

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

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

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:

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

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。

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

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

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

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 $A: X \rightarrow 2^X$ 为一个算法映射 (algorithm mapping).

例: 考虑标准形式的线性规划

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

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

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

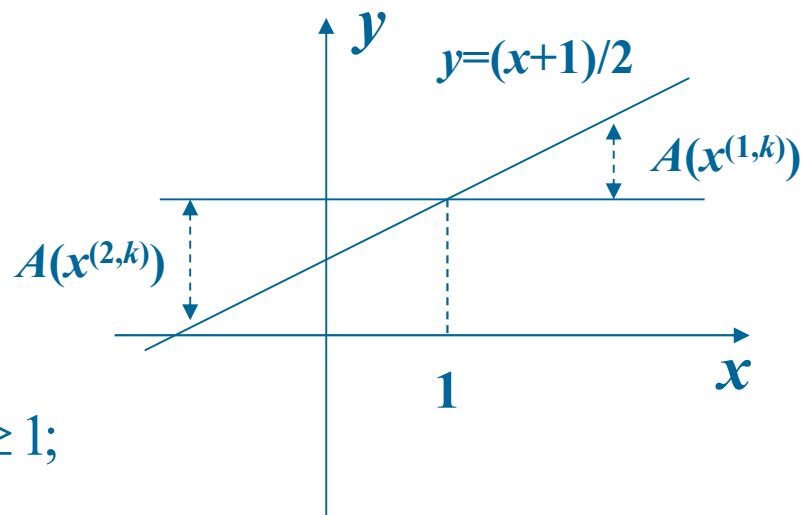
那么对于任意一个基本可行解 $x^{(0)} \in X$, 迭代格式 $x^{(k+1)} \in A(x^{(k)})$ 就生成一个相邻的基本可行解序列。

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：

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

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：

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

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

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

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， $\alpha(y) \leq \alpha(x)$

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**)

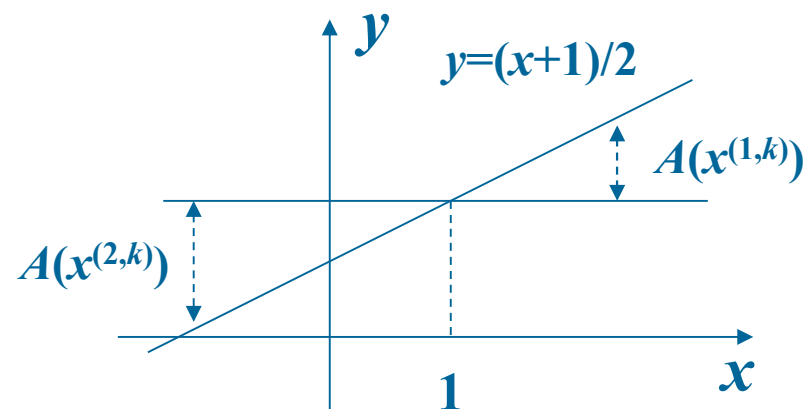
定义:给定两个非空闭集 $X \subset E^m$ 和 $Y \subset E^n$, 以及一个集值映射 $A: X \rightarrow 2^Y$, 假设点列 $\{x^{(k)}\}$ 和 $\{y^{(k)}\}$ 满足 $x^{(k)} \in X$, $y^{(k)} \in A(x^{(k)})$ 。若 $x^{(k)} \rightarrow x$, $y^{(k)} \rightarrow y (k \rightarrow +\infty)$ 蕴涵着 $y \in A(x)$, 则称映射 A 在点 $x \in X$ 处是闭的。若映射 A 在集合 $S \subseteq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

