二阶优化条件

在LICQ（或其他适当的约束规范）前提下，我们可以通过局部最优点x^*处的$\mathcal{F}(x^*)$来分析全部一阶局部信息应该具有的性质，也就是KKT条件。本节我们继续讨论二阶局部信息应该具有的性质，并最终给出二阶必要和充分条件。

已知x^*是局部最优点，由一阶条件，存在λ^*使(x^*, λ^*)满足KKT条件。首先，我们将焦点放在一阶条件中无法判定升降的从x^*出发的可行方向w，也即w满足

$$w^T \nabla f(x^*) = 0.$$

（这里如果大于零，则一阶信息确保了上升，如果小于零，则下降。而严格等于零，表明这是一个待定方向。若x^*是严格局部最优点，我们知道这个方向必须上升，但从一阶条件无法得到这个结论。）

我们在KKT条件的框架下，将其定义为一个锥的形式，称为关键锥(critical cone)：

$$\mathcal{C}(x^*, \lambda^*) = \left\{ w \in \mathcal{F}(x^*) \left| \nabla c_i(x^*)^T w = 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}, \lambda_i > 0 \right. \right\}.$$

（我们仍然保持LICQ，因此$T_\Omega(x^*)$和$\mathcal{F}(x^*)$是一致的。这里$w \in \mathcal{F}(x^*)$首先表明从一阶信息看，w是在x^*邻域内的Ω部分的，也即$\forall i \in \mathcal{E}$，有$\nabla c_i(x^*) = 0$。而对$\forall i \in \mathcal{A}(x^*) \cap \mathcal{I}$，有$\nabla c_i(x^*)^T w \geq 0$，这里注意

$$w^T \nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} w^T \lambda_i \nabla c_i(x^*),$$

对$i \in \mathcal{E}$项，右端部分已经为零。对于$i \in \mathcal{A}(x^*) \cap \mathcal{I}$且$\lambda_i = 0$项，右端部分也为零。所以只要提取$\lambda_i > 0$，且$\nabla c_i(x^*)^T w = 0$的方向，就能确保把全部不确定，也就是$w^T \nabla f(x^*) = 0$的方向都提取出来。）

或等价地，

$$w \in \mathcal{C}(x^*, \lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{E}, \\ \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}, \lambda_i > 0 \\ \nabla c_i(x^*)^T w \geq 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}, \lambda_i = 0. \end{cases} \tag{12.53}$$

由(12.53)，我们马上有结论

$$w \in \mathcal{C}(x^*, \lambda^*) \Rightarrow \lambda_i^* \nabla c_i(x^*)^T w = 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}. \tag{12.54}$$

于是结合(12.34a)和(12.33)，得

$$w \in \mathcal{C}(x^*, \lambda^*) \Rightarrow w^T \nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* w^T \nabla c_i(x^*) = 0. \tag{12.55}$$

也即沿关键锥中的方向，一阶信息已经退化了，和之前无约束优化问题一样，如果要知道$\mathcal{C}$中的方向是上升还是下降，我们必须进一步考虑其二阶信息。

例 12.7

考虑问题

$$\min x_1, \quad \mathrm{s.t.} x_2 \geq 0, 1 - (x_1 - 1)^2 - x_2^2 \geq 0, \tag{12.56}$$

（草图，手算验证一下结果）显然全局最优解是$x^* = (0, 0)^T$，而$\mathcal{A}(x^*) = \{1, 2\}$，对应有唯一的Lagrange乘子$\lambda^* = (0, 0.5)^T$。同时，注意到在$x^*$点，活跃约束的梯度分别为

$$\nabla c_1(x^*) = (0, 1)^T, \quad \nabla c_2(x^*) = (2, 0)^T,$$

满足LICQ条件。该问题的线性化可行方向为

$$\mathcal{F}(x^*) = \{d|d \geq 0\},$$

以及关键锥为

$$\mathcal{C}(x^*, \lambda^*) = \{(0, w_2)^T | w_2 \geq 0\}.$$

我们注意到f在$\mathcal{F}(x^*)$方向内，都是非减的，因此和x^*是最优解不矛盾。但要判定x^*就是最优解，我们缺少了$\mathcal{C}(x^*, \lambda^*)$方向上的信息，因为在这个方向上，$\nabla f(x^*) = 0$，$f$究竟是上升还是下降，由Taylor展开的下一项（二阶项）决定。至此，二阶必要条件和充分条件都已经很清晰了。

