《微积分A2》第八讲

教师　杨利军

清华大学数学科学系

2020年03月11日

考虑线性代数方程组 $Ax = b$，其中 A 为 $m \times n$ 矩阵，$x \in \mathbb{R}^n$，$b \in \mathbb{R}^m$. 当 $m > n$ 时，方程组的形状如下

$$\begin{bmatrix} \quad & \quad & \quad \\ \\ \\ \\ \\ \end{bmatrix} \begin{bmatrix} \quad \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \quad \\ \\ \\ \\ \\ \end{bmatrix},$$

即方程的个数大于变量的个数. 因此方程组可能无解. 回忆方程组 $Ax = b$ 有解，当且仅当右端向量 b 可以表为矩阵 A 的 n 个列向量 $A_i \in \mathbb{R}^m$ 的线性组合，即 $b = \sum_{i=1}^{n} x_i A_i$.

假设方程组 $Ax = b$ 无解，即 $\|Ax - b\| > 0, \forall x \in \mathbb{R}^n$，这里 $\|Ax - b\|$ 代表向量 $b \in \mathbb{R}^m$ 与向量 Ax 的距离. 因此我们可以寻求向量 $x^* \in \mathbb{R}^n$，使得 $\|Ax^* - b\|$ 尽可能地小.

Definition

我们称 $x^* \in \mathbb{R}^n$ 是线性代数方程组 $Ax = b$ 的最小二乘解，如果 $x = x^*$ 是函数 $\|Ax - b\|$ 在全空间 $\mathbb{R}^n$ 上的最小值点，即

$$\|Ax^* - b\| = \inf(\min)\{\|Ax - b\|, x \in \mathbb{R}^n\}$$

Theorem

定理: 向量 $x^* \in \mathbb{R}^n$ 是线性代数方程组 $Ax = b$ 的最小二乘解, 当且仅当 x^* 是方程组 $A^TAx^* = A^Tb$ 的解.

定理的一个代数证明, 可参见《线性代数与几何》(下), 第二版, 俞正光,鲁子群,林润亮编著, 清华大学出版社, 2015, page 65-67. 稍后我们将利用极值理论给出上述定理的一个分析证明.

注: 方程组 $A^TAx^* = A^Tb$ 常称作 $Ax = b$ 的正规方程组.

正规方程组恒有解

Lemma

对任意 $m \times n$ 实矩阵 A 和任意 m 维向量 $b \in \mathbb{R}^m$, 正规方程组 $A^T Ax = A^T b$ 恒有解.

证明: 方程组 $A^T Ax = A^T b$ 有解, 当且仅当 $\text{rank}(A^T A) = \text{rank}(A^T A|A^T b)$. 显然 $\text{rank}(A^T A) \le \text{rank}(A^T A|A^T b)$. 相反的不等式则由 $\text{rank}(A^T A|A^T b) = \text{rank}[A^T(A|b)] \le \text{rank}(A^T) = \text{rank}(A^T A)$ 得到. 证毕. $\square$

注: 等式 $\text{rank}(A^T) = \text{rank}(A^T A)$ 由如下一般结论得到: 对于任意矩阵 B, 线性代数方程组 $Bx = 0$ 与 $B^T Bx = 0$ 同解.

证明：令 $f(x) \triangleq \|Ax-b\|^2$, $x \in \mathbb{R}^n$. 以下我们证明函数 $f(x)$ 在 $\mathbb{R}^n$ 上存在全局最小值点 $x^* \in \mathbb{R}^n$, 并且全局最小值点 x^* 由正规方程组 $A^TAx = A^Tb$ 所确定.

Step 1°: 注意 $f(x) = (Ax-b, Ax-b) = (Ax-b)^T(Ax-b)$

$= (x^TA^T - b^T)(Ax-b) = x^TA^TAx - x^TA^Tb - b^TAx + b^Tb$

$= x^TA^TAx - 2x^TAb + b^Tb = \sum_{i,j=1}^n b_{ij}x_ix_j - 2\sum_{k=1}^n c_kx_k + d$,

这里 $B = [b_{ij}] \triangleq A^TA$, $C = (c_1, \cdots, c_n)^T \triangleq Ab$, $d \triangleq b^Tb$.

