第八章　算法

算法概念

一．下降迭代算法

迭代： 从一点$x^{(k)}$出发，按照某种规则A，求出后继点$x^{(k+1)}$，用$k+1$代替k，重复以上过程，得到一个解的序列$\{x^{(k)}\}$，若该序列有极限点x^*，即

$$\lim_{k\to\infty}\left\|x^{(k)}-x^*\right\|=0$$

则称它收敛于x^*。

下降： 在每次迭代中，后继点处的函数值要有所减少。

下降迭代算法的步骤：

1.选定某一初始点$x^{(0)}$，置$k=0$。

2.确定搜索方向$d^{(k)}$。

3.从$x^{(k)}$出发，沿方向$d^{(k)}$求步长λ_k，以产生下一个迭代点$x^{(k+1)}$。

4.检查$x^{(k+1)}$是否为极小点或近似极小点，若是，则停止迭代；否则，令$k:=k+1$，返回2。

选取搜索方向是最关键的一步，各种算法的区别，主要在于确定搜索方向的方法不同。

确定步长λ_k的主要方法：

1.令它等于某一常数。

2.可接受点算法，即只要能使目标函数值下降，可任意选取步长λ_k。

3.基于沿搜索方向使目标函数值下降最多，即沿射线

$$x = x^{(k)} + \lambda d^{(k)}$$

求目标函数$f(x)$的极小

$$f(x^{(k)} + \lambda_k d^{(k)}) = \min f(x^{(k)} + \lambda d^{(k)}).$$

由于这项工作是求以λ为变量的一元函数的极小点，故常称这一过程为（最优）一维搜索，这样确定的步长为最佳步长。

定理: 设目标函数$f(x)$具有一阶偏导数，$x^{(k+1)}$由下列规则产生:

$$\begin{cases} f(x^{(k)}+\lambda_k d^k)=\min\limits_{\lambda} f(x^{(k)}+\lambda d^k) \\ x^{(k+1)}=x^{(k)}+\lambda_k d^k \end{cases}$$

则有$\nabla f(x^{(k+1)})^T d^k=0$。

证明: 记$\varphi(\lambda)=f(x^{(k)}+\lambda d^k)$

若λ_k是最优步长，则λ_k为$\varphi(\lambda)$的驻点。

$\therefore \quad \varphi'(\lambda_k)=0$

而$\varphi'(\lambda_k)=\nabla f(x^{(k)}+\lambda_k d^k)^T d^k$

$\therefore \quad \nabla f(x^{(k)}+\lambda_k d^k)^T d^k=0$

二．算法映射

定义： 给定集合$X \subset E^n$,记其幂集（即所有子集构成的集合）为2^X，称集值映射$A: X \to 2^X$为一个算法映射(algorithm mapping).

例： 考虑标准形式的线性规划

$$(LP) \quad \min\left\{cx \mid Ax = b, x \geq 0\right\},$$

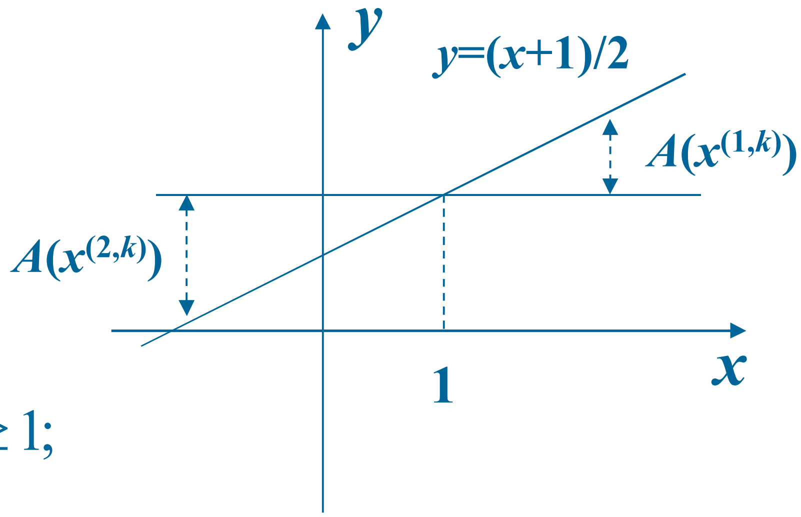
令$X = \{x \in \mathbb{R}^n \mid x为LP的基本可行解\}$，若定义算法映射

