Lagrange Multipliers & KKT 几何关联推导

目的

我们想要找到一个满足一些约束的函数的最大值或最小值

公式描述

给定一个函数 f，不等式约束 $g_1,...,g_m$ 和等式约束 $h_1,...,h_l$ 都在在定义域 $\Omega \subset R^n$ 上的优化问题：

$$\min_{x\in\Omega} \ f(x)$$

$$s.t.\begin{cases}\forall i & g_i(x) \le 0 \\ \forall j & h_j(x) = 0\end{cases}$$

No constraints

Assume: Let $f:\Omega \to \mathbb{R}$ be a continuously differentiable function. 在定义域上连续可微

【局部最小值】的【充要条件】(Necessary and sufficient conditions):

x^* is a local minimum of $f(x)$ 当且仅当

1. f 在 x^* 处是零梯度 (zero gradient):

$$\nabla_x f(x^*) = 0$$

2. f 的 Hessian 矩阵在 x^* 处是半正定的 (positive semi-definite)：保证f 在 x^* 处是"波谷"

$$v^t(\nabla^2 f(x^*))v \ge 0, \forall v \in \mathbb{R}^n$$

$$\nabla^2_{xx} f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \leftarrow Hessian\ Matrix$$

【局部最大值】的【充要条件】(Necessary and sufficient conditions): x^* is a local minimum of $f(x)$ 当且仅当

1. f 在 x^* 处是零梯度 (zero gradient):

$$\nabla_x f(x^*) = 0$$

2. f 的 Hessian 矩阵在 x^* 处是半正定的 (positive semi-definite)：保证f 在 x^* 处是"波谷"

$$v^t(\nabla^2 f(x^*))v \le 0, \forall v \in \mathbb{R}^n$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \leftarrow Hessian\ Matrix$$

Equality Constrainsts

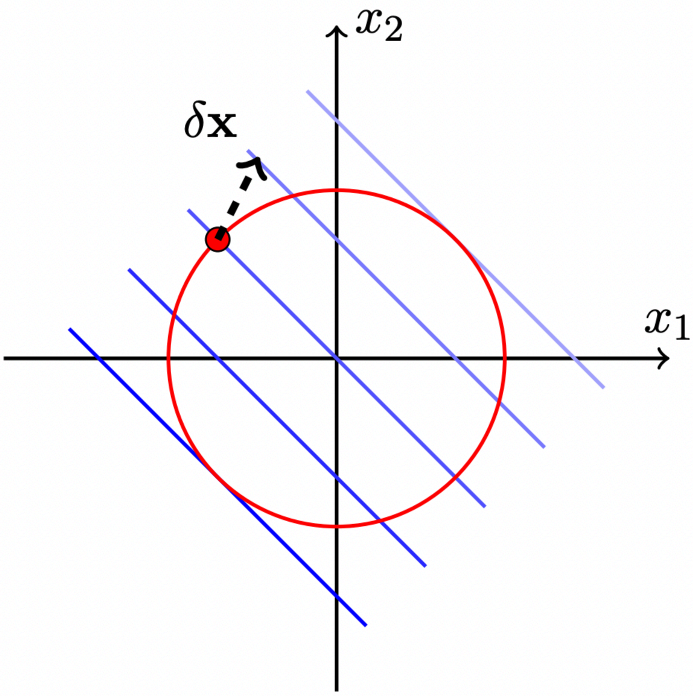
问题提出：

$$\min_{x \in \mathbb{R}^2} \ f(x)$$
$$s.t. \quad h_i(x) = 0, \ \forall i \in \{1, 2, \ldots, l\}$$