Step 2°: 求 $f(x)$ 的临界点. 令

$$\frac{\partial f}{\partial x_p}=\frac{\partial}{\partial x_p}\left[\sum_{i,j=1}^{n} b_{ij}x_ix_j-2\sum_{j=1}^{n} c_jx_j+d\right]=2\sum_{j=1}^{n} b_{pj}x_j-2c_p=0,$$

其中 $p=1,\cdots,n$, 即 $2(Bx-C)=2(A^TAx-A^Tb)=0$. 这表明 x^* 是 $f(x)$ 的临界点, 当且仅当 $A^TAx^*=A^Tb$.

注: 关于求导公式 $\frac{\partial}{\partial x_p}\sum_{i,j=1}^{n} b_{ij}x_ix_j=2\sum_{j=1}^{n} b_{pj}x_j$ 的证明:

$$\frac{\partial}{\partial x_p}\sum_{i,j=1}^{n} b_{ij}x_ix_j=\frac{\partial}{\partial x_p}\sum_{i=1}^{n} x_i\sum_{j=1}^{n} b_{ij}x_j=\sum_{i=1}^{n}\frac{\partial}{\partial x_p}\left[x_i\sum_{j=1}^{n} b_{ij}x_j\right]$$

$$=\sum_{j=1}^{n} b_{pj}x_j+\sum_{i=1}^{n} b_{ip}x_i=2\sum_{j=1}^{n} b_{pj}x_j.$$

证明续二

Step 3°: 证明每个临界点都是全局最小点. 设 x^* 是 $f(x)$ 的临界点, 即 $A^TAx^* = A^Tb$. 由多元函数的二阶 Taylor 公式得

$$f(x) - f(x^*) = \nabla f(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^T H(\xi)(x - x^*),$$

这里 $H(x)$ 记 $f(x)$ 的 Hesse 矩阵. 由前页的注可知

$$\frac{\partial}{\partial x_p}\sum_{i,j=1}^{n} b_{ij}x_ix_j = 2\sum_{j=1}^{n} b_{pj}x_j.$$

由此得

$$\frac{\partial^2}{\partial x_q \partial x_p}\sum_{i,j=1}^{n} b_{ij}x_ix_j = \frac{\partial}{\partial x_q}\left[2\sum_{j=1}^{n} b_{pj}x_j\right] = 2b_{pq}.$$

证明续三

因此 $f(x)$ 的 Hesse 矩阵为常数矩阵，即 $H(x) = 2B = 2A^TA$.

再注意到 $\nabla f(x^*) = 0$，故对 $\forall x \in \mathbb{R}^n$

$$f(x) - f(x^*) = (x - x^*)^T A^T A(x - x^*) = \|A(x - x^*)\|^2 \geq 0.$$

这就证明了每个临界点都是 $f(x)$ 的全局最小点．证毕． $\square$

注：当 $\mathrm{rank}(A) = n$ (此时 $m \geq n$) 时，n 阶矩阵 A^TA 满秩．于是方程组 $A^TAx = A^Tb$ 有唯一解．从而线性方程组 $Ax = b$ 有唯一一个最小二乘解.

例：设已知一组观测数据如下

x	1	2	3	4
y	1.3	1.8	2.2	2.9

在最小二乘的意义下，求最佳直线拟合方程，即求一个线性函数 $y = kx + b$，使得误差的平方和最小.

