

非线性规划

$$\begin{array}{ll}\min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in S \subset \mathbb{R}^n.\end{array}\tag{51}$$

在此，目标函数 f 是定义在 \mathbb{R}^n 上的实值函数， S 是决策变量 \mathbf{x} 的可取值之集合，称为问题的可行域(feasible region).

非线性规划

最优化问题从属性上可以分为两大类：一类是具有连续变量的问题，另一类是离散变量的问题（即组合优化问题）。

非线性规划属于连续型最优化问题的范畴，通常可行域 S 可由一组方程来描述，即

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \geq 0, i = 1, \dots, m; h_j(\mathbf{x}) = 0, j = 1, \dots, \ell\}.$$

$$\begin{array}{ll}\min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \geq 0, i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0, j = 1, \dots, \ell\end{array} \quad (52)$$

这里， $f(\mathbf{x})$, $g_i(\mathbf{x})$, $h_j(\mathbf{x})$ 都是 n 变量、实值、确定的函数，且至少有一个是非线性的。

为求解一个非线性规划问题（即找出其最优解），与此相关的研究分两个方面：一是研究最优解的性质，二是设计有效算法来获得问题的解。

回顾最优解的定义

满足约束条件 $\mathbf{x} \in S$ 的 \mathbf{x} 称为问题的可行解(feasible solution), 如果可行解 $\mathbf{x}^* \in S$ 进一步满足

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in S. \quad (53)$$

则称 \mathbf{x}^* 为问题(51)的全局最优解(global optimal solution). 另外, 在包含可行解 $\mathbf{x}^* \in S$ 的适当邻域 $U(\mathbf{x}^*)$ 里, 成立

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in S \cap U(\mathbf{x}^*). \quad (54)$$

此时称 \mathbf{x}^* 为问题(51)的局部最优解(local optimal solution).

最优性条件

最优性条件：问题的最优解所满足的必要或者充分条件。

最优性条件将为各种求解算法的设计、分析提供必不可少的理论基础。

无约束问题的极值条件

一阶必要条件：设目标函数 $f(\mathbf{x})$ 在点 $\bar{\mathbf{x}}$ 处可微，若 $\bar{\mathbf{x}}$ 是局部极小点，则 $\nabla f(\bar{\mathbf{x}}) = 0$.

二阶必要条件：设目标函数 $f(\mathbf{x})$ 在点 $\bar{\mathbf{x}}$ 处二次可微，若 $\bar{\mathbf{x}}$ 是局部极小点，则 $\nabla f(\bar{\mathbf{x}}) = 0$ ，并且Hesse矩阵 $\nabla^2 f(\bar{\mathbf{x}}) \geq 0$.

二阶充分条件：设目标函数 $f(\mathbf{x})$ 在点 $\bar{\mathbf{x}}$ 处二次可微，若 $\nabla f(\bar{\mathbf{x}}) = 0$ 且 $\nabla^2 f(\bar{\mathbf{x}}) > 0$ ，则 $\bar{\mathbf{x}}$ 是局部极小点。

充要条件：设 $f(\mathbf{x})$ 是定义在 \mathbb{R}^n 上的可微凸函数，则 $\bar{\mathbf{x}}$ 是整体极小点（全局最优解）的充要条件是 $\nabla f(\bar{\mathbf{x}}) = 0$.

约束问题的最优性条件

可行方向

设 $\bar{\mathbf{x}} \in S$, $\mathbf{d} \in \mathbb{R}^n$ 是非零向量。若存在 $\delta > 0$ 使得:

$$\bar{\mathbf{x}} + \lambda \mathbf{d} \in S, \forall \lambda \in (0, \delta) \quad (55)$$

则称 \mathbf{d} 是 S 在 $\bar{\mathbf{x}}$ 处的可行方向。

记 S 在 $\bar{\mathbf{x}}$ 处的所有可行方向的集合为 $F(\bar{\mathbf{x}}, S)$.