定理 12.5（二阶必要条件） 假设x^*是问题(12.1)的局部最优解，且有LICQ成立。令λ^*是对应的Lagrange乘子，则

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad \forall w \in \mathcal{C}(x^*, \lambda^*). \tag{12.57}$$

证明 由x^*是(12.1)的局部最优解，全部趋于x^*的可行序列$\{z_k\}$对充分大的k均有

$$f(z_k) \geq f(x^*).$$

现在我们构建可行序列，使得其极限方向就是w。首先由于$w \in \mathcal{C}(x^*, \lambda^*) \subset \mathcal{F}(x^*)$，因此我们可以用证明引理12.2时的技巧（隐函数定理），构建可行序列和对应正序列，使得

$$\lim_{k \to \infty} \frac{z_k - x^*}{t_k} = w, \tag{12.58}$$

也即

$$z_k - x^* = t_k w + o(t_k). \tag{12.59}$$

而在构建$\{z_k\}$的时候，我们由(12.42)，已知

$$c_i(z_k) = t_k \nabla c_i(x^*)^T w, \forall i \in \mathcal{A}(x^*), \tag{12.60}$$

（做Taylor展开，此时$c_i(x^*) = 0$）结合(12.33)，(12.60)和(12.54)，我们有

$$\begin{aligned}
\mathcal{L}(z_k, \lambda^*) &= f(z_k) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* c_i(z_k) \\
&= f(z_k) - t_k \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*)^T w \\
&= f(z_k),
\end{aligned} \tag{12.61}$$

（然而一阶项实际是零，消失了。）我们再一次Taylor展开，

$$\mathcal{L}(z_k, \lambda^*) = \mathcal{L}(x^*, \lambda^*) + (z_k - x^*)^T \nabla_x \mathcal{L}(x^*, \lambda^*) + \frac{1}{2}(z_k - x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*)(z_k - x^*) + o(\|z_k - x^*\|^2). \tag{12.62}$$

在我们的例子中，考虑到互补性条件(12.34e)，已有$\mathcal{L}(x^*, \lambda^*) = f(x^*)$。而在(12.34a)中（一阶必要条件），全部一阶导数项也全部为零（关键锥），再由(12.59)，我们可以将(12.62)改写为

$$\mathcal{L}(z_k, \lambda^*) = f(x^*) + \frac{1}{2} t_k^2 w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w + o(t_k^2). \tag{12.63}$$

结合(12.61)，有

$$f(z_k) = f(x^*) + \frac{1}{2} t_k^2 w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w + o(t_k^2). \tag{12.64}$$

若$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w < 0$，由(12.64)，存在$f(z) < f(x^*)$对$k$充分大成立，和$x^*$是局部最优解矛盾。因此(12.57)必然成立。□

定理 12.6（二阶充分条件） 假设对可行点$x^* \in \mathbb{R}^n$存在Lagrange乘子λ^*使得KKT条件成立，并且同时有

$$w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w > 0, \quad \forall w \in \mathcal{C}(x^*, \lambda^*), w \neq 0. \tag{12.65}$$

则x^*是问题(12.1)的严格局部最优解。

证明自己看书。实际上书上的证明也不完整。但这个定理的结论还是很显然的。

例12.8 我们检查例12.2中出现的问题(12.8)的二阶局部信息。注意到

$$f(x) = x_1 + x_2, \quad c_1(x) = 2 - x_1^2 - x_2^2 \geq 0,$$

因此

$$\mathcal{L}(x, \lambda) = (x_1 + x_2) - \lambda_1 (2 - x_1^2 - x_2^2),$$

而$x^* = (-1, -1)^T$，$\lambda^* = \frac{1}{2}$。于是Lagrange函数的Hessian阵为：

$$\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 2\lambda_1^* & 0 \\ 0 & 2\lambda_1^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