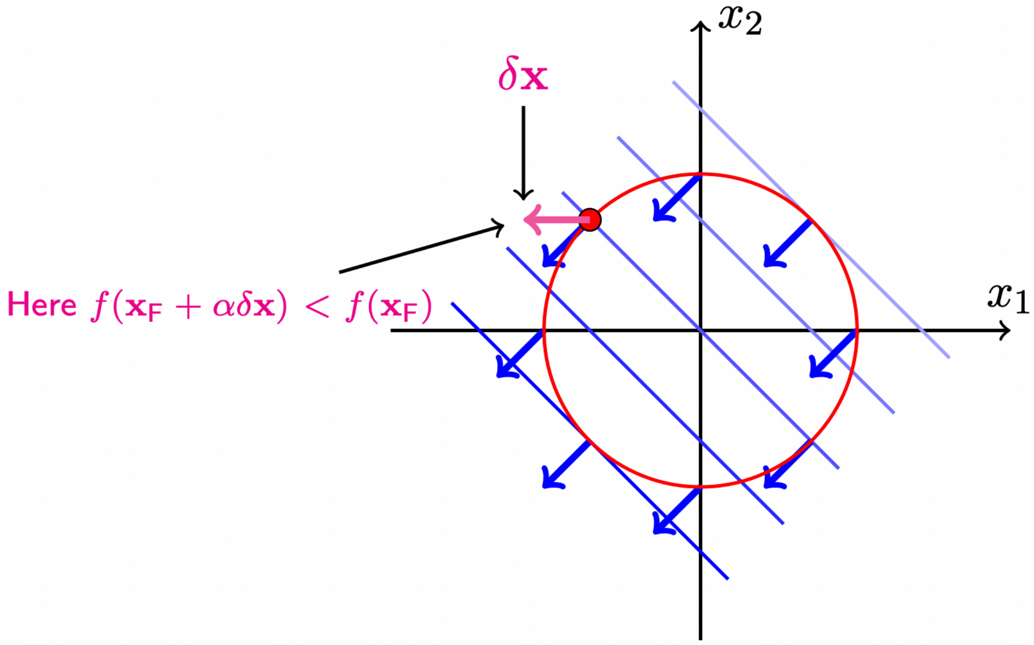
举例：

$$f(x) = x_1 + x_2$$
$$s.t. \ h(x) = x_1^2 + x_2^2 - 2$$

1. 可行点 (feasible point) $x_F \in$ 可行域 (feasible region) 满足约束 $h(x) = x_1^2 + x_2^2 - 2 = 0$，图形上的表示就是在 $x_1^2 + x_2^2 - 2$ 的圆上
2. 目标函数 $f(x) = x_1 + x_2$ 的梯度方向 $\nabla_x f(x) = [1, 1]^T$，所以它的负梯度方向为 $-\nabla_x f(x) = [-1, -1]^T$
3. 我们找到一个点 x_i 满足约束：其中 α 为步长，δx_i 为 x_i 点的运动方向
   - $h(x_F + \alpha \delta x_i) = 0$（确保在圆上）
   - $f(x_F + \alpha \delta x) < f(x_F)$（确保移动后的函数值 < 移动前）



4. 一个点 x_i 要沿着 $f(x)$ 的最速下降 (the steepest descent) 方向，即负梯度方向为 $-\nabla_x f(x) = [-1,-1]^T$；但是 x_i 还要满足等式约束，所以我们要确保 δx_i 与 负梯度方向 $-\nabla_x f(x)$ 的夹角为锐角，即内积 $\delta x_i \cdot (-\nabla_x f(x_F)) > 0$

