MATH 3011 运筹学 **秋季 2023**

Lecture 8: 非线性规划基础

Lecturer: 陈士祥 *Scribes:* 陈士祥

1 问题形式

$$
\begin{aligned}
\min \quad & f(\mathbf{x}) \\
\text{s.t.} \quad & \mathbf{x} \in S \subset \mathbb{R}^n.
\end{aligned}
\tag{8.1}
$$

在此节中，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).

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

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

故，问题(8.1)可以写成如下形式：

$$
\begin{aligned}
\min \quad & f(\mathbf{x}) \\
\text{s.t.} \quad & g_i(\mathbf{x}) \geq 0, i = 1, \cdots, m \\
& h_j(\mathbf{x}) = 0, j = 1, \cdots, \ell
\end{aligned}
\tag{8.2}
$$

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

- 如果求解在约束集合 S 上目标函数 $f(x)$ 的最大值, 则问题 (8.1) 的 “min” 应相应地替换为 “max”.
- 通常来说，我们不考虑严格不等式约束, 如 $g_i(\mathbf{x}) > 0$。因为此时可能没有最优解。
- 为了叙述简便，问题 (8.1) 中 x 为 $\mathbb{R}^n$ 空间中的向量. 实际上，根据具体应用和需求，x 还可以是矩阵、多维数组或张量等.

1.1 最优化问题的类型

- 最优解只有少量非零元素的问题称为**稀疏优化**；

- 最优解是低秩矩阵的问题称为**低秩矩阵优化**.
- 此外还有几何优化、二次锥规划、张量优化、鲁棒优化、全局优化、组合优化、网络规划、随机优化、动态规划、带微分方程约束优化、微分流形约束优化、分布式优化等.
- 就具体应用而言，问题 (8.1) 可涵盖统计学习、压缩感知、最优运输、信号处理、图像处理、机器学习、强化学习、模式识别、金融工程、电力系统等领域的优化模型.

Example 8.1 投资组合优化（*Harry Max Markowitz* 马科维茨）

- r_i, 随机变量, 股票的回报率 i
- x_i, 投资于股票的相对金额 i
- 回报: $r = r_1x_1 + r_2x_2 + \ldots + r_nx_n$
- 期望回报: $R = E(r) = \sum E(r_i)x_i = \sum \mu_i x_i$
- 风险: $V = Var(r) = \sum_{i,j} \sigma_{ij}x_ix_j = x^\top \Sigma x$

$$\begin{aligned} \min \frac{1}{2} x^\top \Sigma x, \\ \text{s.t.} \sum \mu_i x_i \geq r_0 \\ \sum x_i = 1, \\ x_i \geq 0 \end{aligned} \qquad \begin{aligned} \min \quad \text{risk measure}, \\ \text{s.t.} \sum \mu_i x_i \geq r_0 \\ \sum x_i = 1, \\ x_i \geq 0 \end{aligned}$$

1.2 最优化问题的应用

数学建模很容易给出应用问题不同的模型，可以对应性质很不相同的问题，其求解难度和需要的算法也将差别很大. 在投资组合优化中，人们希望通过寻求最优的投资组合以降低风险、提高收益.

- 这时决策变量 x_i 表示在第 i 项资产上的投资额，向量 $x \in \mathbb{R}^n$ 表示整体的投资分配.
- 约束条件可能为总资金数、每项资产的最大（最小）投资额、最低收益等.
- 目标函数通常是某种风险度量.
- 如果是极小化收益的方差，则该问题是典型的二次规划.
- 如果极小化风险价值 (value at risk) 函数，则该问题是混合整数规划
- 如果极小化条件风险价值（conditional value at risk) 函数，则该问题是非光滑优化，也可以进一步化成线性规划.

1.3 最优解

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

$$f(\mathbf{x}^*) \leqslant f(\mathbf{x}), \quad \forall \mathbf{x} \in S. \tag{8.3}$$

则称 $\mathbf{x}^*$ 为问题(8.1)的**全局最优解** (global optimal solution), 对应的函数值叫做**全局最优值**。另外，在包含可行解 $\mathbf{x}^* \in S$ 的适当邻域 $U(\mathbf{x}^*)$ 里，成立

$$f(\mathbf{x}^*) \leqslant f(\mathbf{x}), \forall \mathbf{x} \in S \cap U(\mathbf{x}^*). \tag{8.4}$$

此时称 $\mathbf{x}^*$ 为问题(8.1)的**局部最优解** (local optimal solution)，对应的函数值叫做**局部最优值**。

解的存在唯一性

- 在数学分析课程中, 我们学习过 Weierstrass 定理, 即定义在紧集上的连续函数一定存在最大 (最小) 值点. 下面的推广的 Weierstrass 定理可以保证解存在性。
- 对于一般的凸函数, 其最优解可能不唯一, 比如 $f(x) = \max\{x, 0\}$, 任意 $x \leq 0$ 都是 $f(x)$ 的最优解.

