最优化技术原理及应用

王冰川
bingcwang@csu.edu.cn
https://intleo.csu.edu.cn/index.html
中南大学自动化学院

约束优化基础

主要参考中国科学院大学Xiao Wang老师PPT内容

主要内容

主要内容

1. 介绍
2. 约束规范性条件
3. 一阶最优条件
4. 对偶
5. 参考资料

➢约束优化问题的通常描述如下

$$\min_{x\in\Re^n} f(x)$$

$$\text{s.t.}\ \ c_i(x) = 0, i \in \mathcal{E},$$

$$c_i(x) \geq 0, i \in \mathcal{I}$$

➢$f(x)$代表目标函数， $c_i(x)$代表约束条件，都是平滑的实值函数； $\Re^n$是定义域，也就是搜索空间； $\mathcal{E}$和$\mathcal{I}$是两个有限索引集合； $c_i(x), i \in \mathcal{E}$和$c_i(x), i \in \mathcal{I}$分别代表等式约束和不等式约束

- 满足约束条件的点的集合，也就是可行集合Ω定义如下

$$\Omega = \{x | c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}$$

- 因此，原约束优化问题可以表示如下

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

- 如果$x^* \in \Omega$并且存在邻域$\aleph$，使得下述条件成立，则x^* 是约束优化问题的一个局部最优解

$$f(x^*) \leq f(x), \forall x \in \Omega \cap \aleph$$

- 如果$x^* \in \Omega$并且存在邻域$\aleph$，使得下述条件成立，则x^* 是约束优化问题的一个严格局部最优解

$$f(x^*) < f(x), \forall x \in \Omega \cap \aleph$$

- 如果x^*是$\Omega \cap \aleph$中唯一的最优解，则x^*被称作孤立的局部最优解

中南大学 CENTRAL SOUTH UNIVERSITY

- 平滑性能够保证目标函数和约束具有合理可预测的特性，从而算法可以更好地选择搜索方向

- 非光滑边界通常可以通过光滑约束函数的集合描述

$$\|x\|_1 = |x_1| + |x_2| \le 1$$

$$\Leftrightarrow x_1 + x_2 \le 1, x_1 - x_2 \le 1, x_2 - x_1 \le 1, -x_1 - x_2 \le 1$$

- 非光滑、无约束优化问题有时可以重新表述为光滑约束问题

$$\min f(x) = \max(x^2, x),$$

$$\Leftrightarrow \min t, \qquad s.t.\ t \ge x, t \ge x^2$$

有效集（Active Set）

- 可行解x处的有效集$\mathcal{A}(x)$由等式约束索引，和满足$c_i(x)=0$的不等式约束索引组成

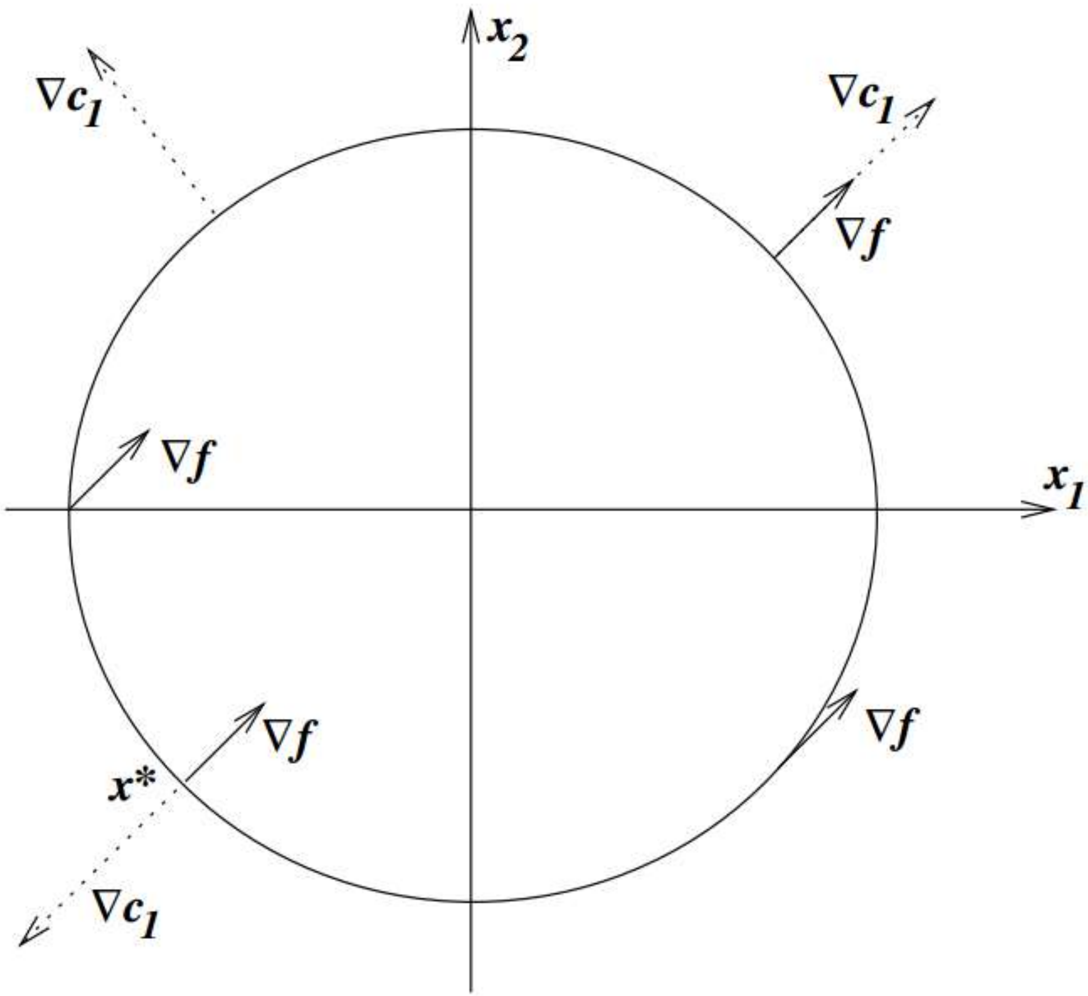
$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} | c_i(x) = 0\}$$