约束问题的最优性条件

下降方向

设 $f(\mathbf{x})$ 是 \mathbb{R}^n 上的实函数, $\bar{\mathbf{x}} \in \mathbb{R}^n$, \mathbf{d} 是非零向量。若存在 $\delta > 0$ 使得:

$$f(\bar{\mathbf{x}} + \lambda \mathbf{d}) < f(\bar{\mathbf{x}}), \forall \lambda \in (0, \delta) \quad (56)$$

则称 \mathbf{d} 为函数 $f(\mathbf{x})$ 在 $\bar{\mathbf{x}}$ 处的下降方向。

下降方向集的子集

如果 $f(\mathbf{x})$ 是可微函数, 且 $\nabla f(\bar{\mathbf{x}})^T \mathbf{d} < 0$. 显然, 此处的 \mathbf{d} 为 $f(\mathbf{x})$ 在 $\bar{\mathbf{x}}$ 处的下降方向. 记这样的方向集合为

$$D(\bar{\mathbf{x}}, f) = \{\mathbf{d} \mid \nabla f(\bar{\mathbf{x}})^T \mathbf{d} < 0\}.$$

约束问题的最优性条件

定理

对于问题 $\min\{f(\mathbf{x}) \mid \mathbf{x} \in S\}$, 设 $\bar{\mathbf{x}} \in S$, $f(\mathbf{x})$ 在 $\bar{\mathbf{x}}$ 处可微。如果 $\bar{\mathbf{x}}$ 是问题的局部最优解, 则可行方向集中无下降方向, 即

$$F(\bar{\mathbf{x}}, S) \cap D(\bar{\mathbf{x}}, f) = \emptyset. \quad (57)$$

然而, 对于一些等式, 例如 $x_1^2 + x_2^2 = 1$, 可行方向不存在。上述定理对于多面体集可以适用。对于一般的非线性等式约束确定的可行域 C , 定义在点 $\mathbf{x} \in C$ 处的切锥为从在 C 中收敛到 \mathbf{x} 的序列获得的极限方向的集合。

定义 (切锥)

$$T(\mathbf{x} \mid S) := \left\{ \mathbf{d} : \exists \tau_i \searrow 0, \{\mathbf{x}_i\} \subset S, \mathbf{x}_i \rightarrow \mathbf{x}, \text{s.t. } \frac{\mathbf{x}_i - \mathbf{x}}{\tau_i} \rightarrow \mathbf{d} \right\}.$$

定理

对于问题 $\min\{f(\mathbf{x}) \mid \mathbf{x} \in S\}$, 设 $\bar{\mathbf{x}} \in S$, $f(\mathbf{x})$ 在 $\bar{\mathbf{x}}$ 处可微。如果 $\bar{\mathbf{x}}$ 是问题的局部最优解, 则

$$\nabla f(\bar{\mathbf{x}})^T d \geq 0, \forall d \in T(\mathbf{x} \mid S). \quad (58)$$

Proof.

若 $d \in T(\bar{\mathbf{x}} \mid S)$, 则存在序列 $\{x_k\} \subset S$, $\tau_k \searrow 0$, 且 $x_k \rightarrow x$ 使得 $\frac{x_k - \bar{x}}{\tau_k} \rightarrow d$, 令 $x_k = \bar{x} + \tau_k d_k$.

$$\nabla f(\bar{\mathbf{x}})^T d = \lim_{k \rightarrow \infty} \nabla f(\bar{\mathbf{x}})^T d_k = \lim_{k \rightarrow \infty} \frac{f(\bar{\mathbf{x}}) + \tau_k \nabla f(\bar{\mathbf{x}})^T d_k + o(\tau_k) - f(\bar{\mathbf{x}})}{\tau_k} = \lim_{k \rightarrow \infty} \frac{f(x_k) - f(\bar{\mathbf{x}})}{\tau_k} \geq 0.$$



推论: 局部最优点 $\bar{\mathbf{x}}$ 满足

$$T(\bar{\mathbf{x}} \mid S) \cap D(\bar{\mathbf{x}}, f) = \emptyset. \quad (59)$$

对于等式约束 $\mathcal{E} := \{h_i(\bar{\mathbf{x}}) = 0, i = 1, \dots, \ell\}$, 若 $d \in T(\bar{\mathbf{x}} \mid h)$, 这里 $T(\bar{\mathbf{x}} \mid h)$ 表示 $h_i(\bar{\mathbf{x}}) = 0, i = 1, \dots, \ell$ 的切锥, 则存在序列 $\{x_k\} \subset \mathcal{E}$, $\tau_k \searrow 0$, 且 $x_k \rightarrow \bar{\mathbf{x}}$ 使得 $\frac{x_k - \bar{\mathbf{x}}}{\tau_k} \rightarrow d$. 令 $d_k = \frac{x_k - \bar{\mathbf{x}}}{\tau_k}$, 则

$$\nabla h_i(\bar{\mathbf{x}})^T d = h_i(\bar{\mathbf{x}}; d) = \lim_{k \rightarrow \infty} \frac{h_i(\bar{\mathbf{x}} + \tau_k d_k) - h_i(\bar{\mathbf{x}})}{\tau_k} = 0. \quad (60)$$

