6-1 约束优化问题与规范条件

约束最优化问题数学模型

一般数学模型：

$$\min \quad f(\boldsymbol{x}) \tag{6.1.1a}$$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, \quad i \in \mathcal{E} = \{1, 2, \cdots, m_e\} \tag{6.1.1b}$$

$$c_i(\boldsymbol{x}) \geq 0, \quad i \in \mathcal{I} = \{m_e + 1, \ldots, m\} \tag{6.1.1c}$$

其中:

- $\boldsymbol{x} \in \mathbb{R}^n$， $f(\boldsymbol{x})$为目标函数, $c_i(\boldsymbol{x})(i \in \mathcal{E} \cup \mathcal{I})$为约束函数;
- $c_i(\boldsymbol{x}) = 0, i \in \mathcal{E}$与$c_i(\boldsymbol{x}) \geq 0, i \in \mathcal{I}$分别称为等式约束和不等式约束.

记可行域：

$$\mathcal{D} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \,\middle|\, c_i(\boldsymbol{x}) = 0, i \in \mathcal{E}; \quad c_i(\boldsymbol{x}) \geq 0, i \in \mathcal{I} \right\}. \tag{6.1.2}$$

求解问题(6.1.1)即在$\mathcal{D}$上寻求使得$f(\boldsymbol{x})$达到最小的可行点$\boldsymbol{x}^*$。

定义6.1.1 局部极小点与全局极小点

对约束优化问题(6.1.1)，若$\boldsymbol{x}^* \in \mathcal{D}$，存在$\varepsilon > 0$，对所有$\boldsymbol{x} \in \mathcal{N}(\boldsymbol{x}^*, \varepsilon) = \{\boldsymbol{x} \in \mathcal{D} | \|\boldsymbol{x} - \boldsymbol{x}^*\|_2 \leq \varepsilon\}$，始终有$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$成立，则称$\boldsymbol{x}^*$是问题(6.1.1)的**局部极小点**。

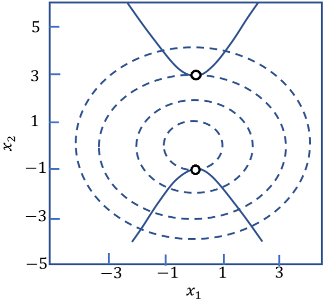
进一步，对$\forall \boldsymbol{x} \in \mathcal{N}(\boldsymbol{x}^*, \varepsilon)$且$\boldsymbol{x} \neq \boldsymbol{x}^*$，若$f(\boldsymbol{x}^*) < f(\boldsymbol{x})$始终成立，则称$\boldsymbol{x}^*$是**严格局部极小点**。

以上定义中若取ε足够大时，即有$\mathcal{D} = \mathcal{N}(\boldsymbol{x}^*, \varepsilon)$，则局部极小点为**全局极小点**。全局(严格)极小点也必定是局部(严格)极小点。

约束产生多极值

例6.1.1 考虑约束条件：

$$\begin{cases} \min & x_1^2 + x_2^2 \\ \text{s.t.} & (x_2 - 1)^2/4 - x_1^2 = 1 \end{cases}$$



$(0, -1)^T$为全局极小点,$(0, 3)^T$为局部极小点。

有效约束与非有效约束

定义6.1.2.

针对约束优化问题(6.1.1)中的约束函数，对任何点$\boldsymbol{x} \in \mathbb{R}^n$处的**有效约束集**或**有效集** $\mathcal{A}(\boldsymbol{x})$定义如下：

$$\mathcal{A}(\boldsymbol{x}) = \mathcal{E} \cup \mathcal{I}(\boldsymbol{x}) \tag{6.1.3}$$

其中$\mathcal{I}(\boldsymbol{x}) = \{i | c_i(\boldsymbol{x}) = 0, i \in \mathcal{I}\}$。