Theorem 8.1 (推广的 Weierstrass 定理) *若函数 $f : \mathcal{X} \to (-\infty, +\infty]$ 适当且闭, 且以下条件中任意一个成立:*

1. $\mathrm{dom} f = \{x \in S : f(x) < +\infty\}$ *是有界的;*
2. *存在一个常数 $\bar{\gamma}$ 使得下水平集*
$$C_{\bar{\gamma}} = \{x \in S : f(x) \leq \bar{\gamma}\}$$
是非空且有界的;
3. f *是强制的, 即对于任一满足 $\|x^k\| \to +\infty$ 的点列 $\{x^k\} \subset \mathcal{X}$, 都有*
$$\lim_{k \to \infty} f(x^k) = +\infty,$$

则函数 f 的最小值点集 $\{x \in \mathcal{X} \mid f(x) \leq f(y), \forall y \in \mathcal{X}\}$ 非空且紧.

Lemma 8.1 *设 S 是 $\mathbb{R}^n$ 的一个非空, 紧且凸的子集, 如果 $f : S \to (-\infty, +\infty]$ 是强凸函数且最小值存在, 那么存在唯一的 x^* 满足*

$$f(x^*) < f(x), \quad \forall x \in S \backslash \{x^*\}.$$

Definition 8.1 *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。

2 无约束最优化条件

Definition 8.2 (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:

$$f(\bar{\mathbf{x}} + \lambda \mathbf{d}) < f(\bar{\mathbf{x}}), \forall \lambda \in (0, \delta) \tag{8.5}$$

则称 $\mathbf{d}$ 为函数 $f(\mathbf{x})$ 在 $\bar{\mathbf{x}}$ 处的下降方向。记下降方向集合为 $\Omega(\bar{\mathbf{x}}, f)$.

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

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

- 一阶最优性条件是利用梯度 (一阶) 信息来判断给定点的最优性.
- 由下降方向的定义, 在局部最优点处不能有下降方向.

Theorem 8.2 (一阶必要条件) 假设 f 在全空间 $\mathbb{R}^n$ 可微. 如果 x^* 是一个局部极小点, 那么

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

Proof:

任取 $v \in \mathbb{R}^n$, 考虑 f 在点 $x = x^*$ 处的泰勒展开

$$f(x^* + tv) = f(x^*) + tv^{\mathrm{T}} \nabla f(x^*) + o(t),$$

整理得

$$\frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}} \nabla f(x^*) + o(1).$$

根据 x^* 的最优性, 在上式中分别对 t 取点 0 处的左, 右极限可知

$$\lim_{t \to 0^+} \frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}} \nabla f(x^*) \geq 0,$$
$$\lim_{t \to 0^-} \frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}} \nabla f(x^*) \leq 0,$$

即对任意的 v 有 $v^{\mathrm{T}} \nabla f(x^*) = 0$, 由 v 的任意性知 $\nabla f(x^*) = 0$. ■

- 在没有额外假设时, 如果一阶必要条件满足, 我们仍然不能确定当前点是否是一个局部极小点. 例如 $f(x) = x^3$, 在 $x = 0$ 处不是局部极小点。
- 假设 f 在点 x 的一个开邻域内是二阶连续可微的. 类似于一阶必要条件的推导, 可以借助当前点处的二阶泰勒展开来逼近该函数在该点附近的取值情况, 从而来判断最优性.

- 在点 x 附近我们考虑泰勒展开

$$f(x+d)=f(x)+\nabla f(x)^{\mathrm{T}}d+\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x)d+o(\|d\|^2).$$

- 当一阶必要条件满足时, $\nabla f(x)=0$, 那么上面的展开式简化为

$$f(x+d)=f(x)+\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x)d+o(\|d\|^2).$$

  由此, 我们可以导出二阶最优性条件.

Theorem 8.3 (二阶最优性条件) *必要条件: 若 x^* 是 f 的一个局部极小点, 则 $\nabla f(x^*)=0, \nabla^2 f(x^*)\succeq 0$.*

充分条件: 若 $\nabla f(x^*)=0, \nabla^2 f(x^*)\succ 0$, 则 x^* 是 f 的一个局部极小点.

Proof:

- **必要性**: 若 $\nabla^2 f(x^*)$ 有负的特征值 $\lambda_-<0$, 设 $\nabla^2 f(x^*)d=\lambda_- d$, 则

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

  当 $\|d\|$ 充分小时, $f(x^*+d)<f(x^*)$, 这和点 x^* 的最优性矛盾.

- **充分性**: 由 $\nabla f(x^*)=0$ 时的二阶展开,

$$\frac{f(x^*+d)-f(x^*)}{\|d\|^2}=\frac{d^{\mathrm{T}}\nabla^2 f(x^*)d+o(\|d\|^2)}{\|d\|^2}\geq\frac{1}{2}\lambda_{\min}+o(1).$$

  这里 $\lambda_{\min}$ 是 $\nabla^2 f(x^*)$ 的最小特征根。当 $\|d\|$ 充分小时有 $f(x^*+d)\geq f(x^*)$, 即二阶充分条件成立.

■

注:

- 设点 $\bar{x}$ 满足一阶最优性条件 (即 $\nabla f(\bar{x})=0$), 且该点处的海瑟矩阵 $\nabla^2 f(\bar{x})$ 不是半正定的, 那么 $\bar{x}$ 不是一个局部极小点.