这是一个正定矩阵，所以定理12.6的条件必然满足。于是$x^* = (-1, -1)^T$必然是严格局部最优点。（这个例子太绝对了一些，任何可行方向，不管是不是关键锥，都是上升的。）

例 12.9 更复杂一些的例子：

$$\min -0.1(x_1 - 4)^2 + x_2^2, \quad \text{s.t.} x_1^2 + x_2^2 - 1 \geq 0, \tag{12.72}$$

（草图）我们的可行域是单位圆和以外部分。现在这是一个无界问题，而且目标函数可以取到$-\infty$，因此也无解。但它在边界上是否存在严格局部最优解呢？我们应该无法从草图上清晰地了解这一点，但仍然可以通过KKT条件和二阶条件来判定：

$$\nabla_x \mathcal{L}(x, \lambda) = \begin{bmatrix} -0.2(x_1 - 4) - 2\lambda_1 x_1 \\ 2x_2 - 2\lambda_1 x_2 \end{bmatrix}, \tag{12.73a}$$

$$\nabla_{xx}^2 \mathcal{L}(x, \lambda) = \begin{bmatrix} -0.2 - 2\lambda_1 & 0 \\ 0 & 2 - 2\lambda_2 \end{bmatrix}, \tag{12.73b}$$

这里$x^* = (1, 0)^T$，$\lambda^* = 0.3$满足KKT条件，且对应$\mathcal{A}(x^*) = \{1\}$。而

$$\nabla c_1(x^*) = \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$

所以关键锥为

$$\mathcal{C}(x^*, \lambda^*) = \left\{ (0, w_2)^T \,|w_2 \in \mathbb{R} \right\}.$$

确实是可行域的边缘切线，我们现在要判定沿这条边缘切线，目标函数是上升还是下降，因为一阶信息在这里退化了，所以要进一步考虑二阶信息。

$$\forall w \neq 0, \quad w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w = [0 \quad w_2] \begin{bmatrix} -0.8 & 0 \\ 0 & 1.4 \end{bmatrix} \begin{bmatrix} 0 \\ w_2 \end{bmatrix} = 1.4 w_2^2 > 0.$$

（上式书中计算有误，但不影响结论）因此由定理12.6，$x^* = (1, 0)^T$是严格局部最优解。

二阶条件的验证比较复杂，为此人们提出了一种稍微弱化但更加方便的形式。当满足KKT条件的λ^*是唯一（比如LICQ），并且严格互补性条件条件成立时，关键锥的定义(12.53)简化为

$$\mathcal{C}(x^*, \lambda^*) = \text{Null}\left[\nabla c_i(x^*)^T\right]_{i \in \mathcal{A}(x^*)} = \text{Null} A(x^*),$$

这里$A(x^*)$和(12.37)一致。换言之，$\mathcal{C}(x^*,\lambda^*)$此时就是$x^*$点全部活跃约束梯度构成的矩阵的零空间。类似(12.39)，我们可以再次定义矩阵Z，它的列向量由$A(x^*)$的零空间的基组成，从而也能张成$\mathcal{C}(x^*,\lambda^*)$，即

$$\mathcal{C}(x^*,\lambda^*)=\left\{Zu\middle|u\in\mathbb{R}^{|\mathcal{A}(x^*)|}\right\},$$

这里$|\mathcal{A}(x^*)|$表示计算其元素个数。因此，定理12.5中的条件(12.57)可以改为

$$u^TZ^T\nabla^2_{xx}\mathcal{L}(x^*,\lambda^*)Zu\geq 0,\forall u\in\mathbb{R}^n,$$

