高等运筹学教程

马良 张惠珍 刘勇 宁爱兵 编著

上海人民出版社

高等运筹学教程

马良 张惠珍 刘勇 宁爱兵 编著

上海人民出版社

图书在版编目(CIP)数据

高等运筹学教程/马良等编著.—上海:上海人民出版社,2015
ISBN 978-7-208-12873-6

Ⅰ.①高…　Ⅱ.①马…　Ⅲ.①运筹学-高等学校-教材　Ⅳ.①O22

中国版本图书馆 CIP 数据核字(2015)第 052422 号

责任编辑　曹怡波
封面装帧　张志全

高等运筹学教程

马　良　张惠珍　刘　勇　宁爱兵　编著

世纪出版集团
上海人民出版社出版
(200001　上海福建中路 193 号　www.ewen.co)
世纪出版集团发行中心发行　　常熟市新骅印刷有限公司印刷
开本 720×1000　1/16　印张 17.75　插页 4　字数 285,000
2015 年 5 月第 1 版　2015 年 5 月第 1 次印刷
ISBN 978-7-208-12873-6/O·2
定价 48.00 元

前　　言

高等运筹学是基础运筹学的后续课程，目前已成为高等院校管理类专业研究生的核心课程，也是许多理工类专业研究生重要的专业基础课程。

本书的撰写是在多年教学实践基础上，参考各种现行教材以及有关专业材料，集体合作完成的。在教材内容和形式的处理上，力求简明精练，着重方法和软件使用，适当增加趣味性，其中一些内容体现了作者近年来在教学与科研上的某些心得和成果。

本书主要包括：各种常用最优化方法及理论基础、组合优化基本理论和经典组合优化难题。结合实例具体讲解了高等运筹学的基本概念与重要算法，力求理论与应用的结合，强调算法的实现与软件的运用。

使用本书的读者假定已初步修习过基础运筹学课程，即已大致了解线性规划、整数规划、目标规划、动态规划、网络图论、决策论、对策论、排队论、库存论等基本知识，但并非要求全部、全面地熟练掌握。

本书部分内容曾作为教材为管理类硕士研究生和博士研究生使用多年，并得到上海市一流学科建设项目的资助，在此谨致谢意。同时，感谢老一辈学者为我们所奠定的基础以及提供的各类有关材料，感谢所有曾修读过该课程的学生及其有益反馈，也向所有被我们直接或间接引用文献资料的同行学者致以由衷的敬意。

鉴于我们学识有限，书中存在的谬误遗漏之处在所避免，敬请读者不吝指正与批评。

参加本书撰写和修订工作的还有王家桢（硕士），高岩教授对本书内容的取舍和编排提出了宝贵的意见。

本书可作为高等院校有关专业研究生和高年级本科生的高等运筹学或最优化方法等课程的教材，亦可供其他科技人员和实际工作者自学或参考。

作　者

2014 年 11 月 1 日

目　　录

第 4 章　无约束优化

第 5 章　有约束优化

第 6 章　组合优化与计算复杂性

第 7 章　旅行商问题

第 8 章　背包问题

第1章 数学基础

唯天下至诚，为能经纶天下之大经，
立天下之大本，知天地之化育。
——《礼记·中庸》

§1-1 非线性规划问题

一、概述

众所周知，线性规划的应用极为普遍，因其具有统一的线性数学结构而导致的通用求解方法也极为有效。然而，一旦目标函数或约束条件中出现了非线性函数，那么这种规划问题就变成了所谓的非线性规划问题。线性规划是目标函数与约束条件全部为线性表达式的数学规划，而非线性规划是目标函数与约束条件不全是线性表达式的数学规划。由于数学结构上的不规则性，非线性规划至今也没有找到可称为“通用”的有效算法。诚然，有时非线性规划可以转化为线性规划来研究，从而在一定条件下能求得最优解。但是，直接研究非线性规划的理论与算法则具有更为一般的意义。

由于现实世界的非线性性质，实际的最优化问题常常应更确切地归结为非线性规划问题。一般而言，解非线性规划问题要比解线性规划问题困难得多。由于非线性规划还没有适用于各种问题的一般算法，因此各个求解方法都有其特定的应用范围。