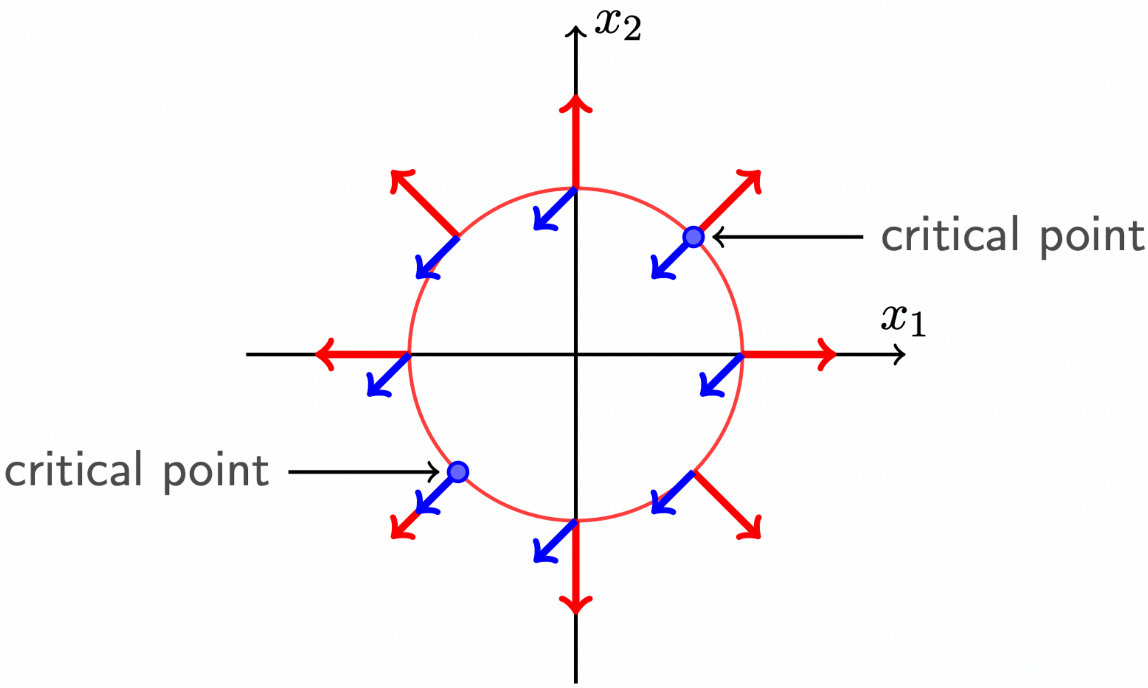


- 至此，我们就找到了满足约的点 x_i 移动的方向，即与目标函数 $f(x)$ 负梯度方向的夹角为锐角的约束函数 $h(x)$ 的切线方向。那么，什么时候停止移动呢？

5. 从图像中我们可以看到，当目标函数 $f(x)$ 与约束函数 $h(x)$ 相切的时候，我们可以取到局部极值点 (临界点 critical point)，即 目标函数 $f(x)$ 的梯度方向与约束函数 $h(x)$ 的梯度方向共线：

$$\nabla_x f(x_F) = \mu \nabla_x h(x_F) \quad \Rightarrow \quad \nabla_x f(x_F) + \mu \nabla_x h(x_F) = 0 \quad ①$$

这个条件【确保局部极值】



而此时，x_i 移动的方向 δx_i 始终与约束函数 $g(x)$ 梯度方向 $\nabla_x h(x)$ 正交，即

$$\delta x_i \cdot \mu \nabla_x h(x_F) = \delta x_i \cdot (-\nabla_x h(x_F))$$

6. 我们重新构造这个优化问题 $(\mathcal{P})$，并推广到多等式约束：

$$\min_{x\in\mathbb{R}^2} \ f(x)$$
$$s.t. \quad h_i(x)=0, \ \forall i\in\{1,2,\dots,l\}$$

我们定义拉格朗日函数 $\mathcal{L}$：

$$\mathcal{L}(x,\mu)=f(x)+\sum_{i=1}^{l}\mu_i h_i(x)$$

当 x^* 是局部最小值时，存在唯一的 μ^* 满足约束：

- $\nabla_x\mathcal{L}(x^*,\mu^*)=0 \qquad \Leftarrow \qquad \nabla_x f(x_F)+\mu\nabla_x h(x_F)=0 \quad ①$
- $\nabla_\mu\mathcal{L}(x^*,\mu^*)=0 \qquad \Leftarrow \qquad \dfrac{\partial\mathcal{L}(x,\mu_i)}{\partial\mu_i}=h_i(x)=0$ 满足约束条件
- $\nabla^2_{xx}\mathcal{L}(x^*,\mu^*)\succeq 0 \qquad \Leftarrow$ Hessain matrix 半正定：满足局部极小