或者更简洁地，

$$Z^T\nabla^2_{xx}\mathcal{L}(x^*,\lambda^*)Z$$

是半正定的，相应地，定理12.6的条件(12.65)可以改为

$$Z^T\nabla^2_{xx}\mathcal{L}(x^*,\lambda^*)Z$$

是正定的。

而Z可以用QR分解计算得到：

$$A(x^*)^T=Q\begin{bmatrix}R\\0\end{bmatrix}=[Q_1\quad Q_2]\begin{bmatrix}R\\0\end{bmatrix}=Q_1R, \tag{12.74}$$

这里R是上三角矩阵，Q是$n\times n$正交阵。若R非奇异，则$Z=Q_2$。当LICQ不成立时，R是奇异的，此时可以对QR分解做适当的列交换来确定Z。

其它形式的约束规范

现在我们比较清楚，所谓约束规范，本质上指的是对Ω的线性化代数表示，是否能够正确地抓住x^*邻域内Ω的几何外形。所以对于退化的情况，比如全部的活跃约束都是线性函数，也即

$$c_i(x)=a_i^Tx+b_i, \tag{12.75}$$

这里$a_i\in\mathbb{R}^n$，以及$b_i\in\mathbb{R}$，那么显然$\mathcal{F}(x^*)$和Ω对活跃约束的表现而言就是一致的。

引理 12.7 假设$x^*\in\Omega$的全部活跃约束$c_i(x^*)$，$i\in\mathcal{A}(x^*)$，都是线性函数，则$\mathcal{F}(x^*)=T_\Omega(x^*)$。

证明 首先由引理12.2，$T_\Omega(x^*)\subset\mathcal{F}(x^*)$。因此只需证$\mathcal{F}(x^*)\subset T_\Omega(x^*)$，也即$\forall w\in\mathcal{F}(x^*)$，有$w\in T_\Omega(x^*)$。由定义及条件(12.75)，

$$\mathcal{F}(x^*)=\left\{d\middle|\begin{array}{ll}a_i^Td=0, & \forall i\in\mathcal{E},\\ a_i^Td\geq 0, & \forall i\in\mathcal{A}(x^*)\cap\mathcal{I}\end{array}\right\}.$$

首先，对于在x^*不活跃的约束，存在常数$\bar{t}>0$使得其在x^*+tw仍然不活跃，$\forall t\in[0,\bar{t}]$，也即

$$c_i(x^*+tw)>0,\quad\forall i\in\mathcal{I}\backslash\mathcal{A}(x^*),t\in[0,\bar{t}].$$

（可以认为是连续性，还没走到对面。）

现定义序列z_k：

$$z_k=x^*+(\bar{t}/k)w,\quad k=1,2,\cdots$$

因为$a_i^Tw\geq 0$，$\forall i\in\mathcal{I}\cap\mathcal{A}(x^*)$，有

$$c_i(z_k)=c_i(z_k)-c_i(x^*)=a_i^T(z_k-x^*)=\frac{\bar{t}}{k}a_i^Tw\geq 0,\quad\forall i\in\mathcal{I}\cap\mathcal{A}(x^*),$$

于是z_k对于活跃的不等值约束c_i可行，$i \in \mathcal{I} \cap \mathcal{A}(x^*)$。由于$\bar{t}$的设置，我们知道$z_k$对不活跃约束$c_i$也是可行的，$i \in \mathcal{I} \backslash \mathcal{A}(x^*)$。而对等值约束，

$$c_i(z_k) = c_i(z_k) - c_i(x^*) = a_i^T(z_k - x^*) = \frac{\bar{t}}{k} a_i^T w = 0, \quad \forall i \in \mathcal{E},$$

