运筹学与优化方法

晁国清
计算机科学与技术学院

课程回顾

- 一般问题的KKT条件
- 举例
  1. Water filling;
  2. Lasso;
  3. SVM

➢给定一个一般约束优化问题：

$$\min_{x} \ f(x)$$
$$s.t.\ \ h_i(x) \le 0, i = 1, \cdots, m$$
$$\ell_j(x) = 0, j = 1, \cdots, r$$

KKT(Karush-Kuhn-Tucker)条件(一阶最优性条件)是：

$0 \in \nabla_x\left(f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x)\right)$ (stationarity)

$u_i \cdot h_i(x) = 0\ for\ all\ i$ (complementary slackness)

$h_i(x) \le 0, \ell_j(x) = 0\ for\ all\ i, j$ (primal feasibility)

$u_i \ge 0\ for\ all\ i$ (dual feasibility)

- 约束形式和拉格朗日形式

➢约束形式：

$$\min_{x} \ f(x)$$
$$s.t.\ \ h(x) \le t \qquad \text{(C)}$$

这里$t \in R$就是一个要调的参数

➢拉格朗日形式：

$$\min_{x} f(x) + \lambda \cdot h(x) \qquad \text{(L)}$$

这里$\lambda \ge 0$是一个可调的参数

课程内容

- 只有等式约束的KT (Kuhn-Tucker)条件
- 只有不等式约束的KT条件
- 举例
- 一般约束的KT条件
- 凸规划的一阶充分条件

等式约束最优化问题的KT条件

考虑等式约束性最优化问题：

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x)=0 \end{array}$$

$f: \mathbb{R}^n \to \mathbb{R}, h: \mathbb{R}^n \to \mathbb{R}^l$均连续可微。可行集：$S = \{\mathrm{x}|\mathrm{h}(\mathrm{x}) = 0\}$

回顾高等数学中所学的条件极值：

问题:在Φ(x,y)=0的条件下， 求z=f(x,y)极值

$$(fh) \quad \begin{cases} \min & f(x,y) \\ \text{s.t.} & \Phi(x,y)=0 \end{cases}$$

引入Lagrange乘子：λ

Lagrange函数 $L(x,y;\lambda)= f(x,y)+ \lambda\, \Phi(x,y)$

等式约束最优化问题的KT条件

若(x^*,y^*)是条件极值，则存在λ^*，使

$$\begin{cases} f_x(x^*,y^*)+\lambda^*\Phi_x(x^*,y^*)=0 \\ f_y(x^*,y^*)+\lambda^*\Phi_y(x^*,y^*)=0 \\ \Phi(x^*,y^*)=0 \end{cases}$$

推广到多个等式约束，可得到对于(fh)的情况：

$$\begin{cases} \min & f(x) \\ \text{s.t.} & hj(x)=0 \quad j=1,2,\ldots,\ell \end{cases}$$

若x*是(fh)的l.opt. ，则存在$v^* \in R^l$使

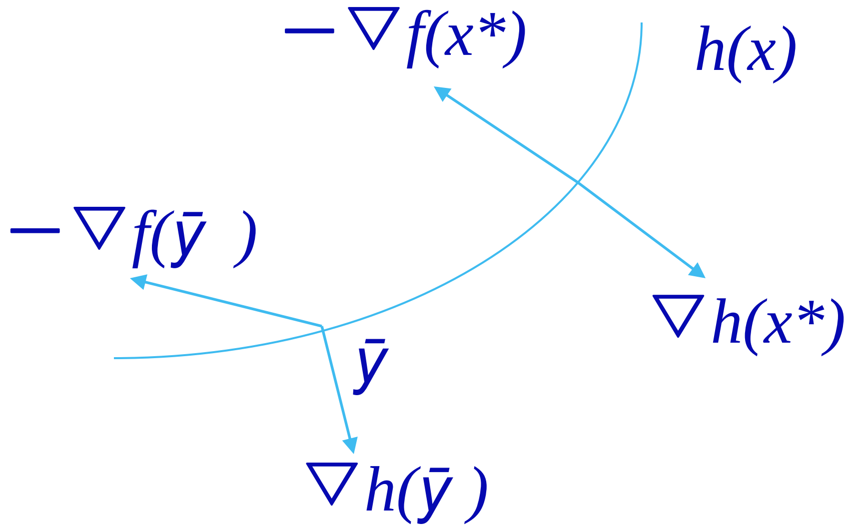
$$\nabla f(x^*)+\sum_{j=1}^{\ell} v_j^* \nabla h_j(x^*)=0$$