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



该算法在每一点 $x \in R^1$ 都是闭的。

考虑下列非线性规划：

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

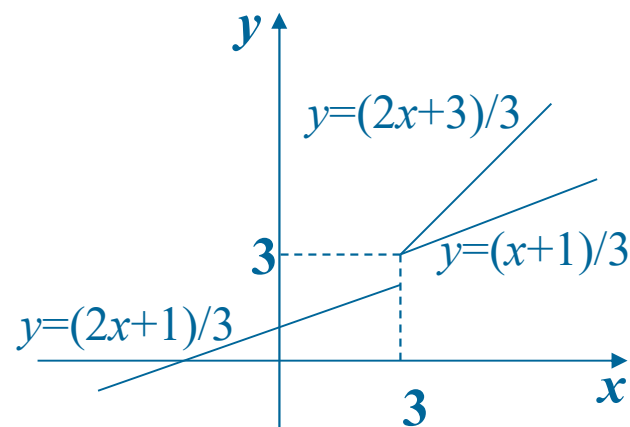
定义算法映射：

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

取点列 $x^{(k)} = 3 - \frac{1}{k}, y^{(k)} = \frac{7}{3} - \frac{2}{3k} \in B(x^{(k)})$, 当 $x^{(k)} \rightarrow \bar{x} = 3$

时, $y^{(k)} \rightarrow \bar{y} = \frac{7}{3} \notin B(\bar{x}) = B(3)$, 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。

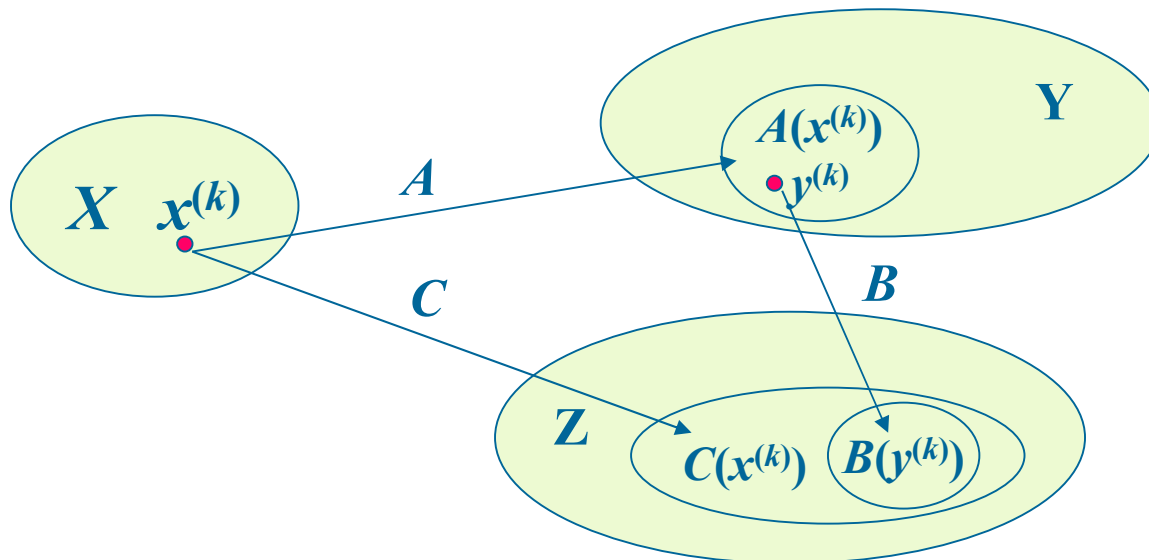


合成映射(**composition mapping**)

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 $A: X \rightarrow 2^Y$ 和 $B: Y \rightarrow 2^Z$ 。若集值映射 $C: X \rightarrow 2^Z$ 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 $A: X \rightarrow 2^Y, B: Y \rightarrow 2^Z$ ，假设 A 在点 $x \in X$ 是闭的， B 在 $A(x)$ 上是闭的。对于 $x^{(k)} \rightarrow x$ ，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 $x^{(k)} \rightarrow x, z^{(k)} \rightarrow z$ 。

由合成映射定义，对于 $k = 1, 2, \dots$ ，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 $\lim_{j \rightarrow +\infty} y^{(k_j)} = y$ 。

根据映射 A 的闭性，有 $y \in A(x)$ 。

根据映射 B 的闭性，有 $\lim_{j \rightarrow +\infty} z^{(k_j)} = z \in B(y)$ 。

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

推论 1 给定集值映射 $A: X \rightarrow 2^Y, B: Y \rightarrow 2^Z$, 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 $A: X \rightarrow Y$ (A 是点到点的映射), $B: Y \rightarrow 2^Z$, 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， $A: X \rightarrow 2^X$ 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 $\{\alpha(x^{(k)})\}$ 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 $\{x^{(k_j)}\}$ ，记其极限为 x ，由于 $\alpha(x)$ 是连续的，

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

$\because \alpha(x)$ 是下降函数，

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

利用反证法证明 $x \in \Omega$.