$A(x) = \{y \in \mathbb{R}^n \mid y为LP的基本可行解, 并且y和x的基矩阵是相邻的\}$,

那么对于任意一个基本可行解$x^{(0)} \in X$，迭代格式$x^{(k+1)} \in A(x^k)$就生成一个相邻的基本可行解序列。

例： 考虑下列非线性规划：

$$\begin{cases} \min x^2 \\ s.t. \quad x \geq 1. \end{cases}$$



定义算法映射：

$$A(x) = \begin{cases} \left[1, \frac{1}{2}(x+1)\right] & x \geq 1; \\ \left[\frac{1}{2}(x+1), 1\right] & x < 1. \end{cases}$$

利用算法A可以产生不同的点列：

$$\left\{3, 2, \frac{3}{2}, \frac{5}{4}, \cdots\right\}, \quad \left\{3, \frac{3}{2}, \frac{9}{8}, \frac{33}{32}, \cdots\right\}, \quad \left\{3, \frac{5}{3}, \frac{7}{6}, \frac{25}{24}, \cdots\right\}$$

解集合

把满足某些条件的点集定义为**解集合**．当迭代点属于该集合时，停止迭代.

常用的解集合：

$$\Omega=\left\{\bar{x} \mid \|\nabla f(\bar{x})\|=0\right\}$$

$$\Omega=\left\{\bar{x} \mid \bar{x}\text{为}KKT\text{点}\right\}$$

$$\Omega=\left\{\bar{x} \mid \bar{x}\in S, f(\bar{x})\le b\right\},$$

其中b是某个可接受的目标函数值。

下降函数

定义： 设$\Omega \subset X$为解集合，A为X上的一个算法，$\alpha(x)$是定义在X上的连续实函数，若满足

1. 当$x \notin \Omega$且$y \in A(x)$时，$\alpha(y) < \alpha(x)$
2. 当$x \in \Omega$且$y \in A(x)$时，$\alpha(y) \le \alpha(x)$

则称α是关于解集合Ω和算法A的下降函数。

一般地，求解非线性规划问题

$$
\begin{cases} \min \ f(x) \\ s.t. \quad x \in S \end{cases}
$$

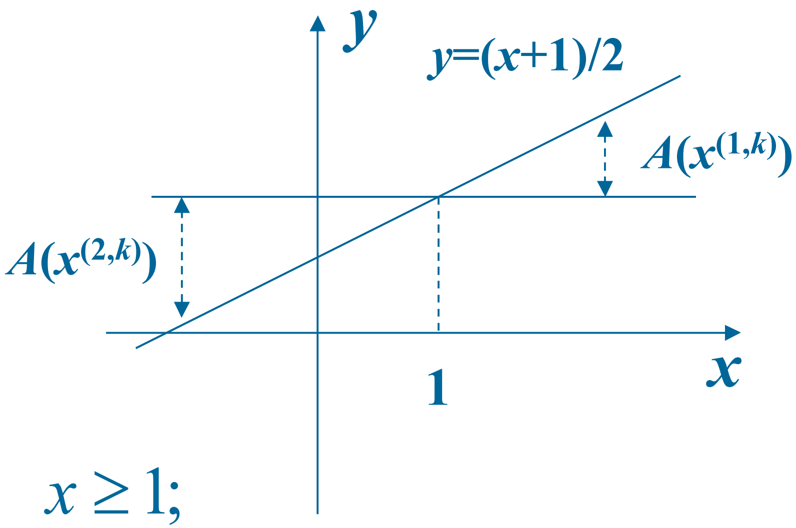
时，通常取$\|\nabla f(x)\|$或$f(x)$为下降函数。

闭映射(closed mapping)