- 进一步地, 如果海瑟矩阵 $\nabla^2 f(\bar{x})$ 既有正特征值又有负特征值, 我们称稳定点 $\bar{x}$ 为**鞍点** (saddle point).

- 事实上, 记 d_1, d_2 为其正负特征值对应的特征向量, 那么对于任意充分小的 $t>0$, 我们都有 $f(\bar{x}+td_1)>f(\bar{x})$ 且 $f(\bar{x}+td_2)<f(\bar{x})$.

- 注意, 二阶最优性条件给出的仍然是关于局部最优性的判断. 对于给定点的全局最优性判断, 我们还需要借助实际问题的性质, 比如目标函数是凸的、非负目标函数值为 0 等.

Example 8.2 线性最小二乘:
$$\min_{x\in\mathbb{R}^n} \quad f(x) \stackrel{\text{def}}{=\!=} \frac{1}{2}\|b-Ax\|_2^2,$$
其中 $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$. 由 f 可微且凸知

$$x^* \text{ 为一个全局最优解} \Leftrightarrow \nabla f(x^*) = A^{\mathrm{T}}(Ax^*-b)=0.$$

3 约束问题的最优性条件

3.1 可行方向，切锥，线性可行锥方向

若考虑到约束，需要先定义可行方向。

Definition 8.3 (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:

$$\bar{\mathbf{x}}+\lambda\mathbf{d}\in S, \forall\lambda\in(0,\delta) \tag{8.6}$$

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 $F(\bar{\mathbf{x}},S)$.

Theorem 8.4 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

$$F(\bar{\mathbf{x}},S)\cap\Omega(\bar{\mathbf{x}},f)=\emptyset. \tag{8.7}$$

Proof: 反证即可。若 $F(\bar{\mathbf{x}},S)\cap\Omega(\bar{\mathbf{x}},f)$ 非空，则 $\mathbf{x}$ 处存在可行下降方向，故不是局部最优。 ■

上面定义的下降方向，以及复杂约束 S 的可行方向，均比较难以刻画。**我们的目标是寻找易于用数学公式表述的定义**。首先，我们先将下降方向的子集刻画如下。

Definition 8.4 (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

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

Theorem 8.5 对于问题(8.1), 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，则下降方向子集 $D(\bar{\mathbf{x}},f)$ 中无可行方向，即
$$F(\bar{\mathbf{x}},S)\cap D(\bar{\mathbf{x}},f)=\emptyset. \tag{8.8}$$

对于不等式约束，我们同样可以定义其可行方向的子集。

Definition 8.5 (不等式可行方向的子集)

$$F_{\bar{\mathbf{x}},g} = \{\mathbf{d} \mid \nabla g_i(\bar{\mathbf{x}})^T \mathbf{d} > 0, i \in \mathcal{A}(\bar{\mathbf{x}})\}.$$

但是，对于有些等式约束，例如二维圆环 $x_1^2 + x_2^2 = 1$ 上, 可行方向不存在（想一想，为什么?）。上述定理 8.5对于多面体集可以适用。但对于一般的非线性等式约束确定的可行域 C, 定义在点 $x \in C$ 处的切锥为从在中收敛到 x 的序列获得的极限方向的集合。

Definition 8.6 (切锥) 我们称方向 $d \in \mathbb{R}^n$ 属于可行点 $x \in S$ 处的切锥，如果存在序列 $\{x_i\} \subset S$, 和实数序列 $\tau_i \searrow 0$, 使得

$$\frac{x_i - x}{\tau_i} \to d,\ i \to \infty.$$

记其切锥集合为

$$T(x \mid S) := \{d : \exists \tau_i \searrow 0, \{x_i\} \subset S, x_i \to x, \text{s.t.} \frac{x_i - x}{\tau_i} \to d\}.$$

根据切锥的定义，显然对于任意可行方向 $d \in F(\bar{\mathbf{x}}, S)$, $d \in T(\bar{\mathbf{x}}, S)$. 故

$$F(\bar{\mathbf{x}}, S) \subset T(\bar{\mathbf{x}}, S).$$

Theorem 8.6 对于问题(8.1), 设 $\bar{\mathbf{x}} \in S$, $f(\mathbf{x})$ 在 $\bar{\mathbf{x}}$ 处可微。如果 $\bar{\mathbf{x}}$ 是局部最优解，则

$$\nabla f(\bar{\mathbf{x}})^T d \geq 0, \forall d \in T(x \mid S). \tag{8.9}$$

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

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

■

Corollary 8.1 问题(8.1)的局部最优点 $\bar{\mathbf{x}}$ 满足

$$T(\bar{\mathbf{x}} \mid S) \cap D(\bar{\mathbf{x}}, f) = \emptyset. \tag{8.10}$$

上述极限定义的切锥仍然不易表达。针对等式和不等式定义的约束，我们分别有下面的结论。

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

$$\nabla h_i(\bar{\mathbf{x}})^T d = h_i(\bar{\mathbf{x}}; d) = \lim_{k\to\infty} \frac{h_i(\bar{\mathbf{x}} + \tau_k d_k) - h_i(\bar{\mathbf{x}})}{\tau_k} = 0. \tag{8.11}$$