- 可行解x处，如果 $c_i(x)=0$，则不等式约束 $i \in \mathcal{I}$被称active；如果满足$c_i(x) > 0$，则称之为inactive

- 接下来通过几个例子定性说明局部最优解所需要满足的一些条件

例子：单个等式约束

$$\min x_1 + x_2 \quad s.t.\, x_1^2 + x_2^2 - 2 = 0$$



- 为了保持约束的可行性$c_1(x) = 0$，任何小的（但非零）步长s需要满足 $c_1(x+s) = 0$：

$$0 = c_1(x+s) \approx c_1(x) + \nabla c_1(x)^T s = \nabla c_1(x)^T s$$

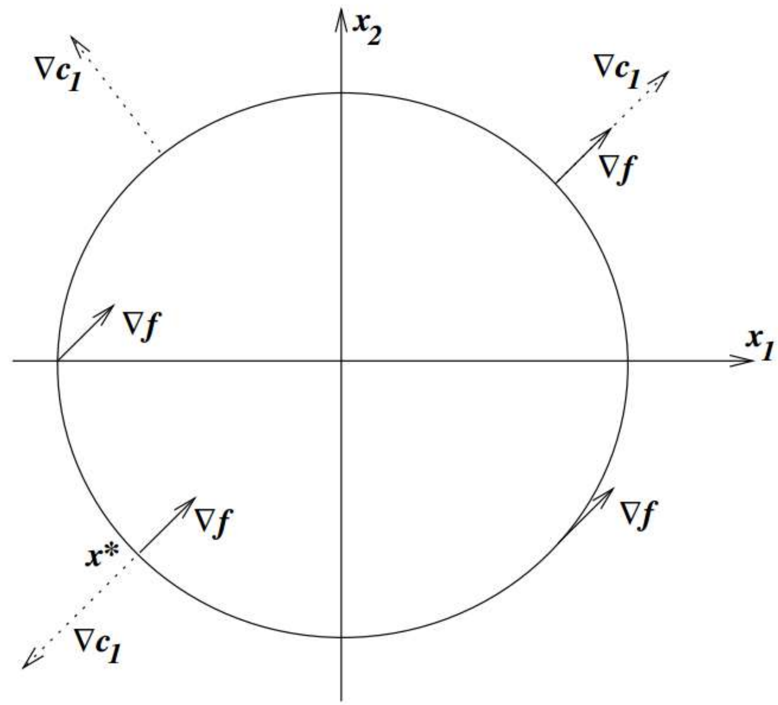
- 如果s能够导致 f 减小，需要满足：

$$0 > f(x+s) - f(x) \approx \nabla f(x)^T s$$

- 只有当满足如下条件的方向向量s不存在时，x^*是一个局部最优解：

$$\nabla c_1(x^*)^T s = 0 \text{ and } \nabla f(x^*)^T s < 0$$

例子：单个等式约束



- 形式化理解，只有当$\nabla f(x^*)$和$\nabla c_1(x^*)$平行时，上述条件才成立，也就是在x^*处满足：

$$\nabla f(x) = \lambda \nabla c_1(x)$$

例子：单个等式约束