二、实例

例 1.1 （数据拟合问题）已知热敏电阻 R 与温度 t 的函数关系为

$$R = x_1 \exp\left(\frac{x_2}{t + x_3}\right)$$

并通过实验测得数据 t_i, $R_i(i=1, 2, \cdots, 15)$。试确定参数 x_1, x_2, x_3,使得偏差的平方和为最小。

解 按题意有

$$\min f(x_1, x_2, x_3)=\sum_{i=1}^{15}\left[R_i-x_1\exp\left(\frac{x_2}{t_i+x_3}\right)\right]^2$$

例 1.2 (投资决策问题)某企业有 n 个项目可供选择投资,并且至少要对其中一个项目投资。已知该企业拥有总资金 A 元,投资于第 $i(i=1, \cdots, n)$ 个项目需花资金 a_i 元,预计可收益 b_i 元。试选择最佳投资方案。

解 设投资决策变量为

$$x_i=\begin{cases}1, & \text{投资于第 } i \text{ 个项目}\\ 0, & \text{不投资第 } i \text{ 个项目}\end{cases}, (i=1, \cdots, n)$$

则投资总额为 $\sum_{i=1}^{n}a_ix_i$,投资总收益为 $\sum_{i=1}^{n}b_ix_i$。

因至少要对一个项目投资,且总投资金额不能超过总资金 A,故有限制条件

$$0<\sum_{i=1}^{n}a_ix_i\leqslant A$$

另外,由于 $x_i(i=1, \cdots, n)$ 只取 0 或 1,故有

$$x_i(1-x_i)=0, (i=1, \cdots, n)$$

最佳投资方案应是投资额最小而总收益最大的方案,所以该最佳投资决策问题归结为,在总资金以及决策变量(取 0 或 1)的限制条件下,极大化总收益和总投资之比。因而,其数学模型可以写为:

$$\max f(x_1, x_2, \cdots, x_n)=\frac{\sum_{i=1}^{n}b_ix_i}{\sum_{i=1}^{n}a_ix_i}$$

$$\text{s.t.}\quad\begin{cases}0<\sum_{i=1}^{n}a_ix_i\leqslant A\\ x_i(1-x_i)=0, (i=1, \cdots, n)\end{cases}$$

三、一般数学模型

非线性规划一般的数学模型可写成如下形式:

$$\min f(X) \tag{1.1}$$

$$\text{s.t.} \quad h_i(X) = 0,\ i = 1,\ 2,\ \cdots,\ p \tag{1.2}$$

$$g_j(X) \geqslant 0,\ j = 1,\ 2,\ \cdots,\ q \tag{1.3}$$

其中,$X = (x_1,\ x_2,\ \cdots,\ x_n)^T \in R^n$,即:目标函数 $f(X)$,约束函数 $h_i(x)(i = 1,\ 2,\ \cdots,\ p)$ 和 $g_j(X)(j = 1,\ 2,\ \cdots,\ q)$ 都是定义在 R^n 上的实函数,且其中至少有一个不是线性函数。(1.2)和(1.3)称为约束条件,满足所有约束条件的点 X 称为可行点或可行解。所有可行点的集合称为可行域,可用 $D = \{X \mid h_i(X) = 0,\ g_j(X) \geqslant 0,\ i = 1,\ 2,\ \cdots,\ p;\ j = 1,\ 2,\ \cdots,\ q\}$ 表示。

显然,利用可行域 D,问题(1.1)~(1.3)可表示为如下简洁的形式:

$$\min_{X \in D} f(X) \tag{1.4}$$

对优化问题(1.4)(或(1.1)~(1.3))的求解,是指在可行域内找一点 X^*,使得目标函数 $f(X)$在该点取得极小值,即

$$f(X^*) = \min_{X \in D} f(X)$$

称 X^* 为问题(1.4)(或(1.1)~(1.3))的最优解,而相应的目标函数 $f(X^*)$称为最优值。

由于

$$h_i(X) = 0 \Leftrightarrow \begin{cases} h_i(X) \geqslant 0 \\ -h_i(X) \geqslant 0 \end{cases},$$