- 同理，对于不等式积极约束集（active set），记 $\mathcal{A}(\bar{\mathbf{x}}) = \{i \mid g_i(\bar{\mathbf{x}}) = 0, i \in \{1, \cdots, m\}\}$,

$$\nabla g_i(\bar{\mathbf{x}})^T d = g_i(\bar{\mathbf{x}}; d) = \lim_{k\to\infty} \frac{g_i(\bar{\mathbf{x}} + \tau_k d_k) - g_i(\bar{\mathbf{x}})}{\tau_k} \geq 0, i \in \mathcal{A}(\bar{\mathbf{x}}). \tag{8.12}$$

另外，我们由(8.10)可知，

$$T(\bar{\mathbf{x}} \mid g) \cap T(\bar{\mathbf{x}} \mid h) \cap D(\bar{\mathbf{x}}, f) = \emptyset. \tag{8.13}$$

由 (8.11)和(8.12)，进一步定义如下**线性可行锥方向**。其优点是可以通过计算约束函数的梯度得到。

Definition 8.7 (等式的线性可行锥方向) $L(\bar{\mathbf{x}}, h) := \{\mathbf{d} \mid \nabla h_j(\bar{\mathbf{x}})^T \mathbf{d} = 0, j = 1, \cdots, \ell\}$.

Definition 8.8 (不等式的线性可行锥方向) $L(\bar{\mathbf{x}}, g) := \{\mathbf{d} \mid \nabla g_i(\bar{\mathbf{x}})^T \mathbf{d} \geq 0, i \in \mathcal{A}(\bar{\mathbf{x}})\}$.

我们有

$$F_{\bar{\mathbf{x}},g} \subset F(\bar{\mathbf{x}}, g) \subset T(\bar{\mathbf{x}} \mid g) \subset L(\bar{\mathbf{x}}, g), \tag{8.14}$$

最后一个子集关系由(8.12)得到。

$$F(\bar{\mathbf{x}}, h) \subset T(\bar{\mathbf{x}} \mid h) \subset L(\bar{\mathbf{x}}, h), \tag{8.15}$$

最后一步由(8.11)得到。下图 8.1描述了几种定义.

假设切锥与线性可行锥方向相等，则我们可以得到易于计算的最优性条件：

$$L(\bar{\mathbf{x}}, g) \cap L(\bar{\mathbf{x}}, h) \cap D(\bar{\mathbf{x}}, f) = \emptyset.$$

然而线性可行锥方向与约束的表示形式有关。同一个可行域，在不同的的表述方式下，线性可行锥方向可能不同。

Example 8.3 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 $T(x, S) = \{(0, 0)^T\}$，然而 $L(x, S) = \{(0, t) \mid t \in \mathbb{R}\}$。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 $T(x, S) = L(x, S)$.

为此，我们引入以下约束规范条件 Mangasarian-Fromovitz Constraint Qualification (MFCQ) 条件, 以排除上面的第一种定义方式影响最优条件。

Definition 8.9 (MFCQ) *在问题(8.2)中，f 和 $g_i, i \in \mathcal{A}(\bar{\mathbf{x}})$ 在点 $\bar{\mathbf{x}}$ 可微，$g_i, i \notin \mathcal{A}(\bar{\mathbf{x}})$ 在点 $\bar{\mathbf{x}}$ 连续，且 $\{\nabla h_j(\bar{\mathbf{x}})\}_{j=1}^{\ell}$ 线性无关，$F_{\bar{\mathbf{x}},g} \cap L(\bar{\mathbf{x}}, h)$ 非空，则称该点处满足 MFCQ 约束规范条件。*

事实上，我们有 LICQ $\implies$ MFCQ. 这个我们不去证明。

Lemma 8.2 (规范性引理) *若在可行点 $\bar{\mathbf{x}} \in S$ 处满足 LICQ 或者 MFCQ，则 $L(\bar{\mathbf{x}}, h) \cap L(\bar{\mathbf{x}}, g) = T(\bar{\mathbf{x}} \mid g) \cap T(\bar{\mathbf{x}} \mid h)$.*

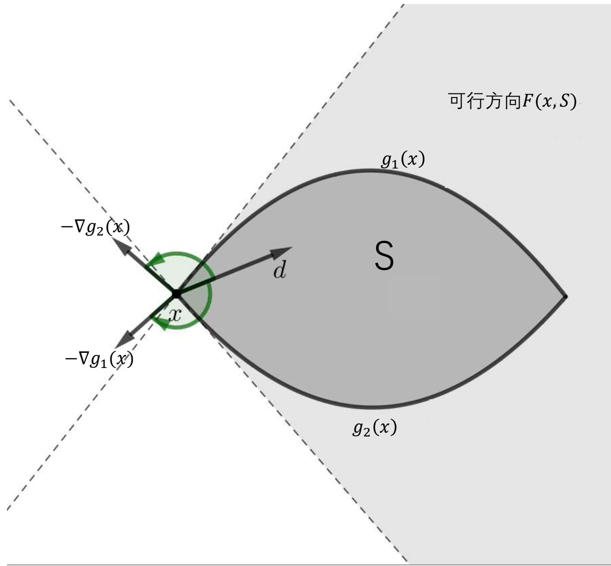


图 8.1: $\mathbb{R}^2$ 上的两个不等式约束集合，深灰色 S 是可行域，可行方向为浅灰色区域，不包括边界虚线。而线性可行锥方向是浅灰色区域包括虚线。