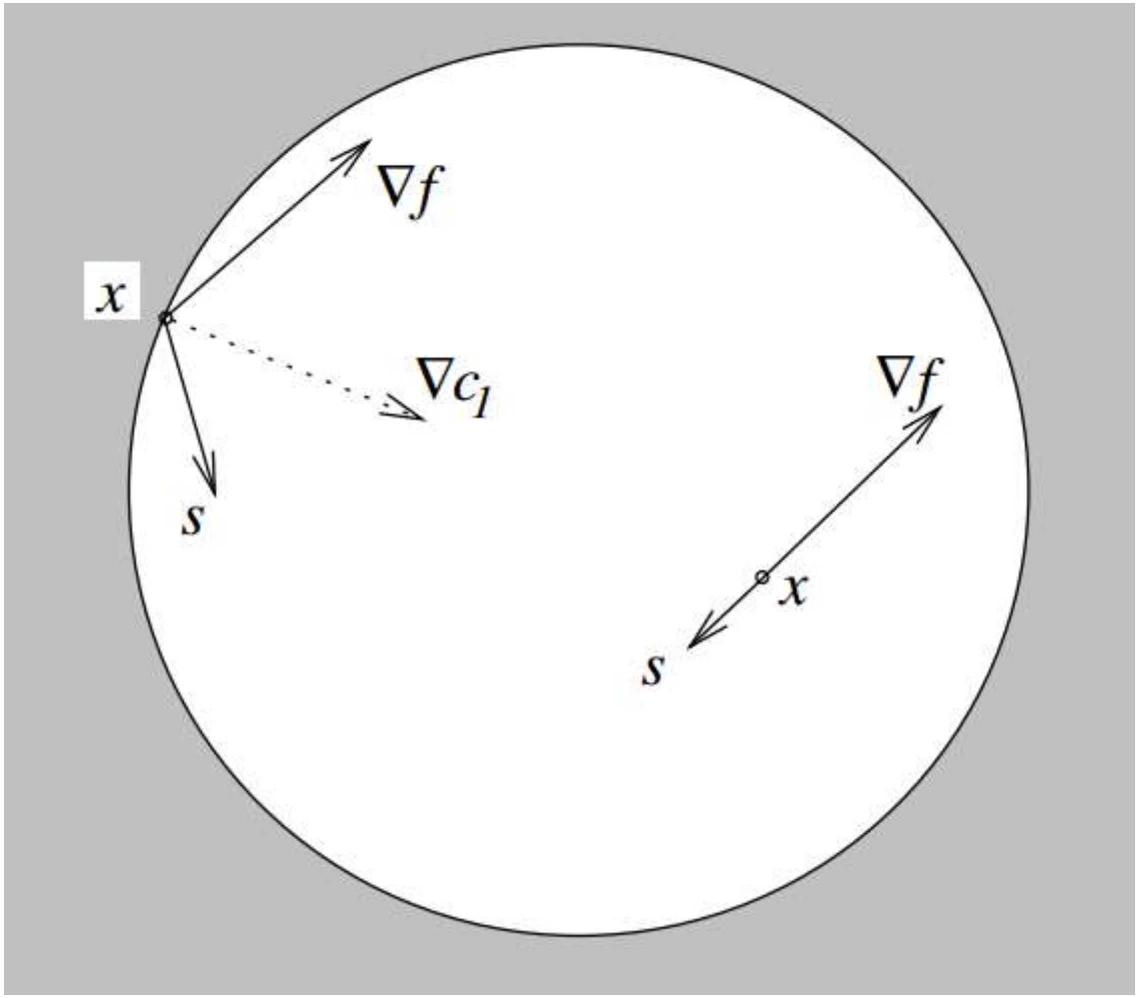
- 拉格朗日函数：

$$\mathcal{L}(x,\lambda) = f(x) - \lambda c_1(x)$$

- $\nabla f(x) = \lambda \nabla c_1(x) \Leftrightarrow \nabla_x \mathcal{L}(x,\lambda) = 0$，此处$\lambda$也被称作拉格朗日乘子
- 上述观察结果表明，可以通过搜索拉格朗日函数的稳定点来搜索等式约束问题的解
- 这是一个必要条件！

例子：单个不等式约束

$$\min x_1 + x_2 \quad s.t.\ 2 - x_1^2 - x_2^2 \geq 0$$



- 和之前一样，如果能够找到既保留可行性又将使目标函数f下降的方向，则给定的可行解不是极值点

- 如果s能够导致f减小，需要满足：

$$f(x)^T s < 0$$

- 如果能够保持可行性，需要满足：

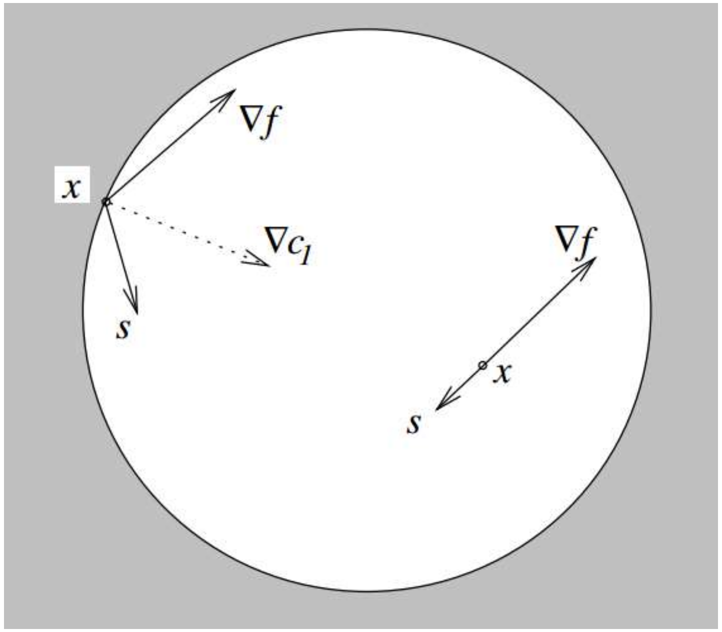
$$0 \le c_1(x+s) \approx c_1(x) + \nabla c_1(x)^T s$$

$$c_1(x) + \nabla c_1(x)^T s \ge 0$$

- Case I：x在圆内，也就是$c_1(x) > 0$，此时当步长足够小的时候，就能保持可行性；当$\nabla f(x) \neq 0$时，可以设置α充分小，使得同时满足可行性和下降条件：