故非线性规划模型又可改写为

$$\min f(X) \tag{1.5}$$

$$\text{s.t.} \quad g_j(X) \geqslant 0,\ j = 1,\ 2,\ \cdots,\ l \tag{1.6}$$

今后讨论时,将视需要而取(1.1)~(1.3)或(1.5)~(1.6)的形式。

由于 $\max f(X) = -\min[-f(X)]$, $g_j(X) \leqslant 0 \Leftrightarrow -g_j(X) \geqslant 0$, 故(1.1)~(1.3)或(1.5)~(1.6)的形式均具有一般性。

非线性规划问题最优解与线性规划问题最优解的不同之处在于:线性规划问

题的最优解 X^* 必在 R 边界特别是顶点上达到，而非线性规划问题的最优解 X^* 既可以在 R 边界也可以在 R 内部达到。

四、非线性优化问题的三种类型

(1) 无约束优化问题

当 $p=q=0$ 时，问题(1.1)～(1.3)为无约束优化问题。若无约束优化问题的目标函数 $f(X)$ 是定义在 R^n 上的可微函数，求其最优解的问题即为高等数学中的函数极值问题，求解方法可从如下含有 n 个变量 $x_1, x_2, \cdots, x_n$ 的非线性方程组

$$\begin{cases} f'_{x_1}(X)=0 \\ f'_{x_2}(X)=0 \\ \quad\cdots\cdots\cdots \\ f'_{x_n}(X)=0 \end{cases}$$

中解出驻点，然后验证这些驻点是不是最优解。

例 1.3 用铁板做一个体积为 3 m^3 的有盖长方体水箱，问如何取长、宽、高的尺寸，使所耗费的铁板最少？

解 设水箱的长宽分别为：x m 和 y m，则其高为 $\dfrac{3}{xy}$ m。该水箱所用铁板的面积为

$$\begin{aligned} f(x, y) &= 2\left(xy + y \cdot \frac{3}{xy} + x \cdot \frac{3}{xy}\right) \\ &= 2\left(xy + \frac{3}{x} + \frac{3}{y}\right) \end{aligned}$$

其中，$x>0, y>0$。

令

$$f'_x = 2\left(y - \frac{3}{x^2}\right) = 0$$

$$f'_y = 2\left(x - \frac{3}{y^2}\right) = 0$$

解该方程组，得

$$x = \sqrt[3]{3}, \; y = \sqrt[3]{3}$$

因此，当水箱的长、宽、高分别为：$\sqrt[3]{3}$ m、$\sqrt[3]{3}$ m、$\dfrac{3}{\sqrt[3]{3}\cdot\sqrt[3]{3}}=\sqrt[3]{3}$ m时，所耗费的铁板最省。

(2) 等式约束优化问题

当 $p\neq 0$ 且 $q=0$ 时，问题(1.1)～(1.3)为等式约束优化问题。该问题的求解通常采用 Lagrange 乘子法，将问题转化为如下求解无约束优化问题的形式：

$$\begin{aligned}&L(x_1,\ x_2,\ \cdots,\ x_n,\ \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_p)\\&=f(x_1,\ x_2,\ \cdots,\ x_n)-\sum_{i=1}^{p}\lambda_i h_i(x_1,\ x_2,\ \cdots,\ x_n)\end{aligned}$$

例 1.4 求表面积为 $2a^2(a>0)$ 体积最大的长方体体积。

解 设长方体的长、宽、高分别为 $x,\ y,\ z$，体积为 v，则问题就是在条件 $\varphi(x,\ y,\ z)=2(xy+yz+xz)-2a^2=0$（即 $xy+yz+xz-a^2=0$）下，求函数 $v=f(x,\ y,\ z)=xyz$ 的最大值。

由 Lagrange 乘子法，考虑函数

$$F(x,\ y,\ z)=xyz+\lambda(xy+yz+xz-a^2)$$

令

$$F'_x=yz+\lambda(y+z)=0$$

$$F'_y=xz+\lambda(x+z)=0$$

$$F'_z=xy+\lambda(x+y)=0$$

因为要求体积最大的长方体，故 $x,\ y,\ z$ 均为正数，由此可知 $\lambda\neq 0$。将上述三式分别乘以 $x,\ y,\ z$，并由 $xy+yz+xz-a^2=0$ 可得：

$$\begin{cases}xyz+\lambda(a^2-yz)=0\\xyz+\lambda(a^2-xz)=0\\xyz+\lambda(a^2-xy)=0\end{cases}$$