Inequality Constraints

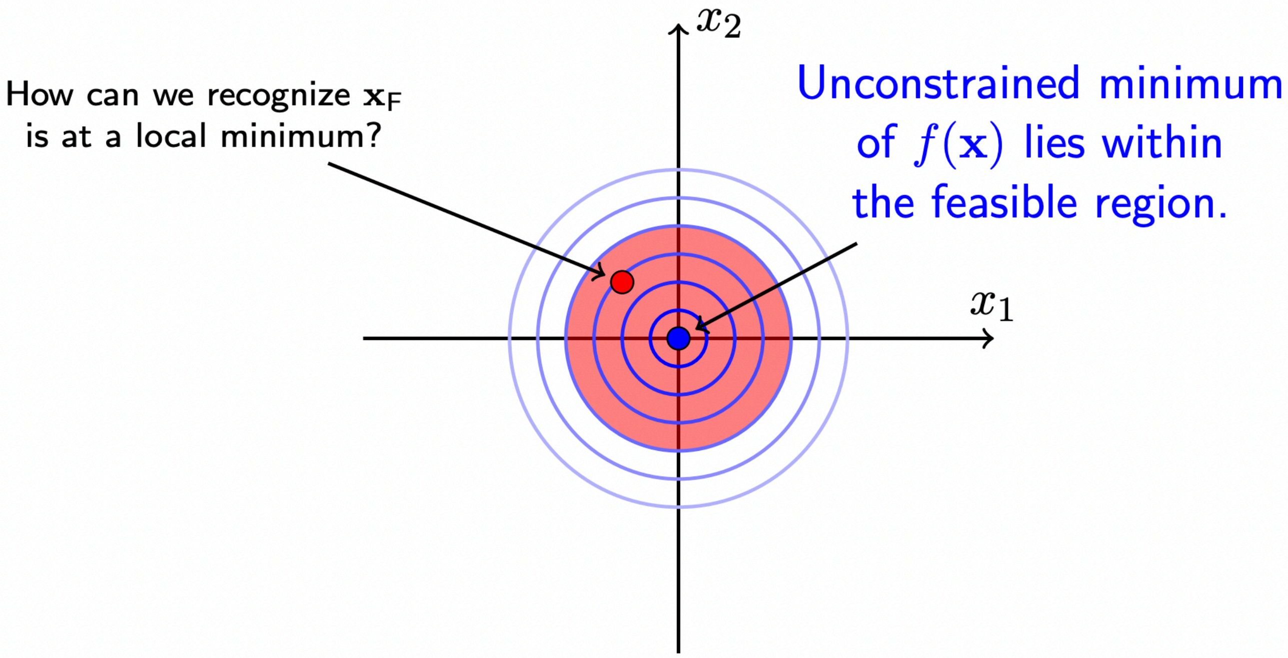
问题提出：

$$\min_{x\in\mathbb{R}^2} \ f(x)$$
$$s.t. \quad g_j(x)\le 0, \ \forall j\in\{1,2,\dots,m\}$$