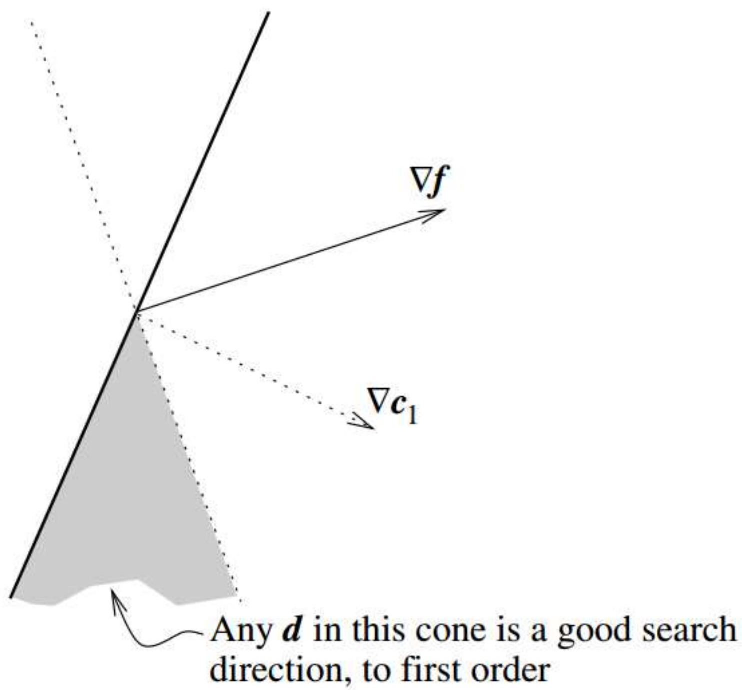
$$s = -\alpha \nabla f(x)$$

- 也就是说，当$\nabla f(x) = 0$时，不能同时满足两个条件



➢Case II：x在圆上，也就是$c_1(x)=0$，能够使得两个条件满足的区域如下图所示。也就是说，当$\nabla f(x)$和$c_1(x)$同方向时，不能同时保持可行性和目标函数下降两个条件：

$$\nabla f(x)=\lambda c_1(x),\lambda\geq 0$$



- 同样，总结上述两种情况，并通过拉格朗日形式表示：

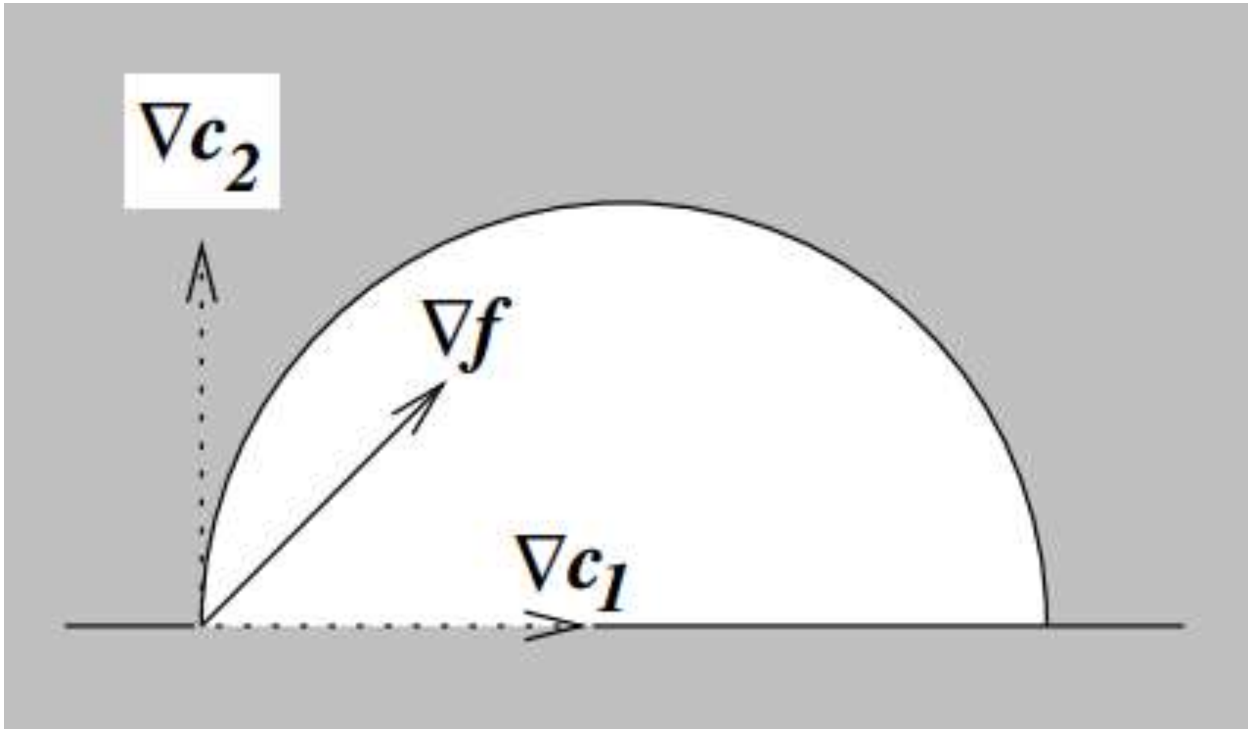
$$\nabla_x \mathcal{L}(x, \lambda) = 0, \qquad \lambda \geq 0$$

$$\lambda c_1(x) = 0$$

- 对于case I，$c_1(x) > 0$，所以$\lambda = 0$，因此以上条件变成了$\nabla f(x) = 0$；
- 对于case II，$\lambda \geq 0$，因此以上条件变成$\nabla f(x) = \lambda c_1(x)$，$\lambda \geq 0$

