

证明. 由条件可知 $p(x)$ 是最高次系数为正数的偶次多项式, 因此存在最小值点. 设 a 为 p 的一个最小值点, 则根据极值原理有 $p'(a) = 0$ 且 $p''(a) \geq 0$, 从而 $p(a) \geq p'(a) + p''(a) \geq 0$. \square

例 1.4.6 若 $[a, b]$ 上的二阶可导函数 f 满足 $f''(x) + p(x)f'(x) + q(x)f(x) = 0$, 其中 $q(x) < 0$. 证明: 若 $f(a) = f(b) = 0$, 则 $f(x) = 0$.

证明. 如果 f 不恒为零, 则它能取到正值或负值. 若它能取到正值, 则可假设它在 $\xi \in (a, b)$ 处取得正的最大值. 根据极值原理有 $f'(\xi) = 0, f''(\xi) \leq 0$. 从而

$$0 = f''(\xi) + p(\xi)f'(\xi) + q(\xi)f(\xi) \geq q(\xi)f(\xi) < 0,$$

矛盾. 因此 f 取不到正值. 类似可知它也取不到负值. 所以 f 恒为零. \square

1.5 凸函数

凸函数的历史可以追溯至 19 世纪末的赫德尔、斯托尔茨、阿达玛 (Hadamard). 詹森 (Jensen) 首次意识到凸函数的重要性并开始了系统性研究. 现在凸分析已经发展为分析领域的一个独立分支, 在最优优化、控制论等领域有着重要应用.

凸函数

设 f 在区间 I 上有定义. 如果对于任意的 $x, y \in I$ 及 $t \in (0, 1)$ 成立

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y),$$

即曲线段位于割线段下方, 则称 f 是 I 上的**凸函数 (convex function)**. 若不等式反向, 则称为**凹函数 (concave function)**. 若不等式是严格的, 则称其为**严格凸/凹函数**.³

虽然凸函数的定义中仅要求曲线段在割线段下方 (不严格), 事实上可以证明**割线段两侧的函数图像位于割线上方 (不严格)**. 不妨设割线段位于区间 $[a, b]$. 若 $c > b$ 处函数图像 $(c, f(c))$ 严格位于割线下方, 则 a, b, c 三点不满足凸性条件, 因为此时 $(b, f(b))$ 严格位于 $[a, c]$ 割线的上方.

部分文献中的凸函数定义为 $f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}$, 本书称之为**中点凸**. 在函数连续的条件下, 凸与中点凸是等价的, 证明留给读者. 值得指出的是, 凸函数定义中的两点可以推广到 n 点.

詹森 (Jensen) 不等式

设 f 是 I 上的凸函数, 则当 $n \geq 2$ 时成立

$$f(t_1x_1 + \cdots + t_nx_n) \leq t_1f(x_1) + \cdots + t_nf(x_n)$$

其中 $x_j \in I, t_j \geq 0$ 且 $t_1 + t_2 + \cdots + t_n = 1$.

3: 凸与凹的命名几乎陷于一种混乱. 这种混乱的原因在于, 凸与凹本质上是同一种性质, 只是观察的角度不同而已. 因此, 本书所指的凸与凹在其它文献中可能有着完全相反的含义. 在英文文献中, convex 和 concave 也称为 concave up 和 concave down, 或 convex down 和 convex up. 在中文语境下, 也有上凹与下凹、上凸与下凸的说法.

证明. 运用数学归纳法. 根据定义 $n = 2$ 时成立. 假设 $n = k$ 时成立, 下证 $n = k + 1$ 时成立.

$$\begin{aligned} f(t_1 x_1 + \cdots + t_{k+1} x_{k+1}) &= f\left((1 - t_{k+1}) \frac{t_1 x_1 + \cdots + t_k x_k}{1 - t_{k+1}} + t_{k+1} x_{k+1}\right) \\ &\leq (1 - t_{k+1}) f\left(\frac{t_1 x_1 + \cdots + t_k x_k}{1 - t_{k+1}}\right) + t_{k+1} f(x_{k+1}) \\ &\leq t_1 f(x_1) + \cdots + t_k f(x_k) + t_{k+1} f(x_{k+1}). \end{aligned}$$

因此成立詹森不等式. \square

下面的性质奠定了凸函数在优化问题中的地位.