同理, 对于不等式约束积极集 (active set), 记 $\mathcal{I}(\bar{\mathbf{x}}) = \{i \in \{1, \dots, m\} \mid g_i(\bar{\mathbf{x}}) = 0\}$,

$$\nabla g_i(\bar{\mathbf{x}})^T d = g_i(\bar{\mathbf{x}}; d) = \lim_{k \rightarrow \infty} \frac{g_i(\bar{\mathbf{x}} + \tau_k d_k) - g_i(\bar{\mathbf{x}})}{\tau_k} \geq 0, i \in \mathcal{I}(\bar{\mathbf{x}}). \quad (61)$$

约束问题的最优性条件

$$D_f = D(\bar{\mathbf{x}}, f) = \{\mathbf{d} \mid \nabla f(\bar{\mathbf{x}})^T \mathbf{d} < 0\}, \quad (62)$$

定义:

$$F_g = \{\mathbf{d} \mid \nabla g_i(\bar{\mathbf{x}})^T \mathbf{d} > 0, i \in \mathcal{I}(\bar{\mathbf{x}})\}, \quad (F_g \text{ 是不等式可行方向的子集})$$

我们有

$$F_g \subset F(\bar{\mathbf{x}}, g) \subset T(\bar{\mathbf{x}} \mid g) \subset L_g := \{\mathbf{d} \mid \nabla g_i(\bar{\mathbf{x}})^T \mathbf{d} \geq 0, i \in \mathcal{I}(\bar{\mathbf{x}})\},$$

(L_g 被称为不等式的线性可行方向)

最后一步由(60)得到。

$$F(\bar{\mathbf{x}}, h) \subset T(\bar{\mathbf{x}} \mid h) \subset L_h := \{\mathbf{d} \mid \nabla h_j(\bar{\mathbf{x}})^T \mathbf{d} = 0, j = 1, \dots, \ell\}.$$

(L_h 被称为等式线性可行方向)

最后一步由(61)得到。

由(63)可知,

$$T(\bar{\mathbf{x}} \mid g) \cap T(\bar{\mathbf{x}} \mid h) \cap D(\bar{\mathbf{x}}, f) = \emptyset. \quad (63)$$

约束问题的最优性条件

线性可行方向与约束的表示形式有关。

例

令 $S = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1, (x_1 - 2)^2 + x_2^2 \leq 1\}$. 则 $S = \{(1, 0)^T\}$. 我们有 $T(S) = \{(0, 0)^T\}$, 然而 $L_S = \{(0, t) \mid t \in \mathbb{R}\}$. 但是, 如果我们直接考虑约束的一种直接表示: $S = \{(1, 0)^T\}$, 那么 $T(S) = L_S$.

以下条件称为 Mangasarian-Fromovitz 约束规范条件 (MFCQ):

定义 (MFCQ)

若 h_j 在点 \bar{x} 连续可微, 且 $\{\nabla h_j(\bar{x})\}_{j=1}^{\ell}$ 线性无关, 并且 $F_g \cap L_h$ 非空, 则称该点处满足 MFCQ 约束规范条件。

事实上, 我们有 $\text{LICQ} \implies \text{MFCQ}$.

引理 (规范性引理)

若在可行点 $\bar{x} \in S$ 处满足 MFCQ, 则 $L_h \cap L_g = T(\bar{x} \mid S)$.

约束问题的最优性条件

引理

设 \bar{x} 为问题(52)的局部最优解, f 和 $g_i, i \in \mathcal{I}(\bar{x})$ 在点 \bar{x} 可微, $g_i, i \notin \mathcal{I}(\bar{x})$ 在点 \bar{x} 连续, h_j 在点 \bar{x} 连续可微, 且 $\{\nabla h_j(\bar{x})\}_{j=1}^{\ell}$ 线性无关, 则

$$D_f \cap F_g \cap L_h = \emptyset. \quad (64)$$

定理

*Fritz-John*条件: 在问题(52)中, 设 \bar{x} 为可行点, f 和 $g_i, i \in \mathcal{I}(\bar{x})$ 在点 \bar{x} 可微, $g_i, i \notin \mathcal{I}(\bar{x})$ 在点 \bar{x} 连续, h_j 在点 \bar{x} 连续可微。如果 \bar{x} 是局部最优解, 则存在不全为零的数 $\lambda_0, \lambda_i, i \in \mathcal{I}(\bar{x})$ 和 $\mu_j, j = 1, \dots, \ell$ 使得

$$\lambda_0 \nabla f(\bar{x}) - \sum_{i \in \mathcal{I}(\bar{x})} \lambda_i \nabla g_i(\bar{x}) - \sum_{j=1}^{\ell} \mu_j \nabla h_j(\bar{x}) = 0. \quad (65)$$

其中 $\lambda_0 \geq 0, \lambda_i \geq 0, i \in \mathcal{I}(\bar{x})$.