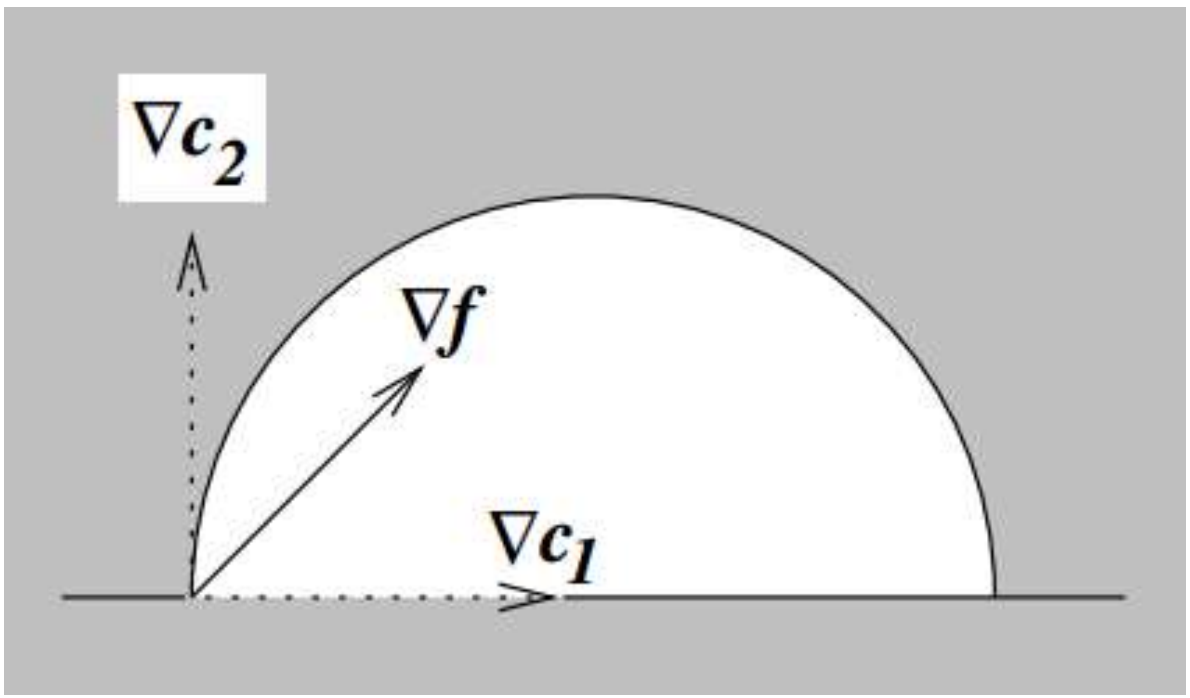
例子：两个不等式约束

$$\min x_1 + x_2 \quad s.t.\ 2 - x_1^2 - x_2^2 \geq 0, x_2 \geq 0$$



$(-\sqrt{2}, 0)$处目标函数和约束条件的梯度方向

➤从下图可以看出，没有方向向量s能够对$(-\sqrt{2}, 0)$改进，因此其是一个局部最优解



- 同样，在$x^* = (-\sqrt{2}, 0)$处能够找到满足拉格朗日条件的值

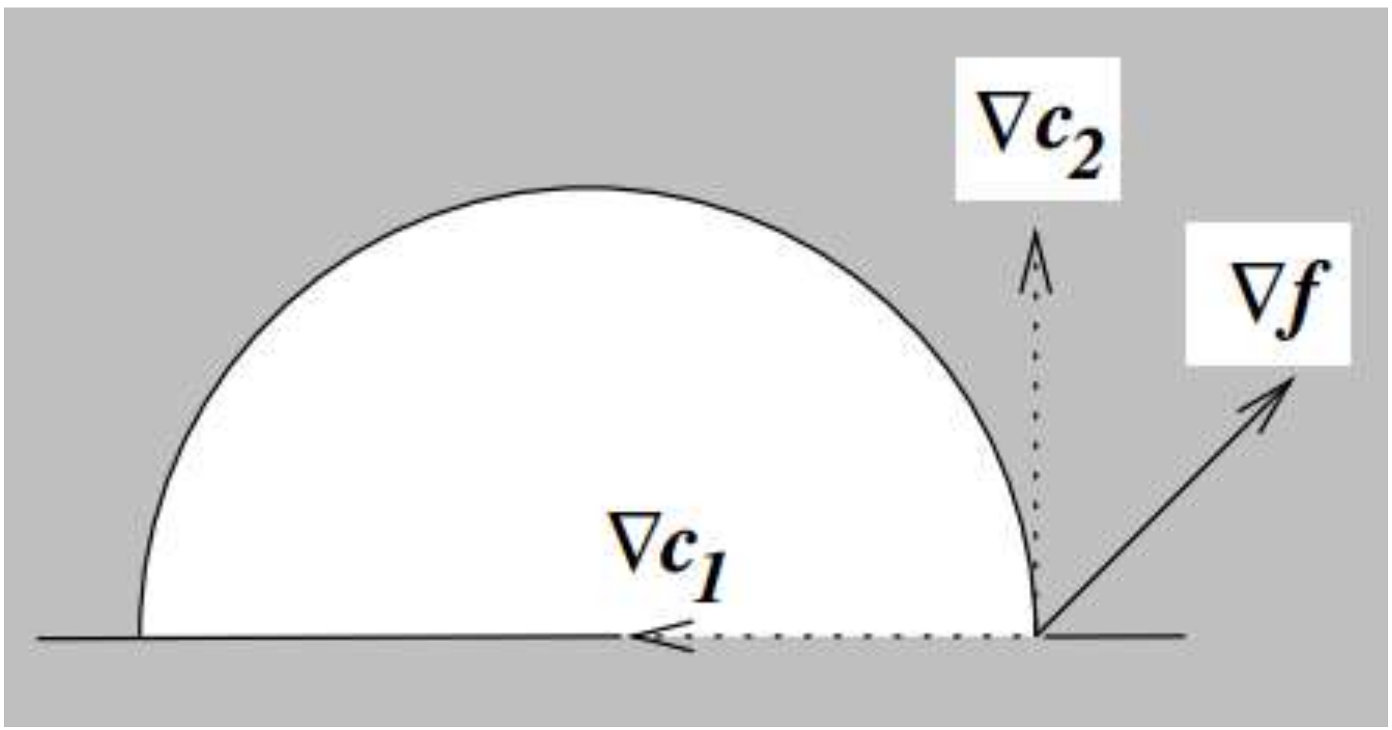
$$\mathcal{L}(x,\lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x)$$