比较上述三式可得：

$$x=y=z$$

将该式代入 $xy+yz+xz-a^2=0$，便得：

$$x = y = z = \frac{\sqrt{3}}{3}a$$

因为由问题本身可知最大值一定存在，所以最大值就在这个可能的极值点取得，即，此表面积为 $2a^2$ 的长方体中以正方体的体积为最大，最大体积为$\frac{\sqrt{3}}{9}a^3$。

(3) 含有不等式约束的优化问题

当 $p = 0$ 且 $q \neq 0$ 时，问题(1.1)～(1.3)为含有不等式约束的优化问题。无论变量数多少，按经典极值方法求解就几乎不可能了。

§1-2　方向导数、梯度与 Taylor 展开

一、方向导数

定义 1.1　对于任意给定的实数 $\delta > 0$，满足不等式 $\| X - X^* \| < \delta$ 的点 X 的集合称为以 X^* 为中心、δ 为半径的邻域，记为

$$N_\delta(X^*) = \{X \mid \| X - X^* \| < \delta\}$$

其中，$\| X - X^* \| = \sqrt{\sum_{i=1}^{n} (x_i - x_i^*)^2}$ 为 X 与 X^* 在 R^n 中的距离。

定义 1.2　设函数 $z = f(X)$ 在点 X^* 处可微，e 是方向 l 上的单位向量，则称极限

$$\frac{\partial f(X^*)}{\partial l} = \lim_{t \to 0^+} \frac{f(X^* + te) - f(X^*)}{t} \tag{1.7}$$

为函数 $f(X)$在点 X^* 处沿方向 l 的方向导数，记作 $\frac{\partial f(X^*)}{\partial l}$。

定义 1.3　设函数 $z = f(X)$ 在点 X^* 处可微，若 $\frac{\partial f(X^*)}{\partial l} < 0$，则称 l 为在点 X^* 附近的下降方向；若 $\frac{\partial f(X^*)}{\partial l} > 0$，则称 l 为在点 X^* 附近的上升方向。

事实上，$X = X^* + te$ 是从 X^* 出发在方向 l 上的点，若 $\frac{\partial f(X^*)}{\partial l} < 0$，则由式(1.7)可知：$f(X^* + te) - f(X^*) < 0$，即 $f(X) < f(X^*)$，$f(X)$ 是下降的。同理，若 $\frac{\partial f(X^*)}{\partial l} > 0$，说明 $f(X)$是上升的。

二、梯度

定义 1.4 设 $f(X)=f(x_1, x_2, \cdots, x_n)$ 是定义在 R^n 上的可微函数,以 $f(X)$的 n 个偏导数为分量的向量称为 $f(X)$在 X 处的梯度,记为

$$\nabla f(X)=\left[\frac{\partial f(X)}{\partial x_1}, \frac{\partial f(X)}{\partial x_2}, \cdots, \frac{\partial f(X)}{\partial x_n}\right]^T$$

其中,∇为 Hamilton 算子(矢量微分算子)。

显然,梯度也可以称为 $f(X)$关于向量 X 的一阶导数。

例 1.5 函数 $f(X)=c$(常数)的梯度为 $\nabla f(X)=0$,即 $\nabla c=0$。

例 1.6 求函数 $f(X)=a^TX$ 的梯度。

解 设 $a=(a_1, a_2, \cdots, a_n)^T$, $X=(x_1, x_2, \cdots, x_n)^T$,则

$$a^TX=\sum_{i=1}^{n}a_ix_i$$

于是

$$\frac{\partial f(X)}{\partial x_j}=\frac{\partial}{\partial x_j}\left(\sum_{i=1}^{n}a_ix_i\right)=a_j, \ j=1, 2, \cdots, n$$