定义: 给定两个非空闭集$X \subset E^m$和$Y \subset E^n$，以及一个集值映射$A: X \to 2^Y$，假设点列$\{x^{(k)}\}$和$\{y^{(k)}\}$满足$x^{(k)} \in X$，$y^{(k)} \in A(x^{(k)})$。若$x^{(k)} \to x$，$y^{(k)} \to y(k \to +\infty)$蕴涵着$y \in A(x)$，则称映射$A$在点$x \in X$处是闭的。若映射$A$在集合$S \subseteq X$上每一点是闭的，则称$A$在集合$S$上是闭映射。

考虑下列非线性规划：

$$\begin{cases} \min x^2 \\ s.t. \quad x \geq 1. \end{cases}$$



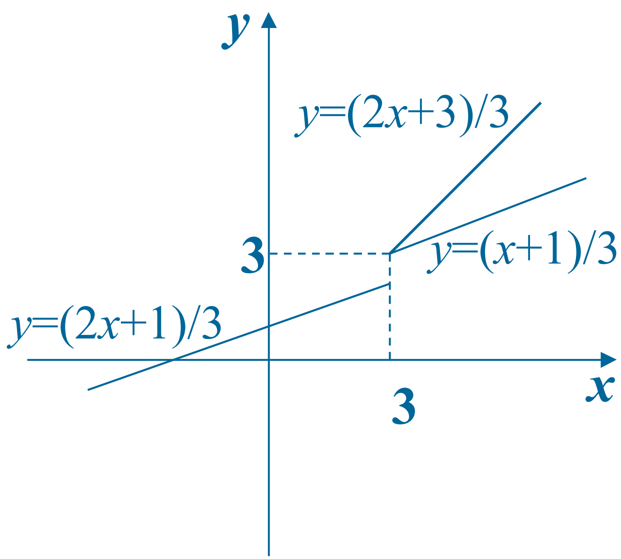
定义算法映射：

$$A(x) = \begin{cases} \left[1, \dfrac{1}{2}(x+1)\right] & x \geq 1; \\ \left[\dfrac{1}{2}(x+1), 1\right] & x < 1. \end{cases}$$

该算法在每一点$x \in R^1$都是闭的。

考虑下列非线性规划：

$$\begin{cases} \min x^2 \\ s.t. \quad x \geq 1. \end{cases}$$



定义算法映射：

$$B(x)=\begin{cases} \left[\frac{1}{2}(x+3), \frac{1}{3}(2x+3)\right] & x \geq 3; \\ \frac{1}{3}(2x+1) & x < 3. \end{cases}$$

取点列$x^{(k)}=3-\frac{1}{k}, y^{(k)}=\frac{7}{3}-\frac{2}{3k} \in B(x^{(k)})$，当$x^{(k)} \to \bar{x}=3$时，$y^{(k)} \to \bar{y}=\frac{7}{3} \notin B(\bar{x})=B(3)$，所以$B(x)$在$\bar{x}=3$处非闭。