解: 设 $y = kx + b$ 为所求的直线方程. 我们考虑如下线性代数方程组

$$\begin{cases} k + b = 1.3 \\ 2k + b = 1.8 \\ 3k + b = 2.2 \\ 4k + b = 2.9 \end{cases} \quad \text{or} \quad \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} k \\ b \end{bmatrix} = \begin{bmatrix} 1.3 \\ 1.8 \\ 2.2 \\ 2.9 \end{bmatrix}$$

的最小二乘解. 记上述线性方程组的系数矩阵为 A, 记右端向量为 b. 根据最小二乘定理可知, 为求所需线性函数, 只需求解正规方程组 $A^TAx = A^Tb$. 简单计算得

例子续二

$$\mathbf{A}^{\mathsf{T}}\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix}$$

$$\mathbf{A}^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1.3 \\ 1.8 \\ 2.2 \\ 2.9 \end{bmatrix} = \begin{bmatrix} 23.1 \\ 7.2 \end{bmatrix}.$$

例子续三

根据上述计算结果得

$$\begin{bmatrix} k \\ b \end{bmatrix} = (A^T A)^{-1} A^T b$$

$$= \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 23.1 \\ 7.2 \end{bmatrix} = \begin{bmatrix} 0.52 \\ 0.75 \end{bmatrix}.$$

于是所求直线为 $y = 0.52x + 0.75$. 解答完毕.

条件极值(带约束的极值)问题

考虑条件极值问题

$$\begin{cases} \min(\max) & f(x_1, \cdots, x_n), \\ \text{s.t.} & g(x_1, \cdots, x_n) = k. \end{cases}$$

这里 s.t.= subject to 意为：受限制于，函数 $f, g : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ 均假设为 C^1 的，Ω 开. 先考虑情形 $n = 2$. 此时上述问题可写作

$$\begin{cases} \min(\max) & f(x, y), \\ \text{s.t.} & g(x, y) = k, \end{cases}$$

函数 $f, g : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ 均为 C^1 的，Ω 开.

考虑极值的必要条件. 设 (x_0, y_0) 是问题的解, 则 $g(x_0, y_0) = k$. 假设 $\nabla g(x_0, y_0) \neq (0, 0)$. 不妨设 $g_y(x_0, y_0) \neq 0$, 则由 IFT 知, 在点(x_0, y_0) 附近曲线 $g(x, y) = k$ 可表为 $y = \xi(x)$, $x \in J_\delta$, 其中 $J_\delta = (x_0 - \delta, x_0 + \delta)$, $\xi(x_0) = y_0$, $g(x, \xi(x)) \equiv k$, 且函数 $\xi(\cdot)$ 是 C^1 的. 记 $\hat{f}(x) \triangleq f(x, \xi(x))$, 则一元函数 $\hat{f}(x)$ 在 x_0 处有极值. 故 $\hat{f}'(x_0) = 0$, 即

$$f_x(x_0, y_0) + f_y(x_0, y_0)\xi'(x_0) = 0. \quad (*)$$

再对恒等式 $g(x, \xi(x)) \equiv k$ 求导得

$$g_x(x_0, y_0) + g_y(x_0, y_0)\xi'(x_0) = 0. \quad (**)$$

比较等式 (*) 和 (**) 可知 $\nabla f(x_0,y_0) \parallel \nabla g(x_0,y_0)$. 故存在常数 λ_0, 使得 $\nabla f(x_0,y_0) = \lambda_0 \nabla g(x_0,y_0)$. 因为已设 $\nabla g(x_0,y_0) \neq (0,0)$. 以下考察必要条件 $\nabla f(x_0) = \lambda_0 \nabla g(x_0)$ 的几何意义.

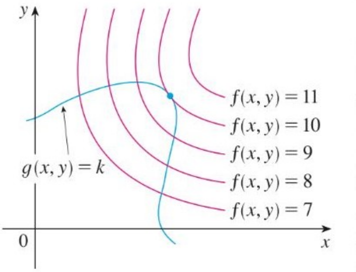
二维条件极值问题

$$\begin{cases} \min(\max) & f(x,y), \\ \text{s.t.} & g(x,y) = k, \end{cases}$$

可看作求函数 $f(x,y)$ 在水平线 $g(x,y)=k$ 上的极值, 比方说极大值. 故问题就是要求尽可能大的 c, 使得水平线 $f(x,y)=c$ 与 $g(x,y)=k$ 相交.