约束问题的最优性条件

证明:(1) 如果 $\{\nabla h_j(\bar{\mathbf{x}})\}_{j=1}^{\ell}$ 线性相关, 则存在不全为零的数 $\mu_j, j = 1, \dots, \ell$ 使得

$$\sum_{j=1}^{\ell} \mu_j \nabla h_j(\bar{\mathbf{x}}) = 0.$$

这时可令 $\lambda_0 = 0, \lambda_i = 0, i \in \mathcal{I}(\bar{\mathbf{x}})$, 结论成立。

约束问题的最优性条件

证明（续）：(2)如果 $\{\nabla h_j(\bar{\mathbf{x}})\}_{j=1}^{\ell}$ 线性无关。利用 $D_f \cap F_g \cap L_h = \emptyset$.
即不等式组

$$\begin{cases} \nabla f(\bar{\mathbf{x}})^T \mathbf{d} < 0 \\ \nabla g_i(\bar{\mathbf{x}})^T \mathbf{d} > 0, i \in \mathcal{I}(\bar{\mathbf{x}}) \\ \nabla h_j(\bar{\mathbf{x}})^T \mathbf{d} = 0, j = 1, \dots, \ell \end{cases} \quad (66)$$

无解。

约束问题的最优性条件

证明(续): 令 A 是以 $\{\nabla f(\bar{\mathbf{x}}), -\nabla g_i(\bar{\mathbf{x}}), i \in \mathcal{I}(\bar{\mathbf{x}})\}$ 为列组成的矩阵, B 是以 $\{-\nabla h_j(\bar{\mathbf{x}}), j = 1, \dots, \ell\}$ 为列组成的矩阵。
于是得

$$\begin{cases} A^T \mathbf{d} < 0 \\ B^T \mathbf{d} = 0 \end{cases} \quad (67)$$

无解。

下证

$$\begin{cases} A\mathbf{p}_1 + B\mathbf{p}_2 = 0 \\ \mathbf{p}_1 \geq 0 \end{cases} \quad (68)$$

有解。

约束问题的最优性条件

定理 (凸集分离定理)

假设 C_1 和 C_2 是两个不相交的非空凸集, 那么存在一个非零向量 w 和一个实数 b 使得对于所有 $x_1 \in \text{CL}(C_1)$ 和 $x_2 \in \text{CL}(C_2)$ 有:

$$w^T x_1 \geq b \quad \text{和} \quad w^T x_2 \leq b,$$

这里 $\text{CL}(C_1)$ 表示 C_1 的闭包。这意味着超平面 $\{x : w^T x = b\}$ 将 $\text{CL}(C_1)$ 和 $\text{CL}(C_2)$ 分开。

证明(续): 现定义

$$S_1 = \left\{ \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \mid \mathbf{y}_1 = A^T \mathbf{d}, \mathbf{y}_2 = B^T \mathbf{d}, \mathbf{d} \in \mathbb{R}^n \right\},$$

$$S_2 = \left\{ \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \mid \mathbf{y}_1 < \mathbf{0}, \mathbf{y}_2 = \mathbf{0} \right\}.$$

显然 S_1 和 S_2 为非空凸集, 且 $S_1 \cap S_2 = \emptyset$.

约束问题的最优性条件

证明(续): 由凸集分离定理知, 对 $\forall \mathbf{d} \in \mathbb{R}^n, \forall \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \in S_2$, 存在非零向量 $\begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{pmatrix}$ 使得 $\mathbf{p}_1^T A^T \mathbf{d} + \mathbf{p}_2^T B^T \mathbf{d} \geq \mathbf{p}_1^T \mathbf{y}_1 + \mathbf{p}_2^T \mathbf{y}_2$.