故 $\nabla f(X)=\nabla(a^TX)=a$。

例 1.7 A 为 n 阶对称矩阵,求函数 $f(X)=X^TAX$ 的梯度。

解 设 $X=(x_1, x_2, \cdots, x_n)^T$,则

$$f(X)=X^TAX=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j$$

因 $a_{ij}=a_{ji}$,于是

$$\frac{\partial f(X)}{\partial x_j}=\frac{\partial}{\partial x_j}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j\right)=\sum_{i=1}^{n}a_{ij}x_i+\sum_{i=1}^{n}a_{ji}x_i=2\sum_{i=1}^{n}a_{ij}x_i,$$
$$j=1, 2, \cdots, n$$

故 $\nabla f(X)=\nabla(X^TAX)=2AX$。

三、梯度与方向导数的关系

定理 1.1 设函数 $f(X)(X\in R^n)$在点 X^* 处可微,e 是方向 l 上的单位向量,则

$$\frac{\partial f(X^*)}{\partial l} = \nabla f\,(X^*)^T e$$

证 由 $f(X^* + te) - f(X) = t\,\nabla f\,(X^*)^T e + O(\|te\|) = t\,\nabla f\,(X^*)^T e + O(t)$，可得

$$\frac{\partial f(X^*)}{\partial l} = \lim_{t\to 0^+}\frac{f(X^* + te) - f(X^*)}{t} = \lim_{t\to 0^+}\frac{t\,\nabla f\,(X^*)^T e + o(t)}{t} = \nabla f\,(X^*)^T e。$$

由于 $\nabla f\,(X^*)^T e = |\nabla f(X^*)|\,|e|\cos\big((\nabla f(X^*), e)\big) = |\nabla f(X^*)|\cos\big((\nabla f(X^*), e)\big)$，可见，方向导数 $\frac{\partial f(X^*)}{\partial l}$ 是梯度$\nabla f(X^*)$在方向 l 上的投影。

方向导数的正负决定了函数值的升降，而它的绝对值大小决定了函数值升降的快慢。绝对值越大，函数值升降的速度就越快。由 $-1 \leqslant \cos\big((\nabla f(X^*), e)\big) \leqslant 1$，可得如下结论：

(1) 方向 l 与梯度的方向一致时，有 $\cos\big((\nabla f(X^*), e)\big) = 1$，$\frac{\partial f(X^*)}{\partial l}$ 有最大值 $|\nabla f(X^*)|$，所以，梯度方向是函数值在这点的最速上升方向；

(2) 方向 l 与梯度的方向正交时，$\cos\big((\nabla f(X^*), e)\big) = 0$，$\frac{\partial f(X^*)}{\partial l} = 0$，所以，函数在与其梯度正交的方向上变化率等于 0；

(3) 方向 l 与梯度的方向成锐角时，$0 < \cos\big((\nabla f(X^*), e)\big) < 1$，$0 < \frac{\partial f(X^*)}{\partial l} < |\nabla f(X^*)|$，所以，函数在与其梯度成锐角的方向上是上升的；

(4) 方向 l 与梯度的方向成钝角时，$-1 < \cos\big((\nabla f(X^*), e)\big) < 0$，$-|\nabla f(X^*)| < \frac{\partial f(X^*)}{\partial l} < 0$，所以，函数在与其梯度成锐角的方向上是下降的；

(5) 方向 l 与梯度的方向相反时，$\cos\big((\nabla f(X^*), e)\big) = -1$，$\frac{\partial f(X^*)}{\partial l}$ 有最小值 $-|\nabla f(X^*)|$，所以，梯度反方向是函数值的最速下降方向。

为尽快得到一个最小化问题的最优解，每一步迭代中总是希望搜索方向 l 尽可能等于或靠近目标函数的负梯度方向，这样才能使函数值下降最快。

四、Taylor 展开

定理 1.2　若 $f(X)$在 X_0 的某邻域内二阶连续可微，则对该邻域内任何点 X，在该邻域中总能找到一点 $Y = X_0 + \theta(X - X_0)$ $(0 < \theta < 1)$，有

$$f(X) = f(X_0) + \nabla f(X_0)^T(X - X_0) + \frac{1}{2}(X - X_0)^T \nabla^2 f(Y)(X - X_0) \tag{1.8}$$

证　设 $P = (X - X_0)$，$\varphi(t) = f(X_0 + tP)$，则

$$\varphi(0) = f(X_0),\ \varphi(1) = f(X_0 + P),\ \varphi'(0) = \nabla f(X_0)^T P$$