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \lambda_1 c_1(x^*) = 0, \lambda_2 c_2(x^*) = 0$$

- $\lambda^* = [\frac{1}{2\sqrt{2}}, 1]$是满足上述条件的值

- 请同学们自己检验$(\sqrt{2}, 0)$处能否找到满足拉格朗日条件的λ值

➢请同学们自己检验$(\sqrt{2},0)$处能否找到满足拉格朗日条件的λ值



主要内容

- 以上例子通过检查f和c_i的一阶导数来确定是否可以从给定的可行解x采取可行的下降步骤

- 具体来说，利用f和c_i在x处的一阶泰勒级数展开得到一个目标和约束都是线性的近似问题

- 然而，只有当线性近似捕获了相关点 x 附近可行集的基本几何特征时，这种方法才有意义

- 如果在x附近，线性化与可行集根本不同，那么线性近似能不能产生有关原始问题的有用信息

$$x^2 = 0$$

在$x = 0$处，线性化对应于整个空间，而可行集是单个点

- 因此，需要对在x处active的约束c_i的性质做出合理假设，以确保线性化近似与x附近的可行集相似

- 约束规范化是确保约束集Ω与其线性化近似在x邻域中保持相似所作出的假设

- 给定可行解x，有效约束集$\mathcal{A}$，线性化可行方向$\mathcal{F}(x)$定义如下：

$$\mathcal{F}(x)=\left\{d\,\middle|\,\begin{array}{c} d^T c_i(x)=0,\forall i\in\mathcal{E} \\ d^T c_i(x)\geq 0,\forall i\in\mathcal{I}\cap\mathcal{A}(x)\end{array}\right\}$$

- 线性化可行方向集取决于约束函数$c_i(\mathrm{x}),i\in\mathcal{E}\cup\mathcal{I}$的定义

- $2-x_1^2+x_2^2=0$和$(2-x_1^2+x_2^2)^2=0$（两种定义）

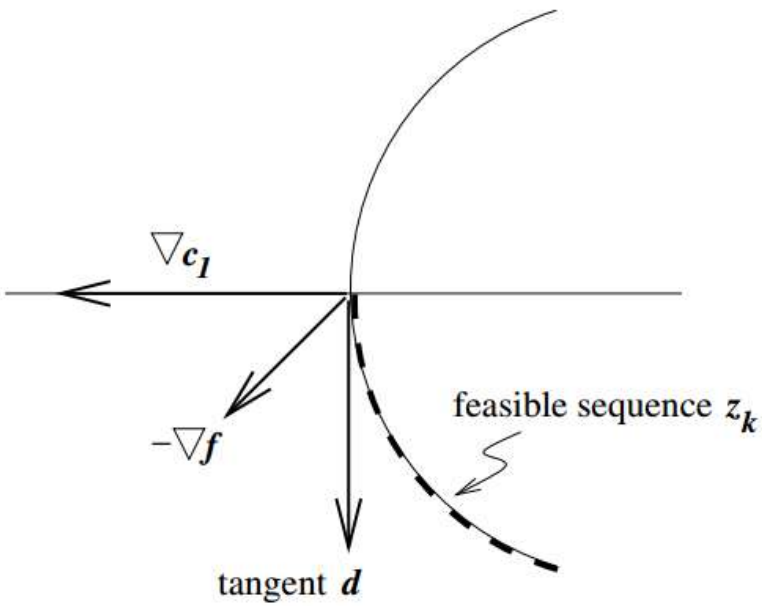
- 给定可行解x，如果 $\{z_k\} \in \Omega$并且对于所有足够大的k满足$z_k \longrightarrow x$，则$\{z_k\}$被称作逼近x的可行序列
- 如果向量d是可行序列$\{z_k\}$的极限方向，则称向量d在点x处与Ω相切。也就说，存在一个接近x的可行序列$\{z_k\}$和一个正数序列$\{t_k\}$，$z_k \longrightarrow 0$，满足如下条件

$$\lim_{k\to\infty} \frac{z_k - x}{t_k} = d$$

- Ω在x处的所有切向量的集合称为切锥：$T_\Omega(x)$
- 切锥的定义不依赖于Ω的代数形式，仅依赖于其几何形状

切锥：例子

$$\min x_1 + x_2 \quad s.t.\, x_1^2 + x_2^2 - 2 = 0$$



- $x = (-\sqrt{2}, 0)^T$
  - $z_k = \left[-\sqrt{2 - 1/k^2} \quad -1/k\right], t_k = \|z_k - x\| \Rightarrow t = (0, -1)^T$
  - $z_k = \left[-\sqrt{2 - 1/k^2} \quad 1/k\right], t_k = \|z_k - x\| \Rightarrow t = (0,1)^T$
- $T_\Omega(x) = \mathcal{F}(x) = \{(0, d_2)^T | d_2 \in \Re\}$