必要条件的几何解释，图示

如果在交点处，两条水平线不相切，那么适当增加 c 的值，还可以使得这两条水平线相交．因此如果参数 c 是一个极大值，使得水平线 $f(x,y)=c$ 与 $g(x,y)=k$ 相交，那么在交点处两条水平线必相切．因此交点处的法方向必平行，即 $\nabla f \parallel \nabla g$．如图．



上述二维条件极值的分析过程对一般 n 维的条件极值问题同样适用. 由此我们有如下定理, 其证明思想同二维情形.

Theorem

定理: 考虑条件极值问题

$$\begin{cases} \min(\max) & f(x_1, \cdots, x_n), \\ \text{s.t.} & g(x_1, \cdots, x_n) = k, \end{cases}$$

其中 $f, g : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ 均为 C^1 函数, Ω 开, 并设当 $g(x) = k$ 时, $\nabla g(x) \neq 0$. 若点 $x_0 \in \Omega$ 是问题的解, 则存在常数 λ_0, 使得 $\nabla f(x_0) = \lambda_0 \nabla g(x_0)$.

由于条件极值问题

$$\begin{cases} \min(\max) & f(x), \\ \text{s.t.} & g(x) = k, \end{cases}$$

的解 x_0 满足必要条件 $\nabla f(x_0) = \lambda_0 \nabla g(x_0)$，受此启发，我们定义函数

$$L(x, \lambda) \triangleq f(x) - \lambda[g(x) - k],$$

并称之为 Lagrange 函数，称 λ 为 Lagrange 乘子 (Lagrange multipliers).

条件极值的解法: Lagrange 乘子法

于是求解原条件极值问题 $\min(\max) f(x)$, s.t. $g(x) = k$, 可转化为求解无约束极值问题

$$(*) \quad \min(\max) L(x, \lambda), \quad x \in \Omega, \lambda \in \mathbb{R}.$$

因为条件极值每个解 $x_0 \in \Omega$, 对应着无约束问题 (*) 的一个临界点 (x_0, λ_0), 其中 $\nabla f(x_0) = \lambda_0 \nabla g(x_0)$. 故函数 $L(x, \lambda)$ 的临界点方程 $L_x = 0$, $L_\lambda = 0$, 即

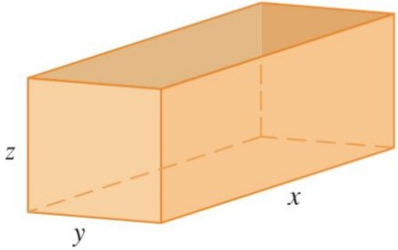
$$\begin{cases} \nabla f(x) = \lambda \nabla g(x), \\ g(x) = k \end{cases}$$

可用来来求解条件极值问题. 这种方法称为 Lagrange 乘子法.

问题: 假设我们有 12m^2 的铁皮, 要做成一个无盖的长方体的盒子, 问怎样确定长方体的尺寸, 可使得盒子的体积最大.

解: 设待求长方体的长宽高分别为 x, y, z, 则盒子的体积为 xyz.

所用的材料为 $2xz + 2yz + xy = 12$. 如图.



因此盒子问题是一个条件极值问题

$$\begin{cases} \max \quad xyz, \\ \text{s.t.} \quad 2xz + 2yz + xy = 12. \end{cases}$$

以下用 Lagrange 乘子法来求解. 为此记 $L(x, y, z, \lambda) = xyz$ $-\lambda(2xz + 2yz + xy - 12)$, 考虑函数 L 的临界点方程组

$$\begin{cases} L_x = yz - \lambda(2z + y) = 0, \\ L_y = xz - \lambda(2z + x) = 0, \\ L_z = xy - \lambda(2x + 2y) = 0, \\ L_\lambda = -(2xz + 2yz + xy - 12) = 0. \end{cases}$$