矩阵形式：

$$\nabla f(x^*)+\frac{\partial h(x^*)}{\partial x}v^*=0$$

几何意义

几何意义是明显的：考虑一个约束的情况：



这里 x^* —l.opt. $\nabla f(x^*)$与 $\nabla h(x^*)$ 共线，而$\bar{y}$非l.opt. $\nabla f(\bar{y})$与$\nabla h(\bar{y})$不共线。

最优性条件即

$$\nabla f(x^*) = -\sum_{j=1}^{h} v_j^* \nabla h_j(x^*)$$

不等式约束最优化问题的KT条件

考虑问题

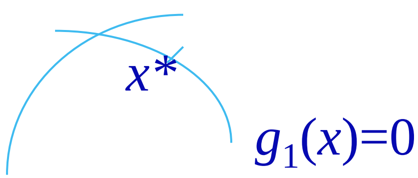
$$(fg)\quad \begin{cases} \min\ f(x) \\ \text{s.t. } g_i(x) \le 0 \quad i=1,2,\ldots,m \end{cases}$$

设$x^* \in S=\{x|g_i(x) \le 0 \quad i=1,2,\ldots,m\}$

令 $I=\{i|\ g_i(x^*) =0 \quad i=1,2,\ldots,m\}$

称I为 x^*点处的起作用集（紧约束集）。

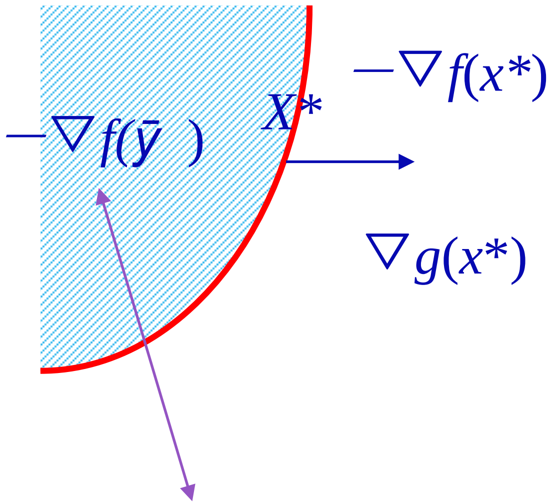
如果x^*是$l.opt.$，对每一个约束函数来说，只有当它是起作用约束时，才产生影响，如：



$g_1(x^*)=0$, g_1为起作用约束或紧约束

几何意义

特别地，　有如下特征：如图(看书上)



在x^*：　$\nabla f(x^*)+u^*\nabla g(x^*)=0$，$u^*>0$

要使函数值下降，必须使$g(x)$值变大，则在$\bar{y}$点使$f(x)$下降的方向（$-\nabla f(\bar{y})$方向）指向约束集合内部，因此$\bar{y}$不是l.opt.