首先令 $\mathbf{y}_2 = \mathbf{0}$, 由 \mathbf{d} 的任意性 (取 $\mathbf{d} = \mathbf{0}$) 及 $\mathbf{y}_1 < \mathbf{0}, \implies \mathbf{p}_1 \geq \mathbf{0}$.

再令 $\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \in \text{CL}(S_2), \implies \mathbf{p}_1^T A^T \mathbf{d} + \mathbf{p}_2^T B^T \mathbf{d} \geq 0$.

最后取 $\mathbf{d} = -(\mathbf{A}\mathbf{p}_1 + \mathbf{B}\mathbf{p}_2), \implies \mathbf{A}\mathbf{p}_1 + \mathbf{B}\mathbf{p}_2 = \mathbf{0}$.

综上所述, 即得(68)有解。

约束问题的最优性条件

证明(续):

把 \mathbf{p}_1 的分量记作 λ_0 和 $\lambda_i, i \in \mathcal{I}(\bar{\mathbf{x}})$, \mathbf{p}_2 的分量记作 $\mu_j, j = 1, \dots, \ell$. 立即得到

$$\lambda_0 \nabla f(\bar{\mathbf{x}}) - \sum_{i \in I(\bar{\mathbf{x}})} \lambda_i \nabla g_i(\bar{\mathbf{x}}) - \sum_{j=1}^{\ell} \mu_j \nabla h_j(\bar{\mathbf{x}}) = 0. \quad (69)$$

约束问题的最优性条件

KKT条件：设 \bar{x} 为约束问题(52)的可行点， f 和 $g_i, i \in \mathcal{I}(\bar{x})$ 在点 \bar{x} 可微， $g_i, i \notin \mathcal{I}(\bar{x})$ 在点 \bar{x} 连续， h_j 在点 \bar{x} 连续可微，向量集 $\{\nabla g_i(\bar{x}), i \in \mathcal{I}(\bar{x}); \nabla h_j(\bar{x}), j = 1, \dots, \ell\}$ 线性无关。如果 \bar{x} 是局部最优解，则存在数 $\lambda_i \geq 0$ 和 μ_j 使得

$$\nabla f(\bar{x}) - \sum_{i \in \mathcal{I}(\bar{x})} \lambda_i \nabla g_i(\bar{x}) - \sum_{j=1}^{\ell} \mu_j \nabla h_j(\bar{x}) = 0. \quad (70)$$

证明：利用Fritz John条件，若LICQ成立，则 $\lambda_0 \neq 0$.两边同除以 λ_0 得证。

约束问题的最优性条件

定义Lagrange函数 $L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) - \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) - \sum_{j=1}^{\ell} \mu_j h_j(\mathbf{x})$.

若 $\bar{\mathbf{x}}$ 为问题局部最优解, 则存在乘子向量 $\bar{\lambda} \geq 0, \bar{\mu}$ 使得

$$\nabla_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mu}) = 0.$$

此时, 一阶必要条件可表达为

$$(KKT) \begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu) = 0 \\ g_i(\mathbf{x}) \geq 0, i = 1, \dots, m \\ \lambda_i g_i(\mathbf{x}) = 0, i = 1, \dots, m \\ \lambda_i \geq 0, i = 1, \dots, m \\ h_j(\mathbf{x}) = 0, j = 1, \dots, \ell \end{cases} \quad (71)$$

理论上可以用最优性条件求“非线性规划问题”的最优解，但在实践中并不切实可行。

在求解最优化问题时最常用的计算方法是“迭代下降算法”。

算法映射：算法 \mathcal{A} 是定义在空间 \mathbf{X} 上的**点到集**的映射，即对每一点 $\mathbf{x}^{(k)} \in \mathbf{X}$ ，经算法 \mathcal{A} 作用后产生一个点集 $\mathcal{A}(\mathbf{x}^{(k)}) \subset \mathbf{X}$ ，任意选择一个点 $\mathbf{x}^{(k+1)} \in \mathcal{A}(\mathbf{x}^{(k)})$ 作为 $\mathbf{x}^{(k)}$ 的后续点。

引入所谓算法的闭性，其实质是点到点映射的连续性的推广。

设 \mathbf{X} 和 \mathbf{Y} 分别是空间 \mathbb{E}^p 和 \mathbb{E}^q 中的非空闭集， $\mathcal{A} : \mathbf{X} \longrightarrow \mathbf{Y}$ 为点到集的映射。如果 $\mathbf{x}^{(k)} \in \mathbf{X}, \mathbf{x}^{(k)} \rightarrow \mathbf{x}, \mathbf{y}^{(k)} \in \mathcal{A}(\mathbf{x}^{(k)}), \mathbf{y}^{(k)} \rightarrow \mathbf{y}$ 蕴涵着 $\mathbf{y} \in \mathcal{A}(\mathbf{x})$ ，则称映射 \mathcal{A} 在 $\mathbf{x} \in \mathbf{X}$ 处是闭的。