按一元函数对 $\varphi(t)$在 $t=0$ 点展开，得

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \frac{1}{2}\varphi''(\theta t)t^2 (0 < \theta < 1)$$

令 $t = 1$，则

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\theta) \tag{1.9}$$

又由于 $\varphi''(\theta) = P^T \nabla^2 f(X_0 + \theta P)P$，将 $\varphi(0)$，$\varphi(1)$，$\varphi'(0)$和 $\varphi''(\theta)$代入(1.9)，得

$$f(X_0 + P) = f(X_0) + \nabla f(X_0)^T P + \frac{1}{2}P^T \nabla^2 f(X_0 + \theta P)P \tag{1.10}$$

将 $X = X_0 + P$，$Y = X_0 + \theta(X - X_0) = X_0 + \theta P$ 代入(1.10)，得

$$f(X) = f(X_0) + \nabla f(X_0)^T(X - X_0) + \frac{1}{2}(X - X_0)^T \nabla^2 f(Y)(X - X_0)$$

式(1.8)还可以写成

$$f(X) = f(X_0) + \nabla f(X_0)^T(X - X_0) + \frac{1}{2}(X - X_0)^T \nabla^2 f(X_0)(X - X_0) + o(\| X - X_0 \|^2)$$

§1-3　Hesse 矩阵与 Jacobi 矩阵

定义 1.5　设函数 $f(X)(X \in R^n)$ 在点 X^* 处二阶连续可微，则称矩阵

$$
\nabla^2 f(X^*) = \begin{bmatrix} \dfrac{\partial^2 f(X^*)}{\partial x_1^2} & \dfrac{\partial^2 f(X^*)}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(X^*)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(X^*)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(X^*)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(X^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial^2 f(X^*)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(X^*)}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(X^*)}{\partial x_n^2} \end{bmatrix}
$$

为函数 $f(X)$在点 X^* 的 Hesse 矩阵,亦记为 $H(X^*)$。

由于 $f(X)$二阶连续可微,故有 $\dfrac{\partial^2 f}{\partial x_i \partial x_j} = \dfrac{\partial^2 f}{\partial x_j \partial x_i}$,于是,$\nabla^2 f(X)$为对称矩阵。

例 1.8 已知 $f(X) = a^T X + b$,求 $f(X)$的梯度和 Hesse 矩阵。

解 因为$\dfrac{\partial f(X)}{\partial x_i} = a_i (i = 1, 2, \cdots, n)$,所以$\nabla f(X) = (a_1, a_2, \cdots, a_n)^T = a$。

由 $f(X)$的梯度进而可知:$\dfrac{\partial^2 f(X)}{\partial x_i \partial x_j} = 0 (i, j = 1, 2, \cdots, n)$,所以$\nabla^2 f(X) = 0$。

例 1.9 已知 $f(X) = \dfrac{1}{2} X^T A X + b^T X + c$,若 A 正定,称 $f(X)$为正定二次函数。求 $f(X)$的梯度和 Hesse 矩阵。

解 设 $A = (a_{ij})_{n \times n}$, $X = (x_1, x_2, \cdots, x_n)^T$, $b = (b_1, b_2, \cdots, b_n)^T$,则

$$
f(X) = \frac{1}{2} X^T A X + b^T X + c = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c
$$

将其对各变量 $x_i (i = 1, 2, \cdots, n)$ 求偏导数,得

$$
\nabla f(X) = \begin{bmatrix} \dfrac{\partial f(X)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(X)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \sum\limits_{j=1}^{n} a_{1j} x_j + b_1 \\ \vdots \\ \sum\limits_{j=1}^{n} a_{nj} x_j + b_n \end{bmatrix} = \begin{bmatrix} \sum\limits_{j=1}^{n} a_{1j} x_j \\ \vdots \\ \sum\limits_{j=1}^{n} a_{nj} x_j \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = AX + b
$$

由上式可知

$$
\frac{\partial f(X)}{\partial x_i} = \sum_{j=1}^{n} a_{ij} x_j + b_i, \ i = 1, 2, \cdots, n
$$