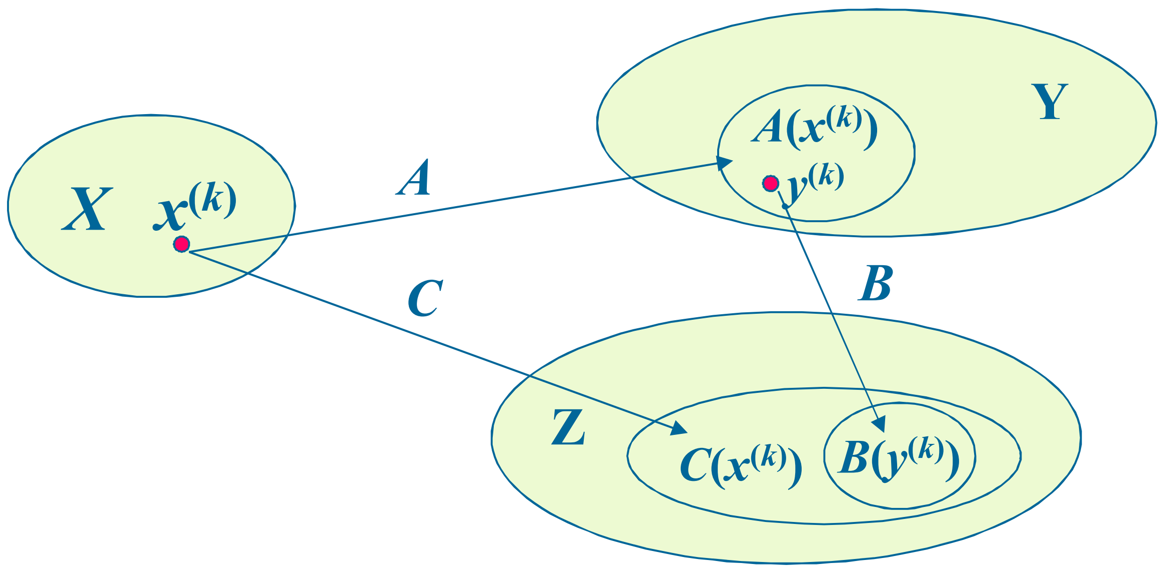
当$B(x)$用于迭代过程时，对于任意初始点$x^{(1)}$，按照$x^{(k+1)} \in B(x^{(k)})$产生的迭代序列收敛于不同的集合：若$x^{(1)} \in (-\infty, 3)$，则算法产生的点列收敛于1；若$x^{(1)} \in [3, \infty)$，则算法产生的点列收敛于3。

合成映射(composition mapping)

定义： 设X，Y和Z分别是空间E^n, E^p和E^q中的非空闭集，集值映射$A: X \to 2^Y$和$B: Y \to 2^Z$。若集值映射$C: X \to 2^Z$定义如下：

$$\forall x \in X, \quad C(x) = \bigcup_{y \in A(x)} B(y),$$

则称C是A和B的合成映射，记为$C = BA$。



定理： 给定集值映射$A: X \to 2^Y, B: Y \to 2^Z$，假设$A$在点$x \in X$是闭的，$B$在$A(x)$上是闭的。对于$x^{(k)} \to x$，若$y^{(k)} \in A(x^{(k)})$，并且$\{y^{(k)}\}$存在收敛子列$\{y^{(k_j)}\}$，则合成映射$C = BA$在点$x$是闭的。

证明： 给定$x^{(k)} \in X, z^{(k)} \in C(x^{(k)})$, 假设$x^{(k)} \to x, z^{(k)} \to z$.

由合成映射定义，对于$k = 1, 2, \cdots$, 存在$y^{(k)} \in A(x^{(k)})$,

使得$z^{(k)} \in B(y^{(k)})$。

$\because \{y^{(k)}\}$存在收敛子列$\{y^{(k_j)}\}$，记$\lim\limits_{j \to +\infty} y^{(k_j)} = y$.

根据映射A的闭性，有$y \in A(x)$.

根据映射B的闭性, 有$\lim\limits_{j \to +\infty} z^{(k_j)} = z \in B(y)$.

$\because B(y) \subseteq C(x), \therefore z \in C(x) \Rightarrow C$在点$x$是闭的。

推论 1 给定集值映射$A: X \to 2^Y, B: Y \to 2^Z$，假设$A$在点$x \in X$是闭的，$B$在$A(x)$上是闭的。若$Y$是紧集，则合成映射$C = BA$在点$x$是闭的。

推论 2 给定映射$A: X \to Y \left(A\text{是点到点的映射} \right), B: Y \to 2^Z$，假设$A$在点$x \in X$连续，$B$在$A(x)$上是闭的，则合成映射$C = BA$在点$x$是闭的。

算法收敛问题

定义： 设Ω为解集合，$A: X \to 2^X$为算法映射。给定一个集合$Y \subset X$，若对于任意的初始点$x^{(1)} \in Y$，算法映射A所产生的序列$\{x^{(k)}\}$中任一收敛子序列的极限都属于Ω，则称算法映射A在Y上收敛。

若集合Y是任意选取的（该集合不必限定在解集合Ω的很小领域内），则相应的收敛性称为全局收敛性(*global convergence*).若集合Y只能取接近Ω的点集，则相应的收敛性称为局部收敛性(*local convergence*).

