2$$
$$s.t.\quad f(a)=g(a),f'(a)=g'(a)$$

$$\min_{\theta}\|A\theta-b\|_2^2$$
$$s.t.\quad C\theta=d$$

13.4 先验假设

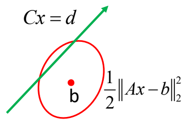
$$\min_{x}\|Ax-b\|_2^2$$
$$s.t.\quad Cx=d$$

- 1. 堆叠矩阵**列线性无关**

$$\begin{bmatrix}A\\C\end{bmatrix}\in\mathbb{R}^{(m+p)\times n}$$

- 2.矩阵C$\in\mathbb{R}^{p\times n}$行线性无关
- 假设“条件1”是一个比$\boldsymbol{A}$可右逆更弱的条件
- $p\le n\le m+p$

13.4 最小二乘法约束KKT条件



$$\min_x \frac{1}{2}\|Ax-b\|_2^2$$

$$s.t. \quad Cx = d$$

引入拉格朗日函数:

$$L(x,z) = \frac{1}{2}\|Ax-b\|_2^2 - z^T(d-Cx), z \in \mathbb{R}^p$$

对拉格朗日函数求导:

$$\nabla_x L(x,z) = A^T(Ax-b) + C^T z = 0$$

$$\nabla_z L(x,z) = Cx - d = 0$$

$$\begin{bmatrix} A^TA & C^T \\ \mathrm{C} & 0 \end{bmatrix}\begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} A^Tb \\ d \end{bmatrix}$$

13.4 KKT最优条件

- 令$\hat{x}$是约束优化问题的解，则有

$$\begin{bmatrix} A^TA & C^T \\ \mathrm{C} & 0 \end{bmatrix}\begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} A^Tb \\ d \end{bmatrix}, \begin{bmatrix} \hat{x} \\ z \end{bmatrix} \in \mathbb{R}^{n+p}$$

- 优化条件Karush-Kuhn-Tucker(KKT)等式
- 特殊情况
  - 最小二乘法问题：当$p = 0$时，即为正规方程$A^TA\hat{x} = A^Tb$
  - 最小范数问题：当$A = I,\ b = 0$时，可以推导得到$C\hat{x} = b, \hat{x} + C^Tz = 0$

13.4 证明最优解

$$\begin{bmatrix} A^TA & C^T \\ \mathrm{C} & 0 \end{bmatrix}\begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} A^Tb \\ d \end{bmatrix}$$

- 假设x满足$Cx = d$, $(\hat{x}, z)$满足KKT等式

$$\begin{aligned}\|Ax-b\|_2^2 &= \|A(x-\hat{x}) + A\hat{x} - b\|_2^2 \\ &= \|A(x-\hat{x})\|_2^2 + \|A\hat{x}-b\|_2^2 + 2(x-\hat{x})^T A^T(A\hat{x}-b) \\ &= \|A(x-\hat{x})\|_2^2 + \|A\hat{x}-b\|_2^2 - 2(x-\hat{x})^T C^T z \\ &= \|A(x-\hat{x})\|_2^2 + \|A\hat{x}-b\|_2^2 \\ &\geq \|A\hat{x}-b\|_2^2\end{aligned}$$

- 第三行$A^TA\hat{x} + C^Tz = A^Tb$
- 第四行$Cx = C\hat{x} = d \Longrightarrow (x-\hat{x})^TC^T = (Cx - C\hat{x})^T = 0$
- $\hat{x}$是唯一性，因为假设矩阵A列线性无关，C行线性无关，即

$$A^TA(\hat{x}-x) = 0 \Rightarrow x = \hat{x}$$

$$C^T(\hat{z}-z) = 0 \Rightarrow \hat{z} = z$$

$$A^TA\hat{x} + C^Tz = A^Tb$$

13.4 非奇异性

- 如果矩阵A列线性无关，C行线性无关，则矩阵

$$\begin{bmatrix} A^TA & C^T \\ \mathrm{C} & 0 \end{bmatrix}$$

为非奇异矩阵

- 证明 $\begin{bmatrix} A^TA & C^T \\ \mathrm{C} & 0 \end{bmatrix}\begin{bmatrix} x \\ z \end{bmatrix} = 0 \Rightarrow x^T(A^TAx + C^Tz) = 0, Cx = 0$

$$\Rightarrow \|Ax\|_2^2 + (Cx)^T z = \|Ax\|_2^2 = 0, \because Cx = 0$$

$$\Rightarrow Ax = 0, Cx = 0$$

$$\Rightarrow x = 0 \quad \text{A列线性无关}$$

由于$A^T\mathrm{A}x + C^T\mathrm{z} = 0$，则当$x = 0$时，有$\mathrm{z} = 0$。

C行线性无关

13.4 奇异性

- 如果矩阵A列线性无关和 C行线性无关不同时成立，则矩阵

$$\begin{bmatrix} A^T A & C^T \\ \mathrm{C} & 0 \end{bmatrix}$$

为奇异矩阵

- 如果A列线性相关，则存在$x \neq 0$,使得$Ax = 0$,则

$$\begin{bmatrix} A^T A & C^T \\ \mathrm{C} & 0 \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} = 0$$

- 如果C行线性相关，则存在$z \neq 0$,使得$C^T z = 0$,则

$$\begin{bmatrix} A^T A & C^T \\ \mathrm{C} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \mathrm{z} \end{bmatrix} = 0$$

