最优化方法

凸集与凸函数

Definition 1 **凸集**:
设集合$D \subset \mathbb{R}^n$ ，若$\forall x, y \in D$ ，

$$\lambda x + (1-\lambda) y \in D, \forall 0 \le \lambda \le 1$$

则称集合D为**凸集**。
凸集的性质：设D_1, D_2是凸集，则$D_1 \cap D_2, D_1 + D_2, D_1 - D_2, \alpha D_1$都是凸集
Examples:

1. 超平面：$H = \{x | p^T x = \alpha\}$是凸集，其中$p \in \mathbb{R}^n$是非零向量，称为超平面的法向量，$\alpha \in \mathbb{R}$。
2. 闭半空间：$H^- = \{x | p^T x \le \alpha\}$和$H^+ = \{x | p^T x \ge \alpha\}$ 是凸集。
3. 开半空见：$H_0^- = \{x | p^T x < \alpha\}$和$H_0^+ = \{x | p^T x > \alpha\}$ 是凸集。
4. 多面集：由有限个闭半空间的交组合成的集合$D = \{x | p_i^T x \le \beta_i, i = 1, \ldots, m\}$。其中$p_i$是非零向量，$\beta_i$是数。*多面集是一个凸闭集。*
   *Statement 1*$D \subset \mathbb{R}^n$是凸集的充分必要条件是D中任意m个点$x^{(i)}(i = 1, 2, \ldots, m)$的凸组合仍然属于$D$，即：

   $$\sum_{i=1}^{m} \alpha_i x^{(i)} \in D, \alpha_i \ge 0 (i = 1, 2, \ldots, m), \sum_{i=1}^{m} \alpha_i = 1$$

   Definition 2 分离，严格分离
   设$D_1, D_2 \subset \mathbb{R}^n$ 为两非空凸集
   若存在非零向量$\alpha \in \mathbb{R}^n$和$\beta \in \mathbb{R}$，使

   $$D_1 \subset H^+ = \{x \in \mathbb{R}^n | \alpha^T x \ge \beta\} D_2 \subset H^- = \{x \in \mathbb{R}^n | \alpha^T x \le \beta\}$$

   则称超平面

   $$H = \{x \in \mathbb{R}^n | \alpha^T x = \beta\}$$

   分离集合D_1, D_2 。
   若存在非零向量$\alpha \in \mathbb{R}^n$和$\beta \in \mathbb{R}$，使

   $$D_1 \subset H_0^+ = \{x \in \mathbb{R}^n | \alpha^T x > \beta\} D_2 \subset H_0^- = \{x \in \mathbb{R}^n | \alpha^T x < \beta\}$$

   则称超平面$H = \{x \in \mathbb{R}^n | \alpha^T x = \beta\}$严格分离$D_1, D_2$ ，其中H_0^+, H_0^-分别表示集合H^+, H^-的内部。
   Statement 2 投影定理
   设$D \subset \mathbb{R}^n$是非空闭凸集，$y \in \mathbb{R}^n$但$y \notin D$ ，则

(1)存在唯一的点$\bar{x} \in D$，使得集合D到点y的距离最小，即

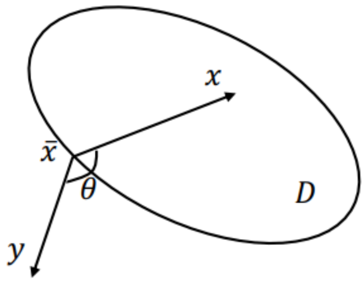
$$\|\bar{x} - y\| = \inf_{x \in D} \|x - y\|$$

(2)$\bar{x} \in D$是点y到集合D的最短距离点的充分必要条件为

$$(x - \bar{x})^{\mathrm{T}}(\bar{x} - y) \geq 0, \forall x \in D$$

或写成

$$\langle x - \bar{x}, y - \bar{x} \rangle \leq 0, \forall x \in D$$



Statement 3 存在超平面严格分离点与凸集

定理5.3 设$D \subset \mathbf{R}^n$为非空闭凸集，$y \in \mathbf{R}^n$，$y \notin D$，则存在非零向量$\alpha \in \mathbf{R}^n$和实数β，使得

$$\alpha^{\mathrm{T}} x \leq \beta < \alpha^{\mathrm{T}} y, \forall x \in D$$

成立，即存在超平面$H = \{x \in \mathbf{R}^n | \alpha^{\mathrm{T}} x = \beta\}$严格分离点$y$与凸集$D$。

*Definition 3*凸函数

定义5.3 设函数$f(x)$在凸集D上有定义，如果对任意x，$y \in D$和任意$\lambda \in [0,1]$，有

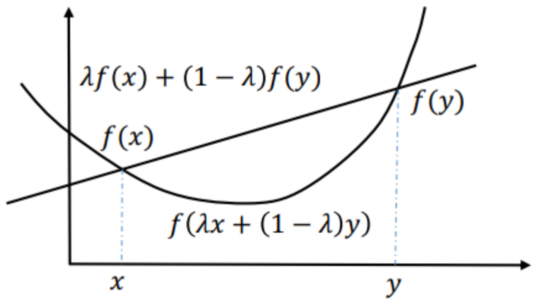
$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