- 如果可行域定义形式如下：

$$\Omega = \{x | (x_1^2 + x_2^2 - 2)^2 = 0\}$$

- 该可行域不变，但是代数描述形式变了，在此基础上计算$\mathcal{F}(x)$

$$0 = d^T c(x) = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}^T \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

此时，$T_\Omega(x) \neq \mathcal{F}(x) = \Re^2$

- 说明线性化可行方向依赖于代数描述形式

- 约束规范性条件是线性化可行集$\mathcal{F}(x)$与切锥$T_\Omega(x)$相似的前提条件

- 事实上，大多数规范性条件能够确保这两个集合是相同的

- 这些条件确保线性化Ω代数描述所得到的集合$\mathcal{F}(x)$能够捕获x处的几何特征$T_\Omega(x)$

线性独立
约束规范性条件

- LICQ (Linear Independent Constraint Qualification)：给定解x和有效集$\mathcal{A}(x)$，LICQ条件如下：

  有效约束集梯度$\{\nabla c_i(\mathrm{x}), i \in \mathcal{A}(x)\}$线性独立

- 通常情况下，当LICQ成立时，则所有active约束梯度都不能为零

主要内容

一阶必要条件

- 定义拉格朗日函数如下：

$$\mathcal{L}(x,\lambda)=f(x)-\sum_{i\in\mathcal{E}\cup\mathcal{I}}\lambda_i c_i(x)$$

- 假设x^*是一个局部最优解并且满足LICQ条件，则存在拉格朗日乘子$\lambda^*=\{\lambda_i^*\},i\in\mathcal{E}\cup\mathcal{I}$使得如下条件在$(x^*,\lambda^*)$处成立

$$\nabla_x\mathcal{L}(x^*,\lambda^*)=0,$$

$$c_i(x^*)=0,\forall i\in\mathcal{E},$$

$$c_i(x^*)\geq 0,\forall i\in\mathcal{I},$$

$$\lambda_i^*\geq 0,$$

$$\lambda_i^*\,c_i(x^*)=0,\forall i\in\mathcal{E}\cup\mathcal{I}$$

- 以上被称作Karush-Kuhn-Tucker (KKT)条件

严格互补

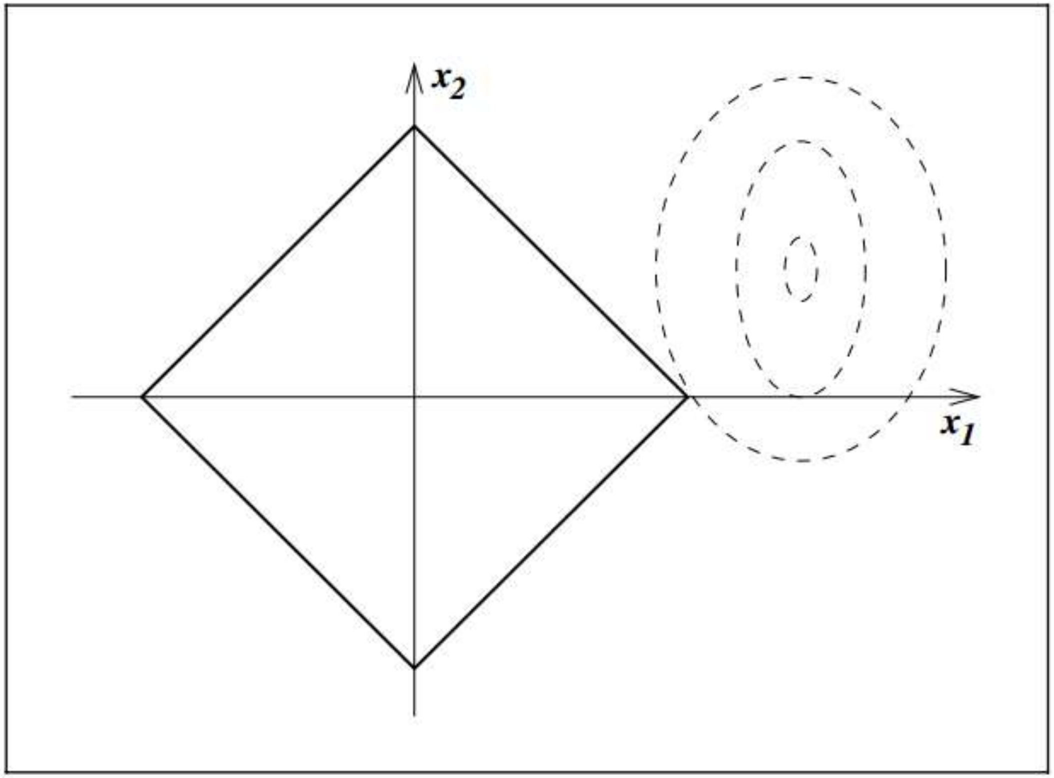
- 给定局部最优解x^*和满足KKT条件的λ^*，如果对于每个不等式约束$i \in \mathcal{I}$，$c_i(x^*)$和λ_i^*刚好只有一个等于0，则严格互补条件成立；换句话说，对于$i \in \mathcal{I} \cap \mathcal{A}(x)$，$\lambda_i^* > 0$

- 一般来说，满足严格互补条件可以使得算法更容易确定有效集$\mathcal{A}(x)$，并快速收敛到解 x^*

- 对于给定问题和解局部最优解x^*，可能有许多向量λ^*满足KKT条件；然而，满足LICQ条件的λ^*是唯一的

一个例子

$$\min_{x} \ (x_1 - 1.5)^2 + (x_2 - 0.5)^4 \quad s.t. \begin{bmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{bmatrix} \geq 0$$