方程组可写为

$$\begin{cases} yz = \lambda(2z + y), \\ xz = \lambda(2z + x), \\ xy = \lambda(2x + 2y), \\ 2xz + 2yz + xy = 12. \end{cases}$$

以 x, y, z 分别依次乘以前三个方程得

$$\begin{cases} xyz = \lambda(2xz + xy), \\ xyz = \lambda(2yz + xy), \\ xyz = \lambda(2xz + 2yz). \end{cases}$$

盒子问题，续三

由此得 $2xz + xy = 2yz + xy = 2xz + 2yz$. 根据问题的实际背景，$x, y, z > 0$. 由此不难解得 $x = y = 2z$. 将其带入约束方程 $2xz + 2yz + xy = 12$，可解得 $z = 1$, $x = y = 2$，且 $\lambda = 1/2$. 这是极值问题唯一一个可能的解. 根据问题的实际背景可知问题的解存在. 因此可以断言 $(x, y, z) = (2, 2, 1)$ 时，盒子的体积最大，其最大体积为 $V = 2 \cdot 2 \cdot 1 = 4$. 后面我们还将证明这是一个极大值点.

盒子问题的另一解法：化为无约束极值问题

我们也可以将盒子问题，即条件极值问题

$$\begin{cases} \max \quad xyz, \\ \text{s.t.} \quad 2xz + 2yz + xy = 12. \end{cases}$$

化为无约束极值问题．由约束条件 $2xz + 2yz + xy = 12$ 可解出 $z = \frac{12-xy}{2(x+y)}$．将其带入目标函数 V 得

$$V = xyz = \frac{(12 - xy)xy}{2(x + y)}.$$

计算 $V(x, y)$ 偏导数得

另解续

$$V_x = \frac{y^2(12 - 2xy - x^2)}{2(x+y)^2}, \quad V_y = \frac{x^2(12 - 2xy - y^2)}{2(x+y)^2}.$$

解方程组 $V_x = V_y = 0$ 可得 $x^2 = y^2$. 由问题的实际背景可知 $x, y > 0$. 于是解得 $x = y$. 再由方程 $12 - 2xy - y^2 = 0$ 可知 $x = y = 2$. 最后由 $z = \frac{12-xy}{2(x+y)}$ 求得 $z = 1$, 从而得到一个临界点 $(2, 2, 1)$, 对应的体积为 $V(2, 2, 1) = 4$. 进一步我们可以计算函数 $V(x, y)$ 的 Hesse 矩阵 $H(x, y)$ 在点 $(2, 2)$ 处是一个负定矩阵. 因此临界点 $(2, 2)$ 是函数 $V(x, y)$ 的极大值点. 解答完毕.

例: 求 $f(x,y)=x^2+2y^2$ 在单位圆周 $x^2+y^2=1$ 上的最大值和最小值.

解: 用 Lagrange 乘子法来求解. 令

$$L(x,y,\lambda)=x^2+2y^2-\lambda(x^2+y^2-1).$$

其临界点方程组为

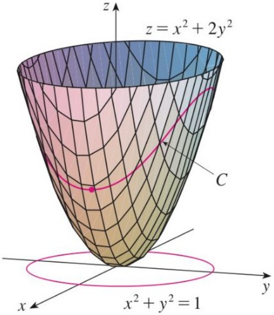
$$\begin{cases} 2x=2\lambda x, \\ 4y=2\lambda y, \\ x^2+y^2=1. \end{cases}$$

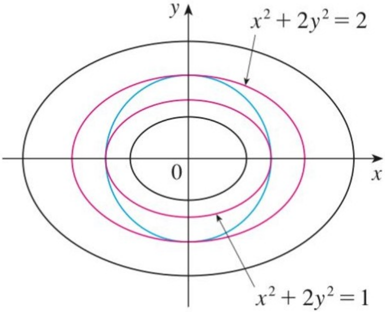
不难求得临界点 $(x,y,\lambda)=(0,\pm1,1),\ (\pm1,0,1)$.