Case 1 : 可退化到无约束

举例：

$$f(x)=x_1^2+x_2^2$$
$$s.t. \ h(x)=x_1^2+x_2^2-1$$

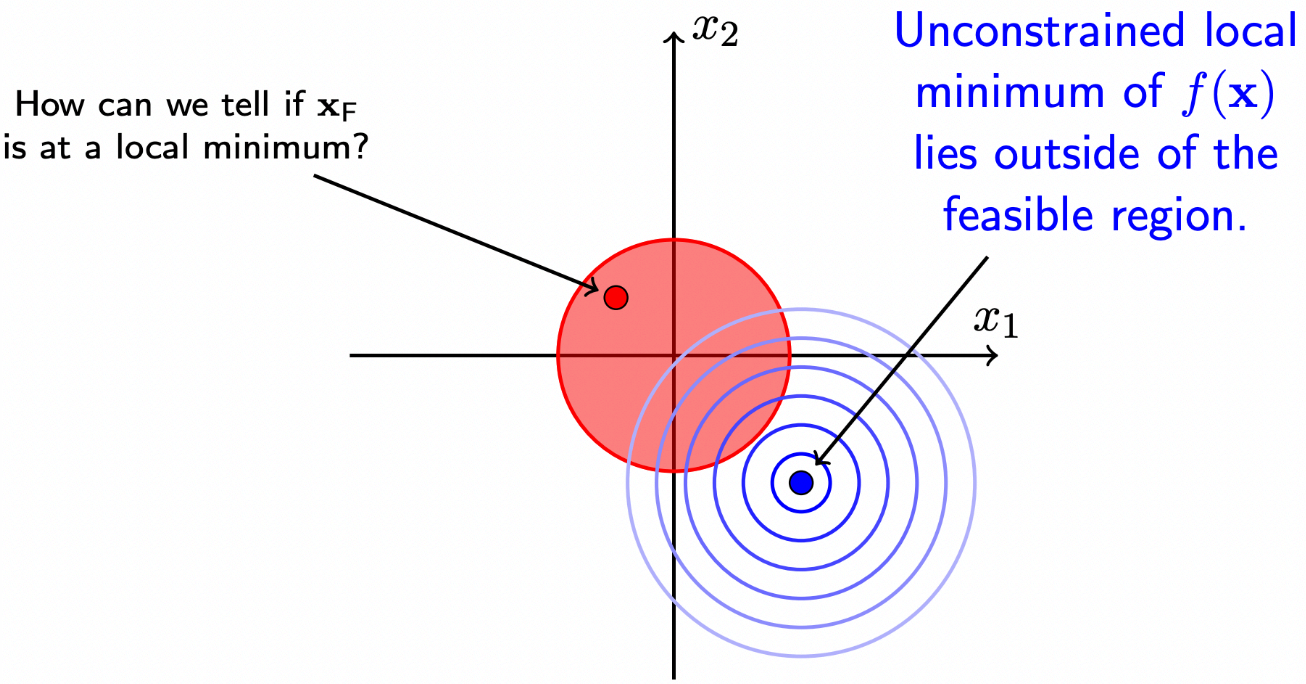


1. 可以从图像中看出，当 $f(x)$ 不加约束条件时的最优点为 $(0,0)$
2. 可行点 (feasible point) $x_F \in$ 可行域 (feasible region) 满足约束 $h(x) = x_1^2 + x_2^2 - 1 = 0$，图形上的表示就是在 $x_1^2 + x_2^2 - 1$ 的圆上
3. 当 $f(x)$ 加入约束条件时的最优点还是 $(0,0)$
4. 说明有无约束条件对这个问题的求解并没有影响
5. 此时我们就可以将这个约束优化问题退化成无约束问题：
   - $f(x)$ 在 x^* 处是零梯度 (zero gradient)：$\nabla_x f(x^*) = 0$
   - f 的 Hessian 矩阵在 x^* 处是半正定的 (positive definite)：$\nabla_{xx}^2 f(x^*) \succeq 0$