KT条件的几何意义：目标函数的负梯度$-f(x^*)$可表示为紧约束函数梯度的非负组合

不等式约束最优化问题的KT条件

定理（最优性必要条件）:（KT条件）

问题(fg), 设$S=\{x|g_i(x)\le 0\}$, $x^* \in S$, I为x^*点处的起作用,

设f, $g_i(x)$,$i \in I$在x^*点可微，$g_i(x), i \notin I$ 在x^*点连续。

向量组$\{\nabla g_i(x^*), i \in I\}$线性无关。

如果x^*—l.opt. 那么， $u_i^* \ge 0, i \in I$使

$$\nabla f(x^*) + \sum_{i\in I} u_i^* \nabla g_i(x^*) = 0$$

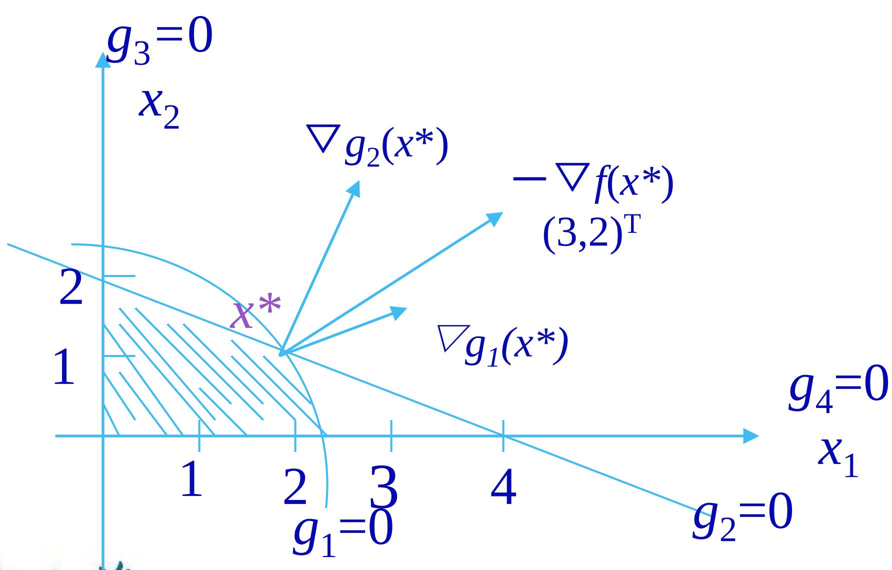
如果在$x^*, g_i(x)$可微，$\forall i$, 那么

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} u_i^* \nabla g_i(x^*) = 0 \\ u_i^* \ge 0, i = 1,2,\cdots,m \\ {u_i^*}^T g_i(x^*) = 0, i = 1,2,\cdots,m \text{ (互补松弛条件)} \end{cases}$$

满足K-T条件的点x^*称K-T点

举例

$$
\begin{cases}
\min f(x_1,x_2) = (x_1-3)^2 + (x_2-2)^2 \\
s.t.\, g_1(x_1,x_2) = x_1^2 + x_2^2 - 5 \le 0 \\
g_2(x_1,x_2) = x_1 + 2x_2 - 4 \le 0 \\
g_3(x_1,x_2) = -x_1 \le 0 \\
g_4(x_1,x_2) = -x_2 \le 0
\end{cases}
$$



举例

在x^*点$\begin{cases} g_1(x_1,x_2)=0 \\ g_2(x_1,x_2)=0 \end{cases}$，交点$(2,1)^T$，起作用集$I=\{1,2\}$

$$\nabla g_1(x^*)=(2x_1^*,2x_2^*)^T=(4,2)^T$$
$$\nabla g_2(x^*)=(1,2)^T$$
$$\nabla f(x^*)=\left(2(x_1^*-3),2(x_2^*-2)\right)^T=(-2,-2)^T$$

计算可得 $u_1^*=\frac{1}{3}\quad u_2^*=\frac{2}{3}$ 使

$$\nabla f(x^*)+\frac{1}{3}\nabla g_1(x^*)+\frac{2}{3}\nabla g_2(x^*)=0$$

用KKT条件求解

$$\nabla f(x)=\begin{pmatrix}2(x_1-3)\\2(x_2-2)\end{pmatrix},\nabla g_1(x)=\begin{pmatrix}2x_1\\2x_2\end{pmatrix},\nabla g_2(x)=\begin{pmatrix}1\\2\end{pmatrix}$$
$$\nabla g_3(x)=\begin{pmatrix}-1\\0\end{pmatrix},\nabla g_4=\begin{pmatrix}0\\-1\end{pmatrix}$$

举例

$$\begin{cases} \nabla f(x) + \sum_{i}^{m} u_i \nabla g_i(x) \\ u_i \geq 0, i = 1,2,\cdots,m \\ u_i g_i(x) = 0 \end{cases}$$

$$\begin{cases} 2(x_1 - 3) + u_1 2x_1 + u_2 - u_3 = 0 & (1) \\ 2(x_2 - 2) + u_1 2x_2 + 2u_2 - u_4 = 0 & (2) \\ u_1, u_2, u_3, u_4 \geq 0 & \\ u_1(x_1^2 + x_2^2 - 5) = 0 & (3) \\ u_2(x_1 + 2x_2 - 4) = 0 & (4) \\ u_3 x_1 = 0 & (5) \\ u_4 x_2 = 0 & (6) \end{cases}$$

举例

可能的K-T点出现在下列情况：

①两约束曲线的交点：g_1与g_2， g_1与g_3， g_1与g_4， g_2与g_3， g_2与g_4， g_3与g_4。

②目标函数与一条曲线相交的情况： g_1， g_2， g_3， g_4

对上述每一个情况容易求得满足(1)~(6)的点$(x_1, x_2)^T$及乘子u_1, u_2, u_3, u_4， 当满足$u_i \geq 0$时，即为KT点。