为研究算法的收敛性，首先要明确解集合的概念。

在许多情况下，要使算法产生的点列收敛于全局最优解是极为困难的。因此，一般把满足某些条件的点集定义为解集合。当迭代点属于这个集合时，就停止迭代。

例如，在无约束最优化问题中，可以定义解集合为

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

在约束最优化问题中，解集合取为

$$\Omega = \{\bar{\mathbf{x}} \mid \bar{\mathbf{x}} \text{ 是 KKT 点}\}.$$

下降函数

一般地，下降算法总是与某个函数在迭代过程中函数值的减小联系在一起，因此需要给出下降函数的概念。

设 $\Omega \subset \mathbf{X}$ 为解集合， \mathcal{A} 为 \mathbf{X} 上的一个算法映射， $\psi(\mathbf{x})$ 是定义在 \mathbf{X} 上的连续实函数，若满足

$$\text{当 } \mathbf{x} \notin \Omega \text{ 且 } \mathbf{y} \in \mathcal{A}(\mathbf{x}) \text{ 时, } \psi(\mathbf{y}) < \psi(\mathbf{x})$$

$$\text{当 } \mathbf{x} \in \Omega \text{ 且 } \mathbf{y} \in \mathcal{A}(\mathbf{x}) \text{ 时, } \psi(\mathbf{y}) \leq \psi(\mathbf{x})$$

则称 ψ 是关于解集合 Ω 和算法 \mathcal{A} 的下降函数。

算法收敛性

设 Ω 为解集合， \mathcal{A} 为 \mathbf{X} 上的算法映射。若以任意初始点 $\mathbf{x}^{(0)} \in \mathbf{Y} \subset \mathbf{X}$ 出发，算法产生的序列的任一收敛子列的极限属于解集合，则称算法映射 \mathcal{A} 在 \mathbf{Y} 上收敛于解集合 Ω 。

算法收敛性

定理： 设 \mathcal{A} 为 \mathbf{X} 上的一个算法， Ω 为解集合，给定初始点 $\mathbf{x}^{(0)} \in \mathbf{X}$ ，进行如下迭代：如果 $\mathbf{x}^{(k)} \in \Omega$ ，则停止迭代；否则取 $\mathbf{x}^{(k+1)} \in \mathcal{A}(\mathbf{x}^{(k)})$ ， $k := k + 1$ ，重复以上过程。这样产生迭代序列 $\{\mathbf{x}^{(k)}\}$ 。又设：

- ① 序列 $\{\mathbf{x}^{(k)}\}$ 含于 \mathbf{X} 的紧子集中；
- ② 存在一个连续函数 ψ ，它是关于 Ω 和 \mathcal{A} 的下降函数；
- ③ 映射 \mathcal{A} 在 Ω 的补集上是闭的。

则序列 $\{\mathbf{x}^{(k)}\}$ 的任一收敛子列的极限属于 Ω 。

算法收敛性定理证明

证明： 先证序列 $\{\mathbf{x}^{(k)}\}$ 对应的下降函数值数列 $\{\psi(\mathbf{x}^{(k)})\}$ 有极限。

$\{\mathbf{x}^{(k)}\}$ 含于 \mathbf{X} 的紧子集，因此有收敛子列 $\{\mathbf{x}^{(k)}\}_K$ ，设其极限为 $\mathbf{x} \in \mathbf{X}$ 。
由 ψ 的连续性得， $\psi(\mathbf{x}^{(k)}) \rightarrow \psi(\mathbf{x}), k \in K$ 。即对 $\forall \epsilon > 0, \exists N$ 使得
当 $k \geq N$ 时，有 $0 < \psi(\mathbf{x}^{(k)}) - \psi(\mathbf{x}) < \epsilon, k \in K$ 。特别地，
 $0 < \psi(\mathbf{x}^{(N)}) - \psi(\mathbf{x}) < \epsilon$ 。