该引理我们也略去证明。因此，在 LICQ 或者 MFCQ 条件下，根据(8.13)我们有

$$L(\bar{\mathbf{x}}, g) \cap L(\bar{\mathbf{x}}, h) \cap D(\bar{\mathbf{x}}, f) = \emptyset. \tag{8.16}$$

3.2 Fritz-John 条件

Fritz-John 条件仅是判别某点是否为最优点的必要条件，而不是充分条件。它是为后面我们证明 Karush-Kuhn-Tucker (KKT) 条件做准备。

根据上一小节的准备工作，我们有下面的结论。

Lemma 8.3 设 $\bar{\mathbf{x}}$ 为问题(8.2)的局部最优解，f 和 $g_i, i \in \mathcal{A}(\bar{\mathbf{x}})$ 在点 $\bar{\mathbf{x}}$ 可微，$g_i, i \notin \mathcal{A}(\bar{\mathbf{x}})$ 在点 $\bar{\mathbf{x}}$ 连续，h_j 在点 $\bar{\mathbf{x}}$ 连续可微，且 $\{\nabla h_j(\bar{\mathbf{x}})\}_{j=1}^{\ell}$ 线性无关，则

$$D(\bar{\mathbf{x}}, f) \cap F_{\bar{\mathbf{x}},g} \cap L(\bar{\mathbf{x}}, h) = \emptyset. \tag{8.17}$$

Proof: 该引理证明分为两种情况。若点 $\bar{\mathbf{x}}$ 处满足 MFCQ，则根据(8.16)和 $F_{\bar{\mathbf{x}},g} \subset L_(\bar{\mathbf{x}}, g)$ 得出结论。若点 $\bar{\mathbf{x}}$ 不满足 MFCQ，则 $F_{\bar{\mathbf{x}},g} \cap L(\bar{\mathbf{x}}, h) = \emptyset$, 得证。 ■

为了证明 Fritz-John 条件，我们有如下凸集分离定理。

Theorem 8.7 (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 $x_1 \in \mathrm{CL}(C_1)$ 和 $x_2 \in \mathrm{CL}(C_2)$ 有：

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

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

Theorem 8.8 (Fritz-John 条件) 在问题(8.2)中，设 $\bar{\mathbf{x}}$ 为可行点，f 和 $g_i, i \in \mathcal{A}(\bar{\mathbf{x}})$ 在点 $\bar{\mathbf{x}}$ 可微，$g_i, i \notin \mathcal{A}(\bar{\mathbf{x}})$ 在点 $\bar{\mathbf{x}}$ 连续，h_j 在点 $\bar{\mathbf{x}}$ 连续可微。如果 $\bar{\mathbf{x}}$ 是局部最优解，则存在不全为零的数 $\lambda_0, \lambda_i, i \in \mathcal{A}(\bar{\mathbf{x}})$ 和 $\mu_j, j = 1, \cdots, \ell$ 使得

$$\lambda_0 \nabla f(\bar{\mathbf{x}}) - \sum_{i \in \mathcal{A}(\bar{\mathbf{x}})} \lambda_i \nabla g_i(\bar{\mathbf{x}}) - \sum_{j=1}^{\ell} \mu_j \nabla h_j(\bar{\mathbf{x}}) = 0. \tag{8.18}$$

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

Proof:

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 $\mu_j, j = 1, \cdots, \ell$ 使得

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

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

(2) 如果 $\{\nabla h_j(\bar{\mathbf{x}})\}_{j=1}^{\ell}$ 线性无关。利用 Lemma 8.3知 $D(\bar{\mathbf{x}}, f) \cap F_{\bar{\mathbf{x}},g} \cap L(\bar{\mathbf{x}}, h) = \emptyset$.
即不等式组

$$\begin{cases} \nabla f(\bar{\mathbf{x}})^T \mathbf{d} < 0 \\ \nabla g_i(\bar{\mathbf{x}})^T \mathbf{d} > 0, i \in \mathcal{A}(\bar{\mathbf{x}}) \\ \nabla h_j(\bar{\mathbf{x}})^T \mathbf{d} = 0, j = 1, \cdots, \ell \end{cases} \tag{8.19}$$

无解。

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

$$\begin{cases} A^T \mathbf{d} < 0 \\ B^T \mathbf{d} = 0 \end{cases} \tag{8.20}$$

无解。

下证

$$\begin{cases} A\mathbf{p}_1 + B\mathbf{p}_2 = 0 \\ \mathbf{p}_1 \geq 0 \end{cases} \tag{8.21}$$

有解。

现定义

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

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

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

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 $\mathbf{y}_1 < \mathbf{0}$, $\Longrightarrow \mathbf{p}_1 \geq \mathbf{0}$.