- $c_i(\boldsymbol{x})(i \in \mathcal{A}(\boldsymbol{x}))$是在$\boldsymbol{x}$点处的**有效约束**;
- $c_i(\boldsymbol{x})(i \notin \mathcal{A}(\boldsymbol{x}))$是在$\boldsymbol{x}$点处的**非有效约束**。

假定$\boldsymbol{x}^*$为问题(6.1.1)的局部最优解，相应的有效约束集为$\mathcal{A}^* = \mathcal{A}(\boldsymbol{x}^*)$。

基于有效约束集优化模型

忽略掉非积极约束后，问题(6.1.1)可转化为等式约束优化问题：

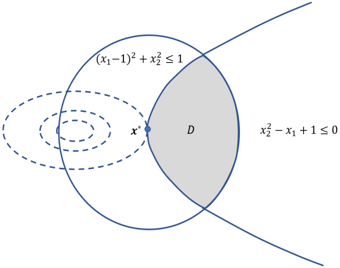
$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}), \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = 0,\ i \in \mathcal{A}^*. \end{aligned} \tag{6.1.4}$$

问题(6.1.4)比问题(6.1.1)更容易求得最优解。

例6.1.2考虑如下约束优化问题：

$$\begin{cases} \min & x_1^2 + x_2^2 \\ \text{s.t.} & (x_1-1)^2 + x_2^2 \le 1 \\ & x_2^2 - x_1 + 1 \le 0 \end{cases}$$

的最优解为$\boldsymbol{x}^* = (1,0)^T$.



去掉第一个约束(最优解处为非有效约束)，问题最优解不变。

例6.1.3考虑约束最优化问题：

$$\begin{cases} \min & x_1^2 + x_2^2 \\ \text{s.t.} & 1-(x_1-1)^2-(x_2-1)^2=0 \end{cases}$$

最优解为: $\boldsymbol{x}^* = (1-\frac{\sqrt{2}}{2}, 1-\frac{\sqrt{2}}{2})^T$.

计算梯度向量：

$$\boldsymbol{g}^* = \nabla f(\boldsymbol{x}^*) = (2-\sqrt{2}, 2-\sqrt{2})^T$$

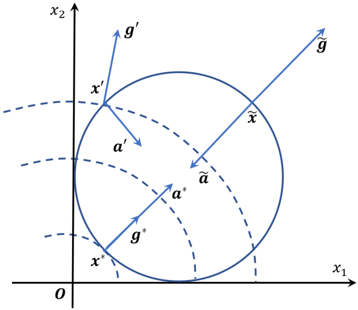
$$\boldsymbol{a}^* = \nabla c(\boldsymbol{x}^*) = (\sqrt{2}, \sqrt{2})$$

$\boldsymbol{g}^*$与$\boldsymbol{a}^*$共线, 即有:

$$\boldsymbol{g}^* = \lambda^* \boldsymbol{a}^*,$$

其中$\lambda^* = \sqrt{2}-1$.

极大点$\tilde{x}$处同样共线。



最优性条件初步认识

例6.1.3考虑约束最优化问题：

$$\begin{cases} \min & x_1^2 + x_2^2 \\ \text{s.t.} & 1 - (x_1 - 1)^2 - (x_2 - 1)^2 = 0 \end{cases}$$

最优解为: $\boldsymbol{x}^* = (1 - \frac{\sqrt{2}}{2}, 1 - \frac{\sqrt{2}}{2})^T$.

若将等式约束改为：

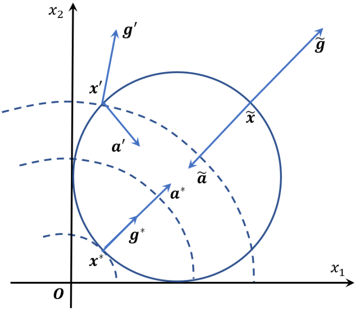
$$(x_1 - 1)^2 + (x_2 - 1)^2 - 1 = 0$$

约束函数梯度向量为：

$$\boldsymbol{a}^* = (-\sqrt{2}, -\sqrt{2})^T$$

此时$\lambda^* = 1 - \sqrt{2} < 0$. $\boldsymbol{g}^*$与$\boldsymbol{a}^*$仍然共线，但反向。

等式约束最优解处λ^*可正可负。



将前面例题中约束变更为：

$$1-(x_1-1)^2-(x_2-1)^2 \geq 0$$

在最优点$\boldsymbol{x}^*$处, $\boldsymbol{g}^*$与$\boldsymbol{a}^*$共线，既有:

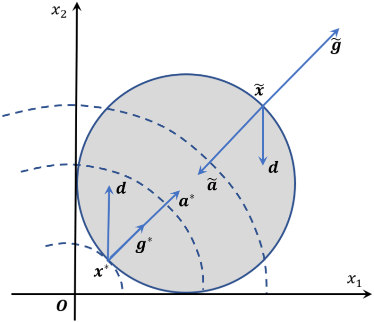
$$\boldsymbol{g}^* = \lambda^* \boldsymbol{a}^*$$

这里$\lambda^* = \sqrt{2}-1 > 0$.

若目标梯度向量与约束梯度向量反向，如在点$\tilde{\boldsymbol{x}}$处，存在可行下降方向$\boldsymbol{d}$满足：

$$\tilde{\boldsymbol{a}}^T\boldsymbol{d} > 0 \text{ 且 } \tilde{\boldsymbol{g}}^T\boldsymbol{d} < 0$$

因此$\tilde{\boldsymbol{x}}$非最优点。



综合以上分析，一般约束优化问题一阶最优性条件为：

$$\boldsymbol{g}^* = \sum_{i\in\mathcal{E}\cup\mathcal{I}^*} \lambda_i^* \boldsymbol{a}_i^*, \quad \lambda_i^* \geq 0, i \in \mathcal{I}^* \tag{6.1.x}$$

这里要求在最优解$\boldsymbol{x}^*$处起作用不等式约束对应的λ^*非负。

注意: 并非任意约束函数在最优点处一阶最优性条件都满足。

例6.1.4考虑约束优化问题：

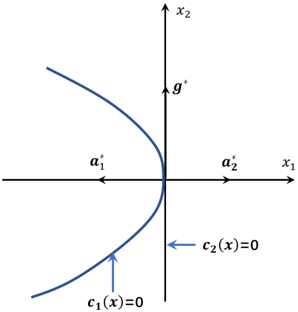
$$\begin{cases} \min & f(\boldsymbol{x}) = x_2 \\ \text{s.t.} & c_1(\boldsymbol{x}) = -x_1 - x_2^2 \geq 0 \\ & c_2(\boldsymbol{x}) = x_1 = 0 \end{cases}$$

该问题最优解为：$\boldsymbol{x}^* = (0,0)^T$.

另外：

$$\boldsymbol{g}^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \boldsymbol{a}_1^* = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \boldsymbol{a}_2^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

显然最优性条件不满足。



若要最优性条件满足，约束函数还需要满足某些条件，称为**约束规范条件**或**约束限制条件**。

可行方向

约束优化局部极小点取决于其目标函数值，还与该点附近其它可行点上的值有关。可行方向是研究最优性条件重要概念。

定义6.1.3 可行方向.

设$\boldsymbol{x}$为约束优化问题(6.1.1)的可行点，$\boldsymbol{d}$为$\mathbb{R}^n$中非零向量，若存在$\delta > 0$使得:

$$\boldsymbol{x} + \alpha \boldsymbol{d} \in \mathcal{D}, \forall \alpha \in [0, \delta],$$

成立，则称$\boldsymbol{d}$是可行域$\mathcal{D}$在$\boldsymbol{x}$处的**可行方向**。$\mathcal{D}$在点$\boldsymbol{x}$处的**可行方向集**:

$$\mathscr{F}_D(\boldsymbol{x}) = \left\{ \boldsymbol{d} \in \mathbb{R}^n \middle| \boldsymbol{x} + \alpha \boldsymbol{d} \in \mathcal{D}, \forall \alpha \in [0, \delta] \right\} \tag{6.1.5}$$

序列可行方向

定义6.1.4 序列可行方向.