又由 ψ 的下降性知， $\psi(\mathbf{x}^{(k)}) - \psi(\mathbf{x}^{(N)}) < 0, \forall k > N$ 。于是有

$$0 < \psi(\mathbf{x}^{(k)}) - \psi(\mathbf{x}) = \psi(\mathbf{x}^{(k)}) - \psi(\mathbf{x}^{(N)}) + \psi(\mathbf{x}^{(N)}) - \psi(\mathbf{x}) < \epsilon, \forall k > N.$$

即得 $\lim_{k \rightarrow \infty} \psi(\mathbf{x}^{(k)}) = \psi(\mathbf{x})$ 。

算法收敛性定理证明

证明（继）： 下证 $\mathbf{x} \in \Omega$ （反证法）。

假设 $\mathbf{x} \notin \Omega$, 考虑序列 $\{\mathbf{x}^{(k+1)}\}_K$. 由于它包含于紧集, 所以也存在收敛子列 $\{\mathbf{x}^{(k+1)}\}_{\bar{K}}, \bar{K} \subset K$, 且设其极限为 $\bar{\mathbf{x}} \in \mathbf{X}$. 显然（同理前述证明）

$\lim_{k \rightarrow \infty} \psi(\mathbf{x}^{(k+1)}) = \psi(\bar{\mathbf{x}})$. 由 $\psi(\mathbf{x}^{(k)})$ 极限的唯一性知,

$$\psi(\mathbf{x}) = \psi(\bar{\mathbf{x}}).$$

另外, 对 $k \in \bar{K} \subset K$ 有

$$\mathbf{x}^{(k)} \longrightarrow \mathbf{x}, \mathbf{x}^{(k+1)} \in \mathcal{A}(\mathbf{x}^{(k)}), \mathbf{x}^{(k+1)} \longrightarrow \bar{\mathbf{x}}.$$

算法 \mathcal{A} 在 Ω 的补集上是闭的, $\mathbf{x} \notin \Omega$, 因此 \mathcal{A} 在 \mathbf{x} 处是闭的, 即有 $\bar{\mathbf{x}} \in \mathcal{A}(\mathbf{x})$.

由于 ψ 是解集合 Ω 和算法 \mathcal{A} 的下降函数, $\mathbf{x} \notin \Omega$, 则有 $\psi(\bar{\mathbf{x}}) < \psi(\mathbf{x})$. 这显然矛盾, 所以 $\mathbf{x} \in \Omega$.

实用收敛准则

在迭代下降算法里，当 $\mathbf{x}^{(k)} \in \Omega$ 时才终止迭代。在实践中许多情况下，这是一个取极限的过程，需要无限次迭代。因此为了解决实际问题，需要规定一些实用的终止迭代过程的准则，一般称为收敛准则或停机准则。

常用的收敛准则有以下几种：

- ① $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < \varepsilon$ or $\frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}{\|\mathbf{x}^{(k)}\|} < \varepsilon$
- ② $f(\mathbf{x}^{(k)}) - f(\mathbf{x}^{(k+1)}) < \varepsilon$ or $\frac{f(\mathbf{x}^{(k)}) - f(\mathbf{x}^{(k+1)})}{|f(\mathbf{x}^{(k)})|} < \varepsilon$
- ③ $\|\nabla f(\mathbf{x}^{(k)})\| < \varepsilon$

在这里， ε 为事先给定的充分小的正数。除此之外，还可以根据收敛定理，制定出其它的收敛准则。

收敛速率

评价算法优劣的标准之一是收敛的快慢，通常称为收敛速率。

一般定义如下：设序列 $\{\mathbf{x}^{(k)}\}$ 收敛与 \mathbf{x}^* ，满足

$$0 \leq \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^p} = \beta < \infty \quad (72)$$

的非负数 p 的上确界称为序列 $\{\mathbf{x}^{(k)}\}$ 的收敛阶。

若在定义式(72)中， $p = 1$ 且 $\beta < 1$ ，则称序列是（收敛比 β ）线性收敛的。

若在定义式(72)中， $p > 1$ ，或者 $\{p = 1, \beta = 0\}$ ，则称序列是超线性收敛的。

最优化可以追溯到十分古老的极值问题，然而它成为一门独立的学科是在二十世纪四十年代末，当时Dantzig提出了求解一般线性规划问题的单纯形法。此后各种最优化问题的理论及应用研究得到迅速发展，特别是线性规划由于其模型的普遍性和实用性，相关算法的进展引起广泛的重视。