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

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

综上所述，即得(8.21)有解。

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

$$\lambda_0 \nabla f(\bar{\mathbf{x}}) - \sum_{i \in I(\bar{\mathbf{x}})} \lambda_i \nabla g_i(\bar{\mathbf{x}}) - \sum_{j=1}^{\ell} \mu_j \nabla h_j(\bar{\mathbf{x}}) = 0. \tag{8.22}$$

■

3.3 KKT 条件

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

3.4 一阶必要条件

若 $\bar{\mathbf{x}}$ 为问题局部最优解，且 $\bar{\mathbf{x}}$ 处满足 LICQ 条件。此时，一阶必要条件可表达为

$$(\mathrm{KKT}) \begin{cases} \textbf{稳定性条件} & \nabla_{\mathbf{x}} \mathcal{L}(\bar{\mathbf{x}}, \lambda, \mu) = 0 \\ \textbf{原始可行性条件} & g_i(\bar{\mathbf{x}}) \geq 0, i = 1, \cdots, m \\ \textbf{互补松弛条件} & \lambda_i g_i(\bar{\mathbf{x}}) = 0, i = 1, \cdots, m \\ \textbf{对偶可行性条件} & \lambda_i \geq 0, i = 1, \cdots, m \\ \textbf{原始可行性条件} & h_j(\bar{\mathbf{x}}) = 0, j = 1, \cdots, \ell \end{cases} \tag{8.23}$$

实际上，任何可以保证切锥和线性可行锥相等的条件，均可推出上述的 KKT 一阶条件。我们这里给出了利用 LICQ 条件的证明。

证明：利用 Fritz-John 条件，若 LICQ 成立，则 $\lambda_0 \neq 0$. 这是因为如果 λ_0, 利用线性无关行，可得 $\lambda_i = 0, \forall i = 1, 2..., m, \mu_j = 0, \forall j = 1, 2, ..., \ell$. (8.22)两边同除以 λ_0 得

$$\nabla f(\bar{\mathbf{x}}) - \sum_{i \in I(\bar{\mathbf{x}})} \lambda_i \nabla g_i(\bar{\mathbf{x}}) - \sum_{j=1}^{\ell} \mu_j \nabla h_j(\bar{\mathbf{x}}) = 0. \tag{8.24}$$

若下标 $i \notin \mathcal{A}(\bar{x})$, 那么 $g_i(\bar(x))$，因此我们可以令相应的 $\lambda_i = 0$. 于是有稳定性条件：$\nabla_{\mathbf{x}} L(\bar{\mathbf{x}}, \lambda, \mu) = 0$, 以及互补松弛性。原始可行性和对偶可行性源自于 Fritz-John 条件结论以及 Lagrange 函数的定义。

作业 8.1 *对于问题(8.2), 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：

$$\max(-\nabla f(\bar{x}))^T d \tag{8.25}$$

$$\text{s.t.} \quad \nabla g_j(\bar{x})^T d \geq 0, j \in \mathcal{I}(\bar{x}) \tag{8.26}$$

$$\nabla h_i(\bar{x})^T d = 0, i = 1, 2, \ldots, \ell. \tag{8.27}$$

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。

注：

- 称满足 KKT 条件的变量对 $(\bar{\mathbf{x}}, \lambda, \mu)$ 为 KKT 对.
- 称 $\bar{\mathbf{x}}$ 为 KKT 点.
- 如果局部最优点 $\bar{\mathbf{x}}$ 处 $T_{\bar{\mathbf{x}},S} \neq L(\bar{\mathbf{x}}, S)$, 那么 x^* 不一定是 KKT 点.
- KKT 条件只是必要的, KKT 点不一定是局部最优点. 也就是说，可以通过计算 KKT 条件，求出最优点的备选集合。

3.5 充分条件

我们这里给出特殊的凸优化问题的充分最优性条件。即问题都可以写为

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \leqslant 0, \quad i = 1, 2, \cdots, m, \\ & Ax = b, \end{aligned} \tag{8.28}$$

其中 $f(x)$ 为适当的凸函数, $\forall i, c_i(x)$ 是凸函数且 $\mathrm{dom} c_i = \mathbb{R}^n$.

Definition 8.10 (Slater 条件) 若对凸优化问题

$$\min_{x\in\mathcal{D}} \quad f(x),\ s.t.\ c_i(x) \leqslant 0,\ i = 1, 2, \cdots, m, \quad Ax = b,$$

存在 $x \in \mathbb{R}^n$ 满足

$$c_i(x) < 0, \quad i = 1, 2, \cdots, m, \quad Ax = b,$$

则称对此问题 Slater 约束品性满足. 该约束品性也称为 Slater **条件**.

对于凸优化问题, 当 Slater 条件满足时, KKT 条件则变为局部最优解的充要条件 (根据凸性, 局部最优解也是全局最优解).