一个例子

$$\min_{x}\ (x_1-1.5)^2+(x_2-0.5)^4 \quad s.t. \begin{bmatrix} 1-x_1-x_2 \\ 1-x_1+x_2 \\ 1+x_1-x_2 \\ 1+x_1+x_2 \end{bmatrix} \geq 0$$

- 一个最优解$x^*=(1,0)^T$，前两个约束条件active，而后两个约束条件inactive，因此有

$$\lambda_3^* = \lambda_4^* = 0$$

- 因为有

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -0.5 \end{bmatrix}, \nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \nabla c_2(x^*) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

- 所以$\lambda_1^* = 0.75, \lambda_2^* = 0.25$

主要内容

对偶

- 基于对偶理论提出了包括增广拉格朗日在内的很多重要算法
- 就其一般性而言，对偶理论不但对非线性规划很重要，也为凸非光滑优化，甚至是离散优化提供了重要的工具
- 对偶理论展示了如何根据原始优化问题构造相应的替代问题
- 在某些情况下，对偶问题比原始问题更容易计算；在一些情况下，对偶可轻松获得原始问题最优值的下界

- 考虑如下约束优化问题（原问题）

$$\min_{x\in\mathfrak{R}^n} f(x), s.t. c(x) \geq 0$$

其中， $c(x) = [c_1(x), c_2(x), \dots, c_m(x)]^T$，并且$f$和$-c_i$都是凸函数

- 上述问题的对偶问题定义如下：

$$\max_{\lambda\in\mathfrak{R}^m} q(\lambda), s.t. \lambda \geq 0$$

其中， $q(\lambda) \equiv \inf_x \mathcal{L}(x,\lambda)$

一个例子

- 考虑如下约束优化问题（原问题）

$$\min_{(x_1,x_2)} 0.5(x_1^2+x_2^2), s.t. x_1-1\geq 0$$

- 其拉格朗日函数如下：

$$\mathcal{L}(x,\lambda)=0.5(x_1^2+x_2^2)-\lambda(x_1-1)$$

- 因为f和c都是凸函数，当满足如下条件时，取到$\inf_x$

$$x_1-\lambda=0, x_2=0$$

- 因此有对偶问题

$$\max_{\lambda\geq 0} -0.5\lambda^2+\lambda$$

- 在许多问题中，某些λ值的下确界为$-\infty$；$q(\lambda)$的定义域是使得其为有限值的λ集合：

$$\mathcal{D} \equiv \{\lambda | q(\lambda) > -\infty\}$$

- 计算下确界需要找到$\mathcal{L}(x,\lambda)$的全局极小值；在实践中，对于一些λ，这可能可能极其困难

- 当f和$-c_i$都是凸函数，并且$\lambda \geq 0$时， $\mathcal{L}(.,\lambda)$也是凸函数，因此$q(\lambda)$的计算更有效

弱对偶定理

- $q(\lambda)$是凹函数，其定义域D是凸集合

- 对于原问题任意一个可行解$\bar{x}$和$\bar{\lambda} \geq 0$，有$q(\bar{\lambda}) \leq f(\bar{x})$

- 弱对偶定理是天然成立的，不需要其他前提

$$\min_{x\in\Re^n} f(x), s.t.\, c(x) \geq 0$$

- 回顾一下该问题的KKT条件

$$\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda} = 0$$

$$c(\bar{x}) \geq 0$$

$$\bar{\lambda} \geq 0$$

$$\bar{\lambda}_i c_i(\bar{x}) = 0, i = 1, \ldots, m$$

- 其中， $\nabla c(\bar{x}) = [\nabla c_1(\bar{x}), \nabla c_2(\bar{x}), \ldots, \nabla c_m(\bar{x})]$

对偶定理

- $\bar{x}$ 是原问题的一个解， f和$-c_i$都是凸函数并且可微，使得$(\bar{x}, \bar{\lambda})$满足KKT条件的任意$\bar{\lambda}$是对偶问题的解

- 假设f和$-c_i$都是凸函数并且连续可微， $\bar{x}$是原问题的一个解，并且满足LICQ条件；假设$\hat{\lambda}$是对偶问题的一个解并且 $\inf_x \mathcal{L}(x, \hat{\lambda})$的下确界在$\hat{x}$处取到，进一步假设$\mathcal{L}(., \hat{\lambda})$是一个严格凸函数，有$\bar{x} = \hat{x}$（ $\hat{x}$是原问题唯一解）， $f(\bar{x}) = \mathcal{L}(\bar{x}, \hat{\lambda})$

Wolfe对偶

$$\max_{\lambda,x} \mathcal{L}(x,\lambda), \qquad s.t. \nabla_x \mathcal{L}(x,\lambda) = 0, \lambda \geq 0$$

- 假设f和$-c_i$都是凸函数并且连续可微，$(\bar{x},\bar{\lambda})$是原问题的解组合，并且满足LICQ条件，则$(\bar{x},\bar{\lambda})$是Wolfe对偶问题的解

中南大學 CENTRAL SOUTH UNIVERSITY

- 推导如下二次规划问题的对偶问题和Wolfe对偶问题：

$$\min_{x\in\Re^n}\frac{1}{2}x^TGx+c^Tx, s.t.\ Ax-b\geq 0$$

Thanks for the attentions!

Q&A