随着实际问题的规模越来越大以及在计算机技术的推动下，人们开始从复杂性角度研究线性规划和非线性规划的算法。

最优化方法通常采用迭代方法求问题的最优解，其基本思想是：

给定一个初始点 $\mathbf{x}^{(0)} \in \mathbb{R}^n$ ，按照某一迭代规则产生一个点列 $\{\mathbf{x}^{(k)}\}$ ，使得当 $\{\mathbf{x}^{(k)}\}$ 是有穷点列时，其最后一个点是最优化模型问题的最优解，当 $\{\mathbf{x}^{(k)}\}$ 是无穷点列时，它有极限点且其极限点是最优化模型问题的最优解。

一个好的迭代算法应具备的典型特征是：

迭代点 $\mathbf{x}^{(k)}$ 能稳定地接近局部极小点 \mathbf{x}^* 的小邻域，然后迅速收敛于 \mathbf{x}^* 。一般地，对于某种算法我们需要证明其迭代点列 $\mathbf{x}^{(k)}$ 的聚点（即子列的极限点）为一局部极小点。在实际计算中，当指定的收敛准则满足时，迭代即终止。

搜索方向与步长因子

设 $\mathbf{x}^{(k)}$ 为第 k 次迭代点， $\mathbf{d}^{(k)}$ 为第 k 次搜索方向， α_k 为第 k 次步长因子，则第 $k+1$ 次迭代为：

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}. \quad (73)$$

从上述迭代格式可以看出，不同的搜索方向和不同的步长策略构成不同的方法。

搜索方向与步长因子

在最优化方法中，搜索方向 $\mathbf{d}^{(k)}$ 一般选取的是某价值函数 (merit function) ψ 在 $\mathbf{x}^{(k)}$ 处的下降方向，即 $\mathbf{d}^{(k)}$ 满足

$$\nabla\psi(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} < 0. \quad (74)$$

步长因子的确定一般归结为解一维最优化问题

$$\min_{\alpha \geq 0} \psi(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \quad (75)$$

最优化迭代算法的基本结构之一

- (a) 给定初始点 $\mathbf{x}^{(0)}$
- (b) 计算搜索方向 $\mathbf{d}^{(k)}$, 即构造某价值函数 ψ 在 $\mathbf{x}^{(k)}$ 点处的下降方向作为搜索方向;
- (c) 确定步长因子 α_k , 使该价值函数值有某种程度的下降;
- (d) 迭代更新, 令 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$.
- (e) 判断停机准则, 若 $\mathbf{x}^{(k+1)}$ 满足某种终止条件, 则停止迭代, 得到近似最优解 $\bar{\mathbf{x}} = \mathbf{x}^{(k+1)}$. 否则, 返回(b)重复以上步骤。

最优化迭代算法的基本结构之二

- (a) 给定初始点 $\mathbf{x}^{(0)}$
- (b) 构造某价值函数 ψ 在 $\mathbf{x}^{(k)}$ 附近（如一定半径内）的二次近似模型；
- (c) 求解该近似模型得到 $\mathbf{s}^{(k)}$ 作为更新位移向量；
- (d) 迭代更新，令 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{s}^{(k)}$.
- (e) 判断停机准则，若 $\mathbf{x}^{(k+1)}$ 满足某种终止条件，则停止迭代，得到近似最优解 $\bar{\mathbf{x}} = \mathbf{x}^{(k+1)}$. 否则，返回(b)重复以上步骤。

习题1: 对于问题(52), 若 \bar{x} 是局部最优解, 并且 \bar{x} 满足MFCQ条件。证明如下结论:

- ① 对于 $\forall d \in T(\bar{x} \mid S)$, 我们有 $\nabla f(\bar{x})^T d \geq 0$ 。
- ② 考虑下面的线性规划问题:

$$\max(-\nabla f(\bar{x}))^T d \quad (76)$$

$$\text{s.t. } \nabla g_j(\bar{x})^T d \geq 0, j \in \mathcal{I}(\bar{x}) \quad (77)$$

$$\nabla h_i(\bar{x})^T d = 0, i = 1, 2, \dots, \ell. \quad (78)$$

写出其对偶问题, 并利用强对偶定理证明, 在MFCQ条件下的KKT条件成立。

习题2: 给定 $c \in \mathbb{R}^n$, $\rho > 0$ 计算如下问题最优解:

$$\min_w c^T w + \rho \sum_{i=1}^n w_i \log(w_i) \quad (79)$$

$$\text{s.t.} \quad \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n. \quad (80)$$

注: $\log w_i$ 表示以 e 为底, 令 $0 \log 0 = 0$.