- 因此该矩阵为奇异矩阵

13.5 LU分解求解

$$\begin{bmatrix} A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} A^T b \\ d \end{bmatrix}$$

$A \in \mathbb{R}^{m\times n}, C \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^m, d \in \mathbb{R}^p$

算法：

- 计算$H = A^T A (2mn^2\ flops)$。
- 计算$c = A^T b (2mn\ flops)$。
- 用LU分解法求解下列线性方程$((2/3)(p+n)^3\ flops)$：

$$\begin{bmatrix} H & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}$$

- 复杂度为：$2mn^2 + (2/3)(p+n)^3\ \ flops$。

13.6 QR分解求解

由于$\hat{x}$满足$C\hat{x} = d$，则有$C^T C\hat{x} = C^T d$，可得

$$L(x,z) = \frac{1}{2}\|Ax-b\|_2^2 - z^T(d - Cx)$$

KKT条件：$\nabla_x L(x,z) = A^T(A\hat{x}-b) + C^T z + C^T C\hat{x} - C^T d = 0$

$$\Rightarrow (A^T A + C^T C)\hat{x} + C^T(z-d) = A^T b$$

$$\nabla_z L(x,z) = C\hat{x} - d = 0$$

- 令$w = z - d$，KKT条件写成：

$$\begin{bmatrix} A^T A + C^T C & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ w \end{bmatrix} = \begin{bmatrix} A^T b \\ d \end{bmatrix}$$

- 假设1保证了$A^T A + C^T C$是非奇异的，即存在以下QR因子分解：

$\begin{bmatrix} A \\ C \end{bmatrix}$列向量无关，则 $\begin{bmatrix} A \\ C \end{bmatrix} = QR = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R = \begin{bmatrix} Q_1 R \\ Q_2 R \end{bmatrix}$

13.6 QR分解的解

$$\begin{bmatrix} A^T A + C^T C & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ w \end{bmatrix} = \begin{bmatrix} A^T b \\ d \end{bmatrix}$$

代入QR分解，可得：

$$\begin{bmatrix} R^T R & R^T Q_2^T \\ Q_2 R & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ w \end{bmatrix} = \begin{bmatrix} R^T Q_1^T b \\ d \end{bmatrix}$$

- 将第一个方程两边乘R^{-T}和并令变量$y = R\hat{x}$相乘，可得：

$$\begin{bmatrix} I & Q_2^T \\ Q_2 & 0 \end{bmatrix} \begin{bmatrix} y \\ w \end{bmatrix} = \begin{bmatrix} Q_1^T b \\ d \end{bmatrix}$$

- 矩阵$C = Q_2 R \Rightarrow Q_2 = CR^{-1}$, Q_2具有行线性无关：

$$Q_2^T u = R^{-T} C^T u = 0 \quad \Rightarrow \quad C^T u = 0 \quad \Rightarrow \quad u = 0$$

因为C行线性无关的(假设2)。

13.6 QR分解的解

利用Q_2^T的QR分解来求解：

$$\left[\begin{array}{cc} I & Q_2^T \\ \hline Q_2 & 0 \end{array}\right]\left[\begin{array}{c} y \\ \hline w \end{array}\right]=\left[\begin{array}{c} Q_1^T b \\ \hline d \end{array}\right]$$

- 方程第1行可得$y = Q_1^T b - Q_2^T w$，并代入第二行：

$$Q_2 y = d \Rightarrow Q_2 Q_2^T w = Q_2 Q_1^T b - d$$

- 用QR分解$Q_2^T = \tilde{Q}\tilde{R}$来解这个关于$w$的方程：

$$\tilde{R}^T \tilde{R} w = \tilde{R}^T \tilde{Q}^T Q_1^T b - d$$

上式可以简化为：

$$\tilde{R} w = \tilde{Q}^T Q_1^T b - \tilde{R}^{-T} d$$

13.6 QR分解综述

$$\begin{bmatrix} A^T A + C^T C & C^T \\ C & 0 \end{bmatrix}\begin{bmatrix} \hat{x} \\ w \end{bmatrix}=\begin{bmatrix} A^T b \\ d \end{bmatrix} \qquad \tilde{R} w = \tilde{Q}^T Q_1^T b - \tilde{R}^{-T} d$$

算法过程：

- 1. 计算两个QR分解：

$$\begin{bmatrix} A \\ C \end{bmatrix}=\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R, \qquad Q_2^T = \tilde{Q}\tilde{R}$$

- 2. 用前代法求解$\tilde{R}^T u = d$，计算$c = \tilde{Q}^T Q_1^T b - u$。
- 3. 用回代法求解$\tilde{R} w = c$，计算$y = Q_1^T b - Q_2^T w$。
- 4. 用回代法计算$R\hat{x} = y$。

复杂度：QR分解有$2(p + m)n^2 + 2np^2$次flops。

13.6 复杂度QR vs LU

假设$p < n$：

作业 16.11

- LU复杂度：$2mn^2 + (2/3)(p + n)^3 < 2mn^2 + (16/3)n^3$
- QR复杂度：$2(p + m)n^2 + 2np^2 < 2mn^2 + 4n^3$
- 稳定性：QR分解避免直接计算 $A^T A$。