下面举几个情况：

- g_1与g_2交点：$x=(2,1)^T \in S$，$I=\{1,2\}$ 则$u_3=u_4=0$ 解

$$\begin{cases} 2(x_1-3)+2u_1x_1+u_2=0 \\ 2(x_2-2)+2u_1x_2+2u_2=0 \end{cases}$$

得到$u_1=\frac{1}{3}, u_2=\frac{2}{3}>0$

故$x=(2,1)^T$是K-T点

举例

●g_1与g_3交点：$\begin{cases} x_1^2+x_2^2-5=0 \\ x_1=0 \end{cases}$

得到$x=\left(0,\pm\sqrt{5}\right)^T$

因为$\left(0,\pm\sqrt{5}\right)^T\notin S$, 不满足$g_2\leq 0$, 所以不是K−T点；

●g_3,g_4交点：$x=(0,0)^T\in S$, $I=\{3,4\}$

故$u_1=u_2=0$

解$\begin{cases} 2(0-3)-u_3=0 \\ 2(0-2)-u_4=0 \end{cases}$ 得到$u_3=-6<0, u_4=-4<0$

所以不是K−T点

举例

目标函数$f(x)$与$g_1(x)$相切的情况：$I=\{1\}$, 则$u_2=u_3=u_4=0$

解

$$\begin{cases} 2(x_1-3)+2x_1u_1=0 \\ 2(x_2-2)+2x_2u_1=0 \\ x_1^2+x_2^2-5=0 \end{cases}$$

得到$\left(\pm\sqrt{\frac{45}{13}},\pm\sqrt{\frac{20}{13}}\right)\notin S$

因为$g_2(x_1,x_2)=\sqrt{\frac{45}{13}}+2\sqrt{\frac{20}{13}}-4=7\sqrt{\frac{5}{13}}-4=0.34>0,$

所以该点不是K-T点

注意：约束最优化问题中的K-T条件低位相当于无约束最优化问题中的逐点（梯度等于0的点）条件。

一般约束最优化问题的KT条件

问题：

$$(\mathrm{fgh})\quad \begin{cases} \min f(x) \\ s.t.\ g_i(x) \le 0, i = 1,2,\cdots,m \\ h_j(x) = 0, j = 1,2,\cdots,\ell \end{cases}$$

$f, g_i, h_j: \mathbb{R}^n \to \mathbb{R}, i = 1,2,\cdots,m; j = 1,2,\cdots,\ell$, 可行集为$S = \{x | g_i(x) \le 0, h_j(x) = 0, i = 1,2,\cdots,m; j = 1,2,\cdots,\ell\}$

也可写成矩阵形式：

$$(\mathrm{fgh})\quad \begin{cases} \min & f(\boldsymbol{x}) \\ s.t. & \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0} \\ & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \end{cases}$$

$f: \mathbb{R}^n \to \mathbb{R}, \boldsymbol{g}: \mathbb{R}^n \to \mathbb{R}^m, \boldsymbol{h}: \mathbb{R}^n \to \mathbb{R}^\ell$, 可行集$S = \{\boldsymbol{x} | \boldsymbol{g}(\boldsymbol{x}) \le 0, \boldsymbol{h}(\boldsymbol{x}) = 0\}$

一般约束最优化问题的KT条件

设$x^* \in S = \{x|g(x) \le 0, h(x) = 0\}$, I为x^*点的紧约束集；
$f, g_i, i \in I, j = 1,2,\cdots,\ell$ 在 x^* 点可微； $g_i, i \notin I$, 在 x^* 点连续。
再设 $\{\nabla g_i(x^*), i \in I, \nabla h_j(x^*), j = 1,2,\cdots,\ell\}$ 线性无关，那么，
存在$u_i \ge 0, i \in I, v_j, j = 1,2,\cdots,\ell$, 使

$$\nabla f(x^*) + \sum_{i\in I} u_i \nabla g_i(x^*) + \sum_{j=1}^{\ell} v_j \nabla h_j(x^*) = 0$$

如进一步假设$g_i(x), i \notin I$, 在x^*点可微, 则

$\nabla f(x^*) + \frac{\partial g(x^*)}{\partial x} u + \frac{\partial h(x^*)}{\partial x} v = 0$ （稳定性条件）

$u \ge 0, u \in \mathbb{R}^m, v \in \mathbb{R}^\ell$ （对偶可行性条件）

$u^T g(x^*) = 0$ （互补松弛条件）

关于凸规划的一阶充分条件

当一般约束最优化问题是凸规划时，KT条件称为充要条件。

充分性：考虑问题(fgh)，设$f:\mathbb{R}^n \to \mathbb{R}$是可微凸函数，$g_i:\mathbb{R}^n \to \mathbb{R}(i=1,2,\cdots,m)$均为可微凸函数，$h(x)=Ax-b, A$为$\ell \times n$矩阵，$b \in \mathbb{R}$^$\ell$。再设$x^* \in S=\{s|g(x)\le 0, h(x)=0\}$, 并且满足
$\nabla f(x^*)+\sum_{i\in I} u_i \nabla g_i(x^*)+\sum_{j=1}^{\ell} v_j \nabla h_j(x^*)=0$
，则x^*为问题(fgh)的g. opt

HARBIN INSTITUTE OF TECHNOLOGY

参考文献

运筹学与最优化方法　　　第6章　　P133-137