Theorem 8.9 对于问题(8.28), 若 *Slater* 条件成立, x^* 是全局最优点当且仅当是 *KKT* 点。

Example 8.4

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \frac{1}{2}\|x-y\|_2^2, \\ s.t. \quad & Ax = b, \end{aligned}$$

其中 $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ 以及 $y \in \mathbb{R}^n$ 为给定的矩阵和向量且 A 满秩.

- 拉格朗日函数: $L(x,\lambda) = \frac{1}{2}\|x-y\|^2 + \lambda^{\mathrm{T}}(Ax-b)$.

- *Slater* 条件成立, x^* 为一个全局最优解当且仅当存在 $\lambda^* \in \mathbb{R}^m$ 使得

$$\begin{cases} x^* - y + A^{\mathrm{T}}\lambda^* = 0, \\ \qquad\qquad Ax^* = b. \end{cases}$$

- 由上述 *KKT* 条件第一式, 等号左右两边同时左乘 A 可得

$$Ax^* - Ay + AA^{\mathrm{T}}\lambda = 0 \Rightarrow \lambda^* = (AA^{\mathrm{T}})^{-1}(Ay-b).$$

- 将 λ^* 代回 *KKT* 条件第一式可知

$$x^* = y - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(Ay-b).$$

  因此点 y 到集合 $\{x \mid Ax = b\}$ 的投影为 $y - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(Ay-b)$.

3.6 二阶最优条件

再次考虑约束问题(8.2)，即

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

假设 x^* 是满足 KKT 条件的点，并且切锥与线性可行锥方向相等，即 $T(x^* \mid S) = L(x^*, S)$, 这里 $L(x^*, S)$ 是线性可行锥。则 $\forall d \in T(x^*, S)$

$$d^T \nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla g_i(x^*)^T d + \sum_{j=1}^{\ell} \mu_j^* \nabla h(x^*)^T d \geq 0.$$

此时一阶条件无法判断 x^* 是否是最优值点.

- 若 $d^T \nabla f(x^*) = 0$, 则需要利用二阶信息来进一步判断在其可行邻域内的目标函数值.
- 拉格朗日函数在这些方向上的曲率即可用来判断 x^* 的最优性.
- 这里引入临界锥来精确刻画这些方向.

Definition 8.11 (临界锥) *设 (x^*, λ^*, μ^*) 是满足 KKT 条件的 KKT 对, 定义临界锥为*

$$C(x^*, \lambda^*, \mu^*) = \{d \in L(x^*, S) | \nabla g_i(x^*)^T d = 0, \forall i \in \mathcal{A}(x^*) \text{且} \lambda_i^* > 0\}$$

其中 $L(x^*, S)$ 为点 x^* 处的线性化可行方向锥.

1. 临界锥是线性化可行锥 $L(x^*, S)$ 的子集.
2. 沿着临界锥中的方向进行优化, 所有等式约束和 $\lambda_i^* > 0$ 对应的不等式约束 (此时这些不等式均取等) 都会尽量保持不变.
3. 当 $d \in C(x^*, \lambda^*, \mu^*)$ 时, $\forall i = 1, ..., m, j = 1, ..., \ell$ 有 $\lambda_i^* \nabla g_i(x^*)^T d = 0$, $\mu_j^* \nabla h_j(x^*)^T d = 0$, 故

$$d^T \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*)^T d + \sum_{j=1}^{\ell} \mu_j^* \nabla h(x^*)^T d = 0$$

4. 临界锥定义了依据一阶导数不能判断是否为下降或上升方向的线性化可行方向, 必须使用高阶导数信息加以判断.

Theorem 8.10 必要性*:* *假设 x^* 是问题的一个局部最优解, 并且 $T(x^*, S) = L(x^*, g) \cap \mathcal{L}(x^*, h)$ 成立. 令 (x^*, λ^*, μ^*) 满足 KKT 条件, 那么*

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) d \geq 0, \quad \forall d \in C(x^*, \lambda^*, \mu^*).$$

充分性: 假设在可行点 x^* 处, 存在一个拉格朗日乘子 λ^*, 使得 (x^*, λ^*, μ^*) 满足 *KKT* 条件. 如果

$$d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*) d > 0, \quad \forall d \neq 0, d \in C(x^*, \lambda^*, \mu^*),$$

那么 x^* 为问题的一个严格局部极小解.

回顾无约束优化问题的二阶最优性条件:

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

必要条件: 若 x^* 是 f 的一个局部极小点, 则 $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$
充分条件: 若 $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succ 0$, 则 x^* 是 f 的一个局部极小点
约束优化问题的二阶最优性条件也要求某种“正定性”, 但只需要考虑临界锥 $C(x^*, \lambda^*, \mu^*)$ 中的向量而无需考虑全空间的向量.