设$\boldsymbol{x}$为约束优化问题(6.1.1)的可行点，若存在可行点列$\{\boldsymbol{x}_k\}$满足：$\boldsymbol{x}_k = \boldsymbol{x} + \alpha_k \boldsymbol{d}_k \in \mathcal{D}$, 这里有:

$$\boldsymbol{x}_k \neq \boldsymbol{x}, \boldsymbol{d}_k \to \boldsymbol{d}, \alpha_k(> 0) \to 0,$$

则称$\boldsymbol{d}$是可行域$\mathcal{D}$在$\boldsymbol{x}$处的序列可行方向。$\mathcal{D}$在$\boldsymbol{x}$处的序列可行方向集:

$$\mathscr{F}_S(\boldsymbol{x}) = \left\{ \boldsymbol{d} \in \mathbb{R}^n \middle| \boldsymbol{x} + \alpha_k \boldsymbol{d}_k \in \mathcal{D}, \forall k; \boldsymbol{d}_k \to \boldsymbol{d}, \alpha_k \to 0 \right\} \tag{6.1.6}$$

定义6.1.5 线性化可行方向.

设$\boldsymbol{x}$为约束优化问题(6.1.1)的可行点，$\boldsymbol{d}$为$\mathbb{R}^n$中非零向量, 若:

$$\boldsymbol{d}^T\nabla c_i(\boldsymbol{x}) = 0, i \in \mathcal{E};\quad \boldsymbol{d}^T\nabla c_i(\boldsymbol{x}) \geq 0, i \in \mathcal{I}(\boldsymbol{x})$$

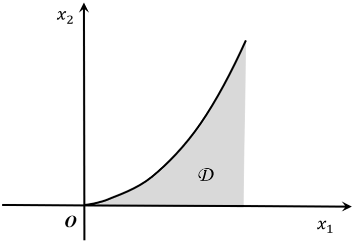
成立，则称$\boldsymbol{d}$是可行域$\mathcal{D}$在$\boldsymbol{x}$处的线性化可行方向。$\mathcal{D}$在$\boldsymbol{x}$处的线性化可行方向集:

$$\mathscr{F}_L(\boldsymbol{x}) = \left\{\boldsymbol{d} \in \mathbb{R}^n \middle| \boldsymbol{d}^T\nabla c_i(\boldsymbol{x}) = 0, i \in \mathcal{E};\quad \boldsymbol{d}^T\nabla c_i(\boldsymbol{x}) \geq 0, i \in \mathcal{I}(\boldsymbol{x})\right\} \tag{6.1.7}$$

例6.1.5 考虑约束

$$c_1(\boldsymbol{x}) = x_1^3 - x_2 \geq 0, c_2(\boldsymbol{x}) = x_2 \geq 0$$

分析$\boldsymbol{x} = (0,0)^{\mathrm{T}}$处序列可行方向与线性化可行方向。



取$\{\boldsymbol{x}_k\}$为$x_1^3 = x_2$上满足$x_2 > 0$的点列，令$\alpha > 0$，取:

$$\alpha_k = \frac{\|\boldsymbol{x}_k - \boldsymbol{x}\|}{\alpha} \to 0, \quad \boldsymbol{d}_k = \alpha\frac{\boldsymbol{x}_k - \boldsymbol{x}}{\|\boldsymbol{x}_k - \boldsymbol{x}\|} \to \boldsymbol{d} = (\alpha, 0)^{\mathrm{T}}$$

计算约束函数在$\boldsymbol{x}$处梯度：

$$\boldsymbol{a}^1 = \nabla c_1(\boldsymbol{x}) = (0,-1)^{\mathrm{T}}, \boldsymbol{a}^2 = \nabla c_2(\boldsymbol{x}) = (0,1)^{\mathrm{T}}$$

所以：$\mathscr{F}_S(\boldsymbol{x}) = \{(\alpha,0)|\alpha > 0\} \subseteq \mathscr{F}_L(\boldsymbol{x}) = \{(\alpha,0)^{\mathrm{T}}|\alpha \in \mathbb{R}\}$，二者不等。

引理 6.1.1.