由此得函数 f 可能的极值点为 $(x,y)=(0,\pm1)$ 和 $(\pm1,0)$. 计算 $f(x,y)$ 在这四个点上值可得 $f(0,\pm1)=2$, $f(\pm1,0)=1$. 于是可断言函数 $f(x,y)$ 在单位圆周 $x^2+y^2=1$ 上的最大值为 2, 且最大值点为 $(0,\pm1)$; 最小值为 1, 且最小值点为 $(\pm1,0)$. 解答完毕.

注: 对于上述条件极值问题, 我们很容易通过消元解除约束, 将条件极值问题化为无约束极值. 由 $x^2+y^2=1$ 可知 $f(x,y)=x^2+2y^2=1+y^2$. 由此可知在单位圆周 $x^2+y^2=1$ 上, 函数 f 的最大值为 2, 最小值为 1.

图示





例: 求 $f(x, y) = x^2 + 2y^2$ 在闭单位圆盘 $x^2 + y^2 \leq 1$ 上的最大值和最小值.

解: 对于这类问题的求解通常分为两步. 第一步求函数在开圆盘上的极值点. 这是无约束极值问题. 第二步求函数在区域边界的极值问题, 这是求解条件极值问题.

例子续

我们来求解上述例题. 先求函数 $f(x,y)=x^2+2y^2$ 的临界点. 令 $f_x=f_y=0$, 即 $2x=4y=0$. 由此得 f 唯一一个临界点原点 $(0,0)$, 且 $f(0,0)=0$. 比较函数 f 在单位圆周 $x^2+y^2=1$ 上的最大值和最小值可知, 函数 $f(x,y)=x^2+2y^2$ 在闭单位圆盘 $x^2+y^2\leq 1$ 上的最大值为 $f(0,\pm 1)=2$, 最小值为 $f(0,0)=0$. 解答完毕.

习题1.9 (page 93-95): 4(2), 7(2)(4), 8, 9(1)(3), 10(2),(4), 11, 13.

补充习题: 证明讲义第 33 页中的引理.

选作题: 设 $P=(p_1,\cdots,p_n)$: $\mathbb{R}^n\to\mathbb{R}^n$ 为一个二次多项式映射, 也就是说, 映射 P 的每个分量 $p_i=p_i(x_1,\cdots,x_n)$ 均为变量 $x_1,\cdots,x_n$ 的二次多项式, $i=1,\cdots,n$. 若映射 P 的 Jacobi 矩阵 $JP(x)=[\frac{\partial p_i}{\partial x_j}(x)]$ 在全空间 $\mathbb{R}^n$ 上处处非奇, 即 $\det JP(x)\neq 0$, $\forall x\in\mathbb{R}^n$, 证明映射 $P(\cdot)$ 是单射 (injective), 即 $P(x)\neq P(y)$, $\forall x,y\in\mathbb{R}^n$, $x\neq y$.

选作题注记

注一: 实际上，我们还可以证明结论：多项式映射 (不必是二次的) $P(\cdot)$, 如果它是单射的话, 那它一定是满射的，即对任意 $y \in \mathbb{R}^n$, 存在 $x \in \mathbb{R}^n$, 使得 $P(x) = y$. 这个结论现有的证明需要用到比较高深的代数几何知识. 不知是否存在比较初等的证明. 有兴趣的同学可以尝试一下. 初等的意思是只使用微积分和线性代数.

注二: 上述结论看起来似乎很特别, 但却是一个很重要的结论. 如果将上述结论再往前推进一步, 即当映射 $P(\cdot)$ 为三次多项式映射, 并且满足如下条件时,

$$\det JP(x) = \det \left[\frac{\partial p_i}{\partial x_j}\right](x) \equiv 1, \quad \forall x \in \mathbb{R}^n.$$

能够证明映射 $P(\cdot)$ 为单射, 那么也就证明了著名的 Jacobian conjecture (雅可比猜想) 成立. 关于这个猜想的详细情况, 建议有兴趣的同学在网上搜一搜.