全局收敛定理

定理： 设A为X上的一个算法，Ω为解集合，给定初点$x^{(1)} \in X$，进行如下迭代：

若$x^{(k)} \in \Omega$，则停止迭代；否则，令$x^{(k+1)} \in A(x^{(k)})$，用$k+1$代替k，重复该过程。

如果下面的条件成立：

1．序列$\{x^{(k)}\}$含于X的某个紧子集中；

2．存在一个连续函数α，它是关于Ω和A的下降函数；

3．映射A在Ω的补集上是闭的。

则序列$\{x^{(k)}\}$的任一收敛子序列的极限属于Ω。

证明： 首先证明序列$\left\{\alpha(x^{(k)})\right\}$有极限。

$\because$ 序列$\{x^{(k)}\}$包含在某个紧子集内，所以存在收敛子列$\left\{x^{(k_j)}\right\}$,记其极限为x，由于$\alpha(x)$是连续的，

$$\therefore \lim_{j\to+\infty} \alpha(x^{(k_j)}) = \alpha(x).$$

$\because \alpha(x)$是下降函数，

$$\therefore \lim_{k\to+\infty} \alpha(x^{(k)}) = \alpha(x).$$

利用反证法证明$x \in \Omega$.

假设某个收敛子序列$\left\{x^{(k_j)}\right\}$的极限$x \notin \Omega$,

则考虑序列$\left\{x^{(k_j+1)}\right\}$，它也存在收敛的子序列，

不妨设 $\lim\limits_{j\to+\infty} x^{(k_j+1)} = \bar{x}$,则$\bar{x} \in X$.

$\because \alpha(x)$连续，$\therefore \lim\limits_{j\to+\infty} \alpha(x^{(k_j+1)}) = \alpha(\bar{x})$,

根据极限的唯一性，有$\alpha(\bar{x}) = \alpha(x)$。

$\because x^{(k_j+1)} \in A(x^{(k_j)}), x^{(k_j)} \to x, x^{(k_j+1)} \to \bar{x}$

由于算法A在Ω的补集上是闭的，$x \notin \Omega$

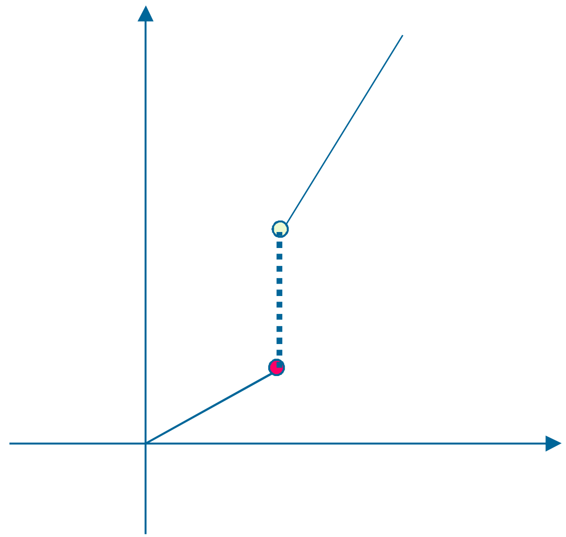
$\therefore A$在x处是闭的 $\quad\Rightarrow\quad \bar{x} \in A(x)$

由于α是关于Ω和A的下降函数，$x \notin \Omega$,

$\therefore \quad \alpha(\bar{x}) < \alpha(x)$矛盾。

考虑如下算法

$$A(x)=\begin{cases}\frac{1}{2}(x-1)+1 & x>1\\ \frac{1}{2}x & x\le 1\end{cases}$$



解集合$\Omega=\{0\}$

$\alpha(x)=|x|$是关于解集合和$A(x)$的下降函数

从$x>1$开始，算法产生一个收敛于$1\notin\Omega$的序列

原因：*A*在解集合外面不是闭的。

实用收敛准则

1. $\left\|x^{(k+1)}-x^{(k)}\right\|<\varepsilon$或者$\dfrac{\left\|x^{(k+1)}-x^{(k)}\right\|}{\left\|x^{(k)}\right\|}<\varepsilon.$

2. $f(x^{(k)})-f(x^{(k+1)})<\varepsilon$或者$\dfrac{f(x^{(k)})-f(x^{(k+1)})}{\left|f(x^{(k)})\right|}<\varepsilon.$

3. $\left\|\nabla f(x^{(k)})\right\|<\varepsilon$(无约束最优化中).