是显然的。所以，z_k对全体$k = 1, 2, \cdots$均可行，进而有

$$\frac{z_k - x^*}{\bar{t}/k} = \frac{(\bar{t}/k)w}{\bar{t}/k} = w,$$

也即w就是$\{z_k\}$的极限方向（切线），因此$w \in T_\Omega(x^*)$，证毕。□

所以我们可以提出一个新的约束规范：全部活跃约束都是线性函数。注意这个约束规范和LICQ互相不覆盖。

定义 12.6 (Mangasarian-Fromovitz constraint qualification, MFCQ) 称x^*点MFCQ成立，若存在$w \in \mathbb{R}^n$使得

$$\begin{aligned} \nabla c_i(x^*)^T w > 0, &\quad \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}, \\ \nabla c_i(x^*)^T w = 0, &\quad \forall i \in \mathcal{E}, \end{aligned}$$

且等值约束的梯度$\{\nabla c_i(x^*), i \in \mathcal{E}\}$都是线性无关的。

注意这里要求不等值约束对应的不等式都是严格成立的。但是MFCQ比LICQ要更弱一点。若x^*满足LICQ，则由$A(x^*)$行满秩，方程组

$$\begin{aligned} \nabla c_i(x^*)^T w = 1, &\quad \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}, \\ \nabla c_i(x^*)^T w = 0, &\quad \forall i \in \mathcal{E}, \end{aligned}$$

必有解w，也即必满足定义12.6。而反面的例子是容易构建的。参见练习12.13。

我们可以在定理12.1（一阶必要条件）中将LICQ替换成MFCQ，如此KKT条件中的Lagrange乘子λ^*将不唯一，但其数量是有限的。此外，约束规范是线性逼近几何区域能够被接受的充分不必要条件（也即哪怕没有约束规范，线性逼近也未必错）。例如对可行域由下列约束构成：

$$x_2 \geq -x_1^2, \quad x_2 \leq x_1^2,$$

在$x^* = (0,0)^T$点，没有任何约束规范成立。但是其线性化方向集

$$\mathcal{F}(x^*) = \left\{(w_1, 0)^T | w_1 \in \mathbb{R}\right\},$$

实际上正确反映了x^*附近可行域的几何特性。

一个几何观点

我们接下来从纯几何的角度观察一阶最优条件，不依赖任何代数形式的描述。我们的原始问题的几何描述是

$$\min f(x), \quad \text{s.t.} x \in \Omega, \tag{12.76}$$

其中Ω是可行域。

首先我们需要定义Ω在x点的法向锥。

定义 12.7 可行域Ω中一点x处的法向锥定义为

$$N_\Omega(x) = \{v \,|\, v^T w \leq 0, \forall w \in T_\Omega(x)\}, \tag{12.77}$$

其中，$T_\Omega(x)$是x处的切锥，如定义12.2。任取$v \in N_\Omega(x)$，我们称v是一个法向量（任取$w \in T_\Omega(x)$，w是一个切向量）。

从几何上看，任何一个法向量v和任何一个切向量w的夹角都至少有$\pi/2$。于是对(12.76)的一阶必要条件为

定理 12.8 假设x^*是f在Ω内的一个局部极小值，则

$$-\nabla f(x^*) \in N_\Omega(x^*). \tag{12.78}$$

（这是最速下降方向。这个定理提供了一个非常几何直观的条件，如果x^*是局部最优点，则最速下降方向一定在法向锥中，反之，我们就能在切锥中找到一个切向是下降的。）

证明 对任意的$d \in T_\Omega(x^*)$，我们有满足定义12.2的$\{t_k\}$和$\{z_k\}$使得

$$z_k \in \Omega, z_k = x^* + t_k d + o(t_k), \forall k. \tag{12.79}$$

因为x^*是局部最优解，故对充分大的k，有

$$f(z_k) \geq f(x^*).$$

由f充分光滑以及Taylor定理，

$$f(z_k) - f(x^*) = t_k \nabla f(x^*)^T d + o(t_k) \geq 0.$$

两边同除以t_k，并取极限$k \to \infty$，有

$$\nabla f(x^*)^T d \geq 0.$$

由于d是$T_\Omega(x^*)$内任何向量，故事实上对$\forall d \in T_\Omega(x^*)$，有

$$-\nabla f(x^*)^T d \leq 0,$$

由定义12.7，$-\nabla f(x^*) \in N_\Omega(x^*)$。□