则称$f(x)$是凸集D上的凸函数

如果对任意x，$y \in D, x \neq y$和任意有$\lambda \in (0,1)$有

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$$

则称$f(x)$是凸集D上的严格凸函数



7

凸函数的性质（类似于凸集）

(1) 如果f是定义在凸集D上的凸函数，实数$\alpha \geq 0$，则αf也是凸集D上的凸函数

(2) 如果f_1，f_2是定义在凸集D上的凸函数，则$f_1 + f_2$也是凸集D上的凸函数

(3) 如果$f_i(x)(i = 1, \cdots, m)$是非空凸集D上的凸函数，则$f(x) = \max\limits_{1 \leq i \leq m} |f_i(x)|$也是凸集$D$上的凸函数

(4) 如果$f_i(x)(i = 1, \cdots, m)$是非空凸集D上的凸函数，则$f(x) = \sum_{i=1}^{m} \alpha_i f_i(x)$也是凸集$D$上的凸函数，其中$\alpha_i \geq 0(i = 1, \cdots, m)$

Statement 4 设x^*是凸规划问题的一个局部最优解，那么x^*也是全局最优解。进一步，如果目标函数严格凸，那么x^*是唯一全局最优解。
Statement 5 设$f(x)$是定义在非空开凸集D的可微函数，则
$f(x)$是D上的凸函数 $\Longleftrightarrow f(y) \geq f(x) + \nabla f(x)^T(y-x), \forall x, y \in D$
$f(x)$是D上的严格凸函数 $\Longleftrightarrow f(y) > f(x) + \nabla f(x)^T(y-x), \forall x, y \in D, x \neq y$
Statement 6 设$f(x)$是定义在非空开凸集D的二阶可微函数，则
$f(x)$是D上的凸函数 $\Longleftrightarrow Hesse_f == \nabla^2 f(x)$在$D$上半正定，即

$$对每一个x \in D, y^T \nabla^2 f(x) y \geq 0, \forall y \in \mathbb{R}^n$$

如果$Hesse_f$在D上正定，则$f(x)$是D上的严格凸函数，**反之不亦然。**

最优化问题

一般形式

$$P: min f(x), x \in \mathbb{R}^n (目标函数)$$

$$s.t.\, h_i(x) = 0 i = 1, 2, \ldots, m (等式约束)$$

$$h_i(x) \geq 0 (不等式约束)$$

tips: 实际应用中，将所有目标函数统一成求**最小值的形式。**
Definition 4 可行解
满足等式约束和不等式约束的x成为可行解（也称可行点或容许点）。
Definition 5 可行域
全体可行解构成的集合为可行域（或称容许集），记为D，即：

$$D = \{x | h_i(x) = 0, i = 1, \ldots, m, h_i(x) \geq 0, i = m+1, \ldots, p, x \in \mathbb{R}^n\}$$

若h_i为连续函数，则D为闭集。
Definition 6 整体最优解（严格整体最优解）
若$x^* \in D, \forall x \in D, f(x^*) \leq f(x)$则称$x^*$为最优化问题(P)的整体最优解。
若$x^* \in D, \forall x \in D \backslash \{x^*\}, f(x^*) < f(x)$则称$x^*$为最优化问题(P)的严格整体最优解。

Definition 7 局部最优解（严格局部最优解）

若$x^* \in D, \exists N_\epsilon(x^*), \forall x \in D \cap N_\epsilon(x^*), f(x^*) \le f(x)$则称$x^*$为最优化问题(P)的整体最优解。

若$x^* \in D, \exists N_\epsilon(x^*), \forall x \in D \cap N_\epsilon(x^*) \backslash \{x^*\}, f(x^*) < f(x)$则称$x^*$为最优化问题(P)的严格整体最优解。

局部最优解不一定是整体最优解；求解最优化问题，就是再求可行域上的整体最优解。一般来说，整体最优解是很难求出的，往往只能求出局部最优解。

Definition 8 局部极小点，全局极小点，严格局部极小点，严格全局极小点（略，见PPT5.Page17）

Definition 9 下降方向

定义5.9 设$f(x)$为定义在空间$\mathbf{R}^n$上的连续函数，点$\bar{x} \in \mathbf{R}^n$，若对于方向$s \in \mathbf{R}^n$存在$\delta > 0$使成立

$$f(\bar{x} + \alpha s) < f(\bar{x}), \forall \alpha \in (0, \delta)$$

则称s为$f(x)$在$\bar{x}$处的一个下降方向。在点$\bar{x}$处的所有下降方向的全体记为$D(\bar{x})$.