收敛速率

定义： 设序列$\{\gamma^{(k)}\}$收敛于$\gamma*$，定义满足

$$\overline{\lim_{k\to+\infty}}\frac{\left\|\gamma^{(k+1)}-\gamma*\right\|}{\left\|\gamma^{(k)}-\gamma*\right\|^{p}}=\beta<\infty$$

的非负数p的上确界为序列$\{\gamma^{(k)}\}$的收敛级。

若序列的收敛级为p，则称序列是p级收敛的。

若$p=1$且$\beta<1$，则称序列是以收敛比β线性收敛的。

若$p>1$,或者$p=1$且$\beta=0$，则称序列是超线性收敛的。

收敛级p越大，序列收敛得越快；当收敛级p相同时，收敛比β越小，序列收敛得越快。

例： $\{a^k\}$ $0<a<1$

$\because \lim\limits_{k\to\infty} a^k = 0$

又 $\lim\limits_{k\to\infty}\dfrac{a^{k+1}}{a^k} = a < 1$, 且 $\lim\limits_{k\to\infty}\dfrac{a^{k+1}}{(a^k)^r} = \infty$(当 $r>1$ 时),

$\therefore \{a^k\}$ 以收敛比 a 线性收敛于0。

例： $\{a^{2^k}\}$ $0<|a|<1$

$\because \lim\limits_{k\to\infty} a^{2^k} = 0$

又 $\lim\limits_{k\to\infty}\dfrac{a^{2^{k+1}}}{\left(a^{2^k}\right)^2} = \lim\limits_{k\to\infty}\dfrac{a^{2^{k+1}}}{a^{2^{k+1}}} = 1$, 且 $\lim\limits_{k\to\infty}\dfrac{a^{2^{k+1}}}{\left(a^{2^k}\right)^r} = \infty$(当 $r>2$ 时),

$\therefore \{a^{2^k}\}$ 是2级收敛的。

例： $\left\{\left(\frac{1}{k}\right)^k\right\}$

$$\because \lim_{k\to\infty}\left(\frac{1}{k}\right)^k = 0$$

$$\text{又} \lim_{k\to\infty}\frac{\left(\frac{1}{k+1}\right)^{k+1}}{\left(\frac{1}{k}\right)^k} = \lim_{k\to\infty}\left(\frac{k}{k+1}\right)^k \times \frac{1}{k+1} = 0$$

$$\lim_{k\to\infty}\frac{\left(\frac{1}{k+1}\right)^{k+1}}{\left(\left(\frac{1}{k}\right)^k\right)^p} = \lim_{k\to\infty}\left(\frac{k}{k+1}\right)^{k+1} \times \frac{k^{(p-1)k}}{k} = \infty (p > 1)$$

$\therefore \left\{\left(\frac{1}{k}\right)^k\right\}$是超线性收敛的。

算法的二次终止性

- 定义：若某个算法对于任意的正定二次函数，从任意的初始点出发，都能经有限步迭代达到其极小点，则称该算法具有二次终止性。

- 用二次终止性作为判断算法优劣的原因：
- **(1)**正定二次函数具有某些较好的性质，因此一个好的算法应能够在有限步内达到其极小点。
- **(2)**对于一个一般的目标函数，若在其极小点处的**Hesse**矩阵正定，由于

$$f(x)=f\left(x^*\right)+\nabla f\left(x^*\right)^T\left(x-x^*\right)$$

$$+\frac{1}{2}\left(x-x^*\right)^T\nabla^2 f\left(x^*\right)\left(x-x^*\right)+o\left(\|x-x^*\|\right)$$

- 因此可以猜想，对正定二次函数好的算法，对于一般目标函数也应具有较好的性质。