可行域$\mathcal{D}$中点$\boldsymbol{x}$的三种可行方向集满足如下关系：

$$\mathscr{F}_D(\boldsymbol{x}) \subseteq \mathscr{F}_S(\boldsymbol{x}) \subseteq \mathscr{F}_L(\boldsymbol{x}). \tag{6.1.8}$$

证明： (1) 考虑$\boldsymbol{d} \in \mathscr{F}_D(\boldsymbol{x})$, 由可行方向定义知，存在$\alpha > 0$使得: $\boldsymbol{x} + \alpha \boldsymbol{d} \in \mathcal{D}$成立. 令$\boldsymbol{d}_k = \boldsymbol{d}$和$\alpha_k = \alpha/2^k, (k = 1, 2, \ldots)$则有:

$$\forall k, \boldsymbol{x} + \alpha_k \boldsymbol{d}_k \in \mathcal{D}\text{成立, 且显然有}\boldsymbol{d}_k \to \boldsymbol{d}, \alpha_k \to 0$$

故$\boldsymbol{d} \in \mathscr{F}_S(\boldsymbol{x})$, 由$\boldsymbol{d}$的任意性, 即知$\mathscr{F}_D(\boldsymbol{x}) \subseteq \mathscr{F}_S(\boldsymbol{x})$.

引理6.1.1证明

证明： (2) 考虑$\boldsymbol{d} \in \mathscr{F}_S(\boldsymbol{x})$, 若$\boldsymbol{d} = 0$, 显然$\boldsymbol{d} \in \mathscr{F}_L(\boldsymbol{x})$. 由序列可行方向定义, 存在序列$\boldsymbol{d}_k$和$\alpha_k > 0(k = 1, 2, \ldots)$使得:

$$\forall k, \boldsymbol{x} + \alpha_k \boldsymbol{d}_k \in \mathcal{D}\text{成立，}\quad\text{且}\boldsymbol{d}_k \to \boldsymbol{d}\text{和}\alpha_k \to 0.$$

考虑约束函数$c_i(\boldsymbol{x})$泰勒展开:

$$\begin{aligned} 0 &= c_i(\boldsymbol{x} + \alpha_k \boldsymbol{d}_k) = \alpha_k \boldsymbol{d}_k^T \nabla c_i(\boldsymbol{x}) + o(\|\alpha_k \boldsymbol{d}_k\|), \quad i \in \mathcal{E} \\ 0 &\le c_i(\boldsymbol{x} + \alpha_k \boldsymbol{d}_k) = \alpha_k \boldsymbol{d}_k^T \nabla c_i(\boldsymbol{x}) + o(\|\alpha_k \boldsymbol{d}_k\|), \quad i \in \mathcal{I}(\boldsymbol{x}) \end{aligned}$$

两边除以α_k, 令$k \to \infty$, 即可得到：$\boldsymbol{d} \in \mathscr{F}_L(\boldsymbol{x})$, 所以$\mathscr{F}_S(\boldsymbol{x}) \subseteq \mathscr{F}_L(\boldsymbol{x})$. ■

约束规范条件

引入序列可行方向与线性化可行方向的目的是建立约束规范条件，从而建立最优性条件。约束规范条件有多种，常用的一种为KT(Kuhn-Tucker)约束规范条件，即：

$$\mathscr{F}_S = \mathscr{F}_L$$

以上约束规范条件不容易检验，但可以证明以下条件之一成立:

- $\mathcal{A}$中所有约束为线性约束;
- $\{\boldsymbol{a}_i, i \in \mathcal{A}\}$线性无关.