Case 2 : 不等式约束

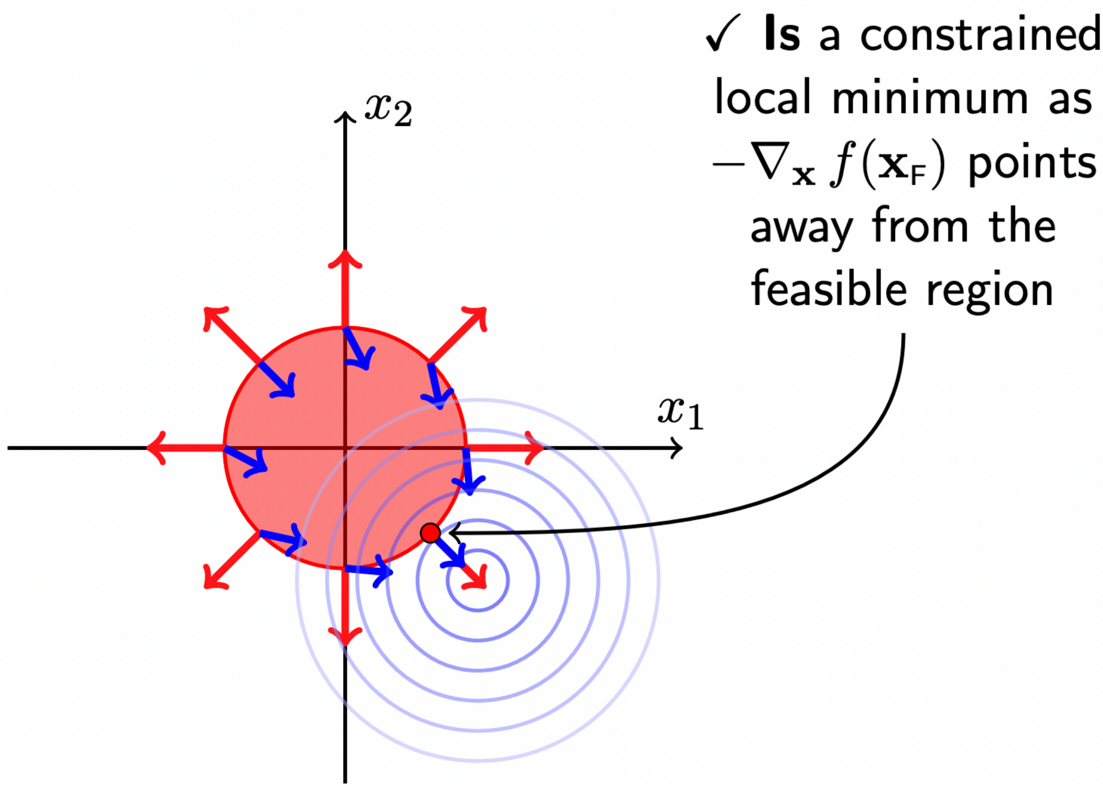
我们更改上一例题的条件：

$$f(x) = (x_1 - 1.1)^2 + (x_2 - 1.1)^2$$
$$s.t.\ h(x) = x_1^2 + x_2^2 - 1$$



1. 可以从图像中看出，当 $f(x)$ 不加约束条件时的最优点为 $(1.1, -1.1)$
2. 可行点 (feasible point) $x_F \in$ 可行域 (feasible region) 满足约束 $h(x) = x_1^2 + x_2^2 - 1 = 0$，图形上的表示就是在 $x_1^2 + x_2^2 - 1$ 的圆上
3. 当 $f(x)$ 加入约束条件时的最优点并不在原来的 $(1.1, -1.1)$ 点
4. 在这种情况下，极值在约束面上，即 $g(x^*) = 0$，此时就与等式约束条件一致了
5. 所以参考等式约束问题，最优值出现在目标函数 $f(x)$ 的梯度方向与约束函数 $h(x)$ 的梯度方向共线：

$$-\nabla_x f(x) = \lambda \nabla_x g(x), \lambda > 0 \quad \Rightarrow \quad \nabla_x f(x) + \lambda \nabla_x g(x) = 0 \quad ②$$



总结两种情况

$$\min_{x\in\mathbb{R}^2} \ f(x)$$
$$s.t. \quad g_j(x) \le 0,\ \forall j \in \{1,2,\dots,m\}$$

Case 1 : 无约束时的局部极小值的可行域【中】

1. $g(x^*) < 0 \qquad \Leftarrow \qquad$ 在可行域里
2. $\nabla_x f(x^*) = \nabla_x f(x^*) + \lambda_{(=0)} \nabla_x g(x^*) = 0 \qquad \Leftarrow \qquad \mathcal{KKT}$条件1
3. $\nabla^2_{xx} f(x^*) \succeq 0 \qquad \Leftarrow \qquad Hessian$矩阵是半正定的

Case 2 : 无约束时的局部极小值的可行域【外】

1. $g(x^*) = 0 \qquad \Leftarrow \qquad \mathcal{KKT}$条件3
2. $-\nabla_x f(x^*) + \lambda \nabla_x g(x^*) = 0,\ \lambda > 0 \qquad \Leftarrow \qquad \mathcal{KKT}$条件1
3. $\nabla^2_{xx} f(x^*) \succeq 0 \qquad \Leftarrow \qquad Hessian$矩阵是半正定的

$\mathcal{KKT}$ 条件4： x^* 是可行点

多等式和多不等式的 $\mathcal{KKT}$ 条件

给定一个函数 f，不等式约束 $g_1,\dots,g_m$ 和等式约束 $h_1,\dots,h_l$ 都在在定义域 $\Omega \subset \mathbb{R}^n$ 上的优化问题：

$$(\mathcal{P}):\ \min_{x\in\Omega} f(x)$$
$$s.t. \quad \begin{cases} h_i(x) = 0,\ \forall i \in \{1,2,\dots,l\} \\ g_j(x) \le 0,\ \forall j \in \{1,2,\dots,m\} \end{cases}$$

我们定义拉格朗日函数 $\mathcal{L}$:

$$\mathcal{L}(x,\lambda,\mu) = f(x) + \sum_{i=1}^{l} \mu_i h_i(x) + \sum_{j=1}^{m} \lambda_j g_j(x))$$

当 x^* 是局部最小值时，存在唯一的 μ^* 满足约束：

- $\nabla_x \mathcal{L}(x^*,\mu^*,\lambda^*) = 0 \qquad \Leftarrow \qquad \nabla_x f(x^*) + \sum_{i=1}^{l} \mu_i \nabla_x h_i(x^*) + \sum_{j=1}^{m} \lambda_j \nabla_x g_j(x^*) = 0$
- $\lambda_j^* \ge 0$ for $j = 1,\dots,m$
- $\lambda_j^* g_j(x^*) = 0$ for $j = 1,\dots,m$
- $g_j(x^*) \le 0$ for $j = 1,\dots,m$
- $g_i(x^*) = 0$ for $i = 1,\dots,l$
- $\nabla^2_{xx} \mathcal{L}(x^*,\lambda^*) \succ 0 \qquad \Leftarrow \qquad$ Hessain matrix 正定：满足局部极小