Example 8.5

$$\min x_1^2 + x_2^2$$

$$s.t.\ \frac{x_1^2}{4} + x_2^2 - 1 = 0$$

其拉格朗日函数为

$$\mathcal{L}(x, \lambda) = x_1^2 + x_2^2 + \lambda(\frac{x_1^2}{4} + x_2^2 - 1)$$

该问题可行域在任意一点 $x = (x_1, x_2)^T$ 处的线性化可行方向锥为

$$L(x, S) = \{(d_1, d_2) | \frac{x_1}{4} d_1 + x_2 d_2 = 0\}$$

因为只有一个等式约束且其对应函数的梯度非零, 故有 *LICQ* 成立, 于是

$$L(x, S) = T(x, S)$$

若 (x, λ) 为 *KKT* 对, 由于无不等式约束, 故临界锥等于线性可行锥, 即

$$C(x, \lambda) = L(x, S)$$

可以计算出其 *4* 个 *KKT* 对

$$(x^T, \lambda) = (2, 0, -4), (-2, 0, -4), (0, 1, -1), (0, -1, -1)$$

考虑第一个 KKT 对 $y = (2, 0, -4)^T$, 计算可得

$$\nabla^2_{xx}\mathcal{L}(y, \lambda) = \begin{bmatrix} 0 & 0 \\ 0 & -6 \end{bmatrix}$$

$$C(y) = \{(d_1, d_2) | d_1 = 0\}$$

取 $d = (0, 1)$, 则

$$d^T \nabla^2_{xx}\mathcal{L}(y, \lambda) d = -6 < 0$$

因此 y 不是局部最优点. 类似地, 对第三个 KKT 对 $z = (0, 1, -1)$,

$$\nabla^2_{xx}\mathcal{L}(z, \lambda) = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 0 \end{bmatrix}$$

$$C(z) = \{(d_1, d_2) | d_2 = 0\}$$

对于任意的 $d = (d_1, 0)$ 且 $d_1 \neq 0$,

$$d^T \nabla^2_{xx}\mathcal{L}(z, \lambda) d = \frac{3}{2} d_1^2 > 0$$

因此, z 为一个严格局部最优点.

总结:

问题	一阶条件	二阶条件
可微问题	$\nabla f(x^*) = 0$ (必要)	$\nabla^2 f(x^*) \succeq 0$ (必要) $\nabla^2 f(x^*) \succ 0$ (充分)
凸问题	$\nabla f(x^*) = 0$ (充要)	—

问题	一阶条件	二阶条件	约束品性
一般问题	KKT 条件 (必要)	$d^{\mathrm{T}} \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d \geq 0, \quad \forall d \in \mathcal{C}(x^*, \lambda^*, \mu^*)$ (必要) $d^{\mathrm{T}} \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d > 0, \quad \forall d \in \mathcal{C}(x^*, \lambda^*, \mu^*),\ d \neq 0$, (充分)[1]	LICQ[2]
凸问题	KKT 条件 (充要)	—	Slater

1. 一般约束优化问题的二阶充分条件不需要 LICQ 作为前提.

2. 或其他可推出切锥等于线性可行锥的约束规范性条件.

4 优化算法前置知识

4.1 收敛速率

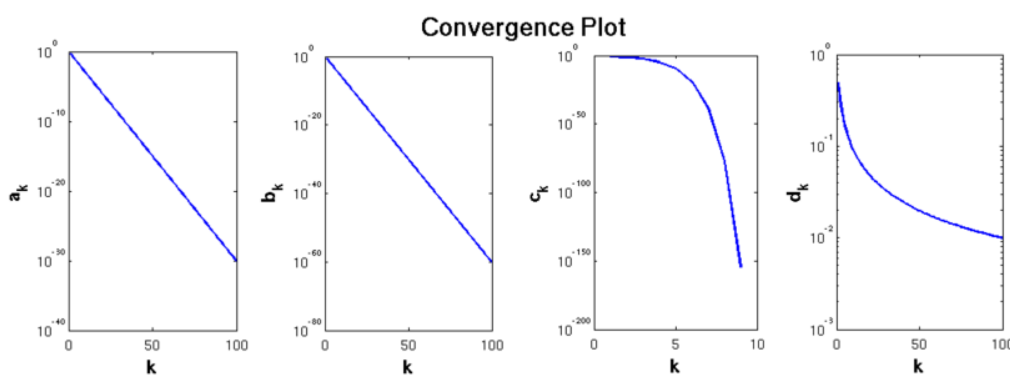
评价算法优劣的标准之一是收敛的快慢，通常称为收敛速率。

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

$$0 \leq \overline{\lim_{k\to\infty}} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^p} = \beta < \infty \tag{8.29}$$

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(8.29)中，$p = 1$ 且 $\beta < 1$, 则称序列是（收敛比 β）线性收敛的。例如 $\mathbf{x}^{(k)} = \frac{1}{2^k}$
- 若在定义式(8.29)中，$p > 1$, 或者 $\{p = 1, \beta = 0\}$, 则称序列是超线性收敛的。
- 若在定义式(8.29)中，$p = 2$, 或则称序列是二次收敛的。例如 $\mathbf{x}^{(k)} = \frac{1}{2^{2^k}}$
- 次线性收敛（sublinear）：一般形如 $\mathbf{x}^{(k)} = \mathcal{O}(1/k), \mathcal{O}(1/k^2)$ 的序列。



上述图像，分别表示线性收敛、超线性收敛、二次收敛和次线性收敛。图中的纵坐标是以 $\log_1 0$ 为基准做放缩后的值。

4.2 优化算法基本格式

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。

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。