Statement 7 设函数$f(x)$在点x_0处连续可微，如存在非零向量$s \in \mathbb{R}^n$ 使成立

$$\nabla f(x_0)^T s < 0$$

，则s是$f(x)$在x_0处的一个下降方向。

Statement 8~12，一阶必要条件 二阶必要条件 二阶充分条件 凸最优性定理

定理5.8 (一阶必要条件) 设$f: \mathrm{D} \subset \mathbf{R}^n \to \mathbf{R}^1$在开集$D$上连续可微，若$x^* \in \mathrm{D}$是$\min\limits_{x \in \mathbf{R}^n} f(x)$的局部极小点，则

$$g(x^*) = \nabla f(x^*) = 0$$

定理5.9 (二阶必要条件) 设$f: \mathrm{D} \subset \mathbf{R}^n \to \mathbf{R}^1$在开集$D$上二阶连续可微，若$x^* \in \mathrm{D}$是$\min\limits_{x \in \mathbf{R}^n} f(x)$的局部极小点，则

$$g(x^*) = 0, G(x^*) = \nabla^2 f(x^*) \ge 0$$

定理5.10 (二阶充分条件) 设$f: \mathrm{D} \subset \mathbf{R}^n \to \mathbf{R}^1$在开集$D$上二阶连续可微，则$x^* \in D$是$f$的一个严格局部极小点的充分条件是

$$g(x^*) = 0 \text{和} G(x^*) \text{是正定矩阵}$$

定理5.11 (凸最优性定理) 设$f: \mathrm{D} \subset \mathbf{R}^n \to \mathbf{R}^1$是凸函数，且$f \in C^1$. 则$x^*$是总体极小点的充分必要条件是$g(x^*) = 0$

迭代法基本格式：

1. 给定初始点$x^{(0)}$，令$k = 0$
2. 如果$x^{(k)}$满足对最优解估计的终止条件，停止迭代。

3. 确定一个改善$x^{(k)}$的修正量$s^{(k)}$
4. 得到一个更优的估计$x^{(k+1)}=x^{(k)}+s^{(k)}$
5. $k++, goto(2)$
   如果一个算法只有当初始点充分接近最优解时，产生的点列才收敛，则这种算法为局部收敛的算法。如果对任意初始点，都收敛，则是具有全局收敛的算法。由于一般情况最优解是未知的，所有只有具有全局收敛的算法才有实用意义。
   Definition 9 收敛速度（与方程迭代法收敛速度相似）

定义5.9 设向量序列$\{x^{(k)}\}\subset \mathbf{R}^n$收敛于$x^*$，定义误差序列

$$e_k = x^{(k)} - x^*$$

如果存在正的常数C和r使成立

$$\lim_{k\to\infty}\frac{\|e_{k+1}\|}{\|e_k\|^r}=C$$

则称序列$\{x^{(k)}\}$ r阶收敛于x^*(以C为因子)

- 当$r=1, 0<C<1$时称为线性收敛，这时的误差序列满足

$$\|e_{k+1}\|\le C\|e_k\|$$

- 当$r=1, C=0$时，称序列$\{x^{(k)}\}$超线性收敛于x^*，超线性收敛是一种比线性收敛更快的收敛，在上述收敛率的定义中，所有$r>1$的收敛都属于超线性收敛

黄金分割法

1. 选取初始数据，确定搜索区间$[a_0,b_0]$，和精度要求δ>0。计算初始点
   $\lambda_0=a_0+(1-0.618)(b_0-a_0)$ $\mu_0=a_0+0.618(b_0-a_0)$ 计算$\phi(\lambda_0),\phi(\mu_0)$, 令k=0
2. 比较$\phi(\lambda_k),\phi(\mu_k)$，若$\phi(\lambda_0)>\phi(\mu_0)$，转3，否则转4
3. 若$b_k-\lambda_k\le\delta$，则停止计算，输出μ_k; 否则，令

$$a_{k+1}:=\lambda_k, b_{k+1}:=b_k, \lambda_{k+1}:=\mu_k, \phi(\lambda_{k+1})=\phi(\mu_k), \mu_{k+1}:=a_{k+1}+0.618(b_{k+1}-a_{k+1})$$

   计算$\phi(\mu_{k+1}), k++$，转2。
4. 若$\mu_k-a_k\le\delta$，则停止计算，输出λ_k; 否则，令

$$a_{k+1}:=a_k, b_{k+1}:=\mu_k, \mu_{k+1}:=\lambda_k, \phi(\mu_{k+1})=\phi(\lambda_k), \lambda_{k+1}:=a_{k+1}+0.382(b_{k+1}-a_{k+1})$$

   计算$\phi(\lambda_{k+1}), k++$，转2。

总结：即每一次取两个黄金分割点，取舍掉较大的一个，形成新的区间，进行新一轮搜索。

二分法

1. 设初始区间$[a_1,b_1]$，第k步的搜索区间$[a_k,b_k]$，满足$\phi'(a_k)\le 0, \phi'(b_k)\ge 0$
2. 取中点c_k，$if \phi'(c_k)<0, then a_{k+1}=a_k, b_{k+1}=c_k; else, then a_{k+1}=c_k, b_{k+1}=b_k$。
   一直进行2，知道搜索区间长度小于预定的限度为止。

二分法收敛速度为**线性**，收敛比1/2。

$$(b_1 - a_1)(1/2)^n \le \delta$$