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

则考虑序列 $\{x^{(k_j+1)}\}$, 它也存在收敛的子序列,

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

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

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

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

由于算法 A 在 Ω 的补集上是闭的, $x \notin \Omega$

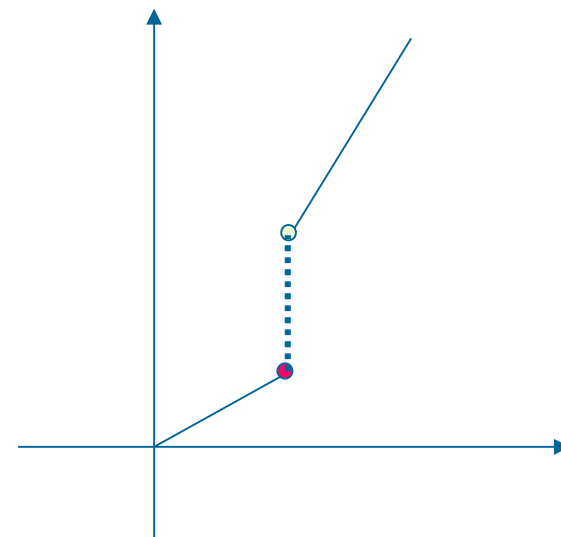
$\therefore A$ 在 x 处是闭的 $\Rightarrow \bar{x} \in A(x)$

由于 α 是关于 Ω 和 A 的下降函数, $x \notin \Omega$,

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

考虑如下算法

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



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. $\|x^{(k+1)} - x^{(k)}\| < \varepsilon$ 或者 $\frac{\|x^{(k+1)} - x^{(k)}\|}{\|x^{(k)}\|} < \varepsilon$.
2. $f(x^{(k)}) - f(x^{(k+1)}) < \varepsilon$ 或者 $\frac{f(x^{(k)}) - f(x^{(k+1)})}{|f(x^{(k)})|} < \varepsilon$.
3. $\|\nabla f(x^{(k)})\| < \varepsilon$ (无约束最优化中).

若取 $\varepsilon = 0$, 算法会产生无穷点列. 实际计算时应避免. 一般取 $\varepsilon = 10^{-4}, 10^{-6}, 10^{-8}$. 理论分析时, 取 $\varepsilon = 0$.

收敛速率

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

$$\overline{\lim}_{k \rightarrow +\infty} \frac{\|\gamma^{(k+1)} - \gamma^*\|}{\|\gamma^{(k)} - \gamma^*\|^p} = \beta < \infty$$

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: $\{a^k\} \quad 0 < a < 1$

$$\because \lim_{k \rightarrow \infty} a^k = 0$$

$$\text{又 } \lim_{k \rightarrow \infty} \frac{a^{k+1}}{a^k} = a < 1, \text{ 且 } \lim_{k \rightarrow \infty} \frac{a^{k+1}}{(a^k)^r} = \infty (\text{当 } r > 1 \text{ 时}),$$

$\therefore \{a^k\}$ 以收敛比 a 线性收敛于 0。

例: $\{a^{2^k}\} \quad 0 < |a| < 1$

$$\because \lim_{k \rightarrow \infty} a^{2^k} = 0$$

$$\text{又 } \lim_{k \rightarrow \infty} \frac{a^{2^{k+1}}}{(a^{2^k})^2} = \lim_{k \rightarrow \infty} \frac{a^{2^{k+1}}}{a^{2^{k+1}}} = 1, \text{ 且 } \lim_{k \rightarrow \infty} \frac{a^{2^{k+1}}}{(a^{2^k})^r} = \infty (\text{当 } r > 2 \text{ 时}),$$

$\therefore \{a^{2^k}\}$ 是 2 级收敛的。

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

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

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

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

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

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

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

- 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。