则KT约束规范条件成立。

例6.1.4中在点$(0,0)^T$处$\boldsymbol{a}_1$与$\boldsymbol{a}_2$线性相关，不满足KT约束规范条件。

约束规范条件

另一种约束规范条件为正则性假设，它建立在序列可行方向与下降方向的基础之上。

定义6.1.6 下降方向与正则性.

设$\boldsymbol{g}$为$f(\boldsymbol{x})$在$\boldsymbol{x}$处的梯度向量，定义：

$$\mathscr{D} = \mathscr{D}(\boldsymbol{x}) = \{\boldsymbol{d} \in \mathbb{R}^n | \boldsymbol{d}^T \boldsymbol{g} < 0\}$$

为$f(\boldsymbol{x})$在$\boldsymbol{x}$处的下降方向集合，向量$\boldsymbol{d}$称为$\boldsymbol{x}$处的下降方向。正则性假设为：

$$\mathscr{F}_S \cap \mathscr{D} = \mathscr{F}_L \cap \mathscr{D}$$

正则性假设考虑$\mathscr{F}_S$与$\mathscr{F}_L$中下降方向的部分，显然这个条件比KT约束规范条件弱。

根据序列可行方向与下降方向定义的概念，可以建立下面最优性必要条件：

定理6.1.2 几何一阶必要条件.

若$\boldsymbol{x}^*$为问题(6.1.1)的局部最优解，则：

$$\mathscr{F}_S^* \cap \mathscr{D}^* = \emptyset \tag{6.1.9}$$

其中: $\mathscr{F}_S^* = \mathscr{F}_S(\boldsymbol{x}^*), \mathscr{D}^* = \mathscr{D}(\boldsymbol{x}^*)$.

证明：需证明$\boldsymbol{d} \in \mathscr{F}_S^*, \boldsymbol{d} \notin \mathscr{D}^*$. 由$\mathscr{F}_S^*$的定义，存在可行点列$\{\boldsymbol{x}_k\}$, $\boldsymbol{x}_k = \boldsymbol{x}^* + \alpha_k \boldsymbol{d}_k$, 使得$\alpha \to 0, \boldsymbol{d}_k \to \boldsymbol{d}$. 由Taylor公式有：

$$f_k = f^* + \alpha_k {\boldsymbol{g}^*}^T \boldsymbol{d}_k + o(\alpha)$$

一阶必要条件

证明续：由于$\boldsymbol{x}^*$为局部最优解，当k充分大时，有$f_k \geq f^*$，即

$$\boldsymbol{g}^{*T}\boldsymbol{d}_k + o(1) \geq 0$$

令$k \to 0$，得$\boldsymbol{g}^{*T}\boldsymbol{d} \geq 0$，所以$\boldsymbol{d} \notin \mathscr{D}^*$。 ■

约束规范条件

问题(6.1.1)的最优解$\boldsymbol{x}^*$处不存在可行的下降方向, 在正则性假设下

- 对等式约束优化问题：线性化可行方向条件$\boldsymbol{a}_i^T\boldsymbol{d} = 0(i \in \mathcal{E})$和下降方向条件${\boldsymbol{g}^*}^T\boldsymbol{d} < 0$不会同时成立，当且仅当：

$$\boldsymbol{g}^* = \sum_{i\in\mathcal{E}} \lambda_i^* \boldsymbol{a}_i^*$$

- 对不等式约束优化问题：线性化可行方向条件$\boldsymbol{a}_i^T\boldsymbol{d} = 0(i \in \mathcal{I}^*)$和下降方向条件${\boldsymbol{g}^*}^T\boldsymbol{d} < 0$不会同时成立，当且仅当：

$$\boldsymbol{g}^* = \sum_{i\in\mathcal{I}^*} \lambda_i^* \boldsymbol{a}_i^*, \lambda_i^* \geq 0, i \in \mathcal{I}^*$$

若λ_i^*出现负值，则与$\boldsymbol{x}^*$为极小点矛盾。