凸函数的极值是最值

- ▶ 设 f 是 (a, b) 上的凸函数, 则 f 的极值点必是最小值点.
- ▶ 设 f 在 (a, b) 上严格凸, 则 f 至多只有一个极值点.

证明. 先来证明第一条. 设 x_* 是 f 的极值点, 若它不是最小值点, 则存在 ξ 使得 $f(\xi) < f(x_*)$. 不妨假设 $\xi < x_*$.

- (1) 若 x_* 是极小值点, 则在 (ξ, x_*) 上存在 x' , 使得 $f(x') \geq f(x_*) > f(\xi)$, 从而导致 ξ, x', x_* 不满足凸性条件.
- (2) 若 x_* 是极大值点, 则在 (x_*, b) 上存在 x'' , 使得 $f(x'') \leq f(x_*)$, 从而导致 ξ, x_*, x'' 不满足凸性条件.

因此 x_* 必是 f 的最小值点.

下面证明第二条. 如果函数有两个极值点, 则它们都是最小值点. 但根据严格凸性, 它们之间的函数值比最小值更小, 矛盾. 因此至多只有一个极值点. \square

凸函数单侧可导

设 f 是 I 上的凸函数, 如果 ξ 是 I 的内点, 则 f 在 ξ 点单侧可导, 进而连续.

证明. 仅证 f 在 ξ 左可导. 在区间 I 内取 $a < \xi < b$. 考察过 $(\xi, f(\xi))$ 的割线的斜率

$$k(x) = \frac{f(x) - f(\xi)}{x - \xi}.$$

因为曲线段在割线段下方, 所以当 $x \rightarrow \xi^-$ 时 $k(x)$ 单调递增; 再者, $[a, \xi]$ 的函数图像位于 $[\xi, b]$ 割线的上方, 所以对任意 $x \in (a, \xi)$ 有 $k(x) \geq k(b)$. 故而, 根据单调收敛定理, $\lim_{x \rightarrow a^-} k(x)$ 存在, 极限即为 $f'(a)$. \square

要注意的是, 的确存在不可导的凸函数, 比如 $|x|$. 此外, 端点处的连续性也无法保证, 比如将凸函数 $x^2 (-1 \leq x \leq 1)$ 在端点处的值增大 1 之后仍然是凸函数, 但端点处不再连续.

3 课时/51 课时

根据定义判断函数凸性并不容易, 下面给出两个导数判别法.

凸性的一阶导数判定

设 f 在区间 I 上可导, 则下述三条等价:

- (i) f 是 I 上的凸函数;
- (ii) f' 在 I 上单调递增;
- (iii) f 的图像在其任何一点的切线上方.

证明. 下面按 (i) 推 (ii)、(ii) 推 (iii)、(iii) 推 (i) 的方式来证明.

(i) \Rightarrow (ii) 设 $x < y$, 过两点的割线斜率记为 k . 由单侧可导性的证明过程可知

$$f'(x) = f'_+(x) \leq k \leq f'_-(y) = f'(y).$$

因此 f' 单调递增.

(ii) \Rightarrow (iii) 考虑 $(a, f(a))$ 处的图景. 切线为 $y = f'(a)(x - a) + f(a)$, 要证明 $f(x) \geq f'(a)(x - a) + f(a)$. 根据 Lagrange 中值定理以及单调性, 可知

$$f(x) - f(a) = f'(\xi)(x - a) \geq f'(a)(x - a).$$

因此曲线位于切线上方.

(iii) \Rightarrow (i) 在区间段 $[a, b]$ 上考虑. 过 $(a, f(a)), (b, f(b))$ 的割线记为 $s(x)$. 点 $c \in (a, b)$ 处的切线记为 $t(x)$. 由条件可知

$$s(a) = f(a) \geq t(a), \quad s(b) = f(b) \geq t(b),$$

即在两个端点处有 $s(x) \geq t(x)$, 从而在 $[a, b]$ 上恒有 $s(x) \geq t(x)$. 特别地, $s(c) \geq t(c) = f(c)$, 即曲线段位于割线下方, 函数是凸函数.

综上, 三者等价. \square

根据上述定理 (iii), 直接有下述结论.

设 f 是 (a, b) 上可微的凸函数, 则 f 的驻点必是极值点, 进而是最小值点.

利用一阶导数判别法, 下面的二阶导数判别法也是显然的.

凸性的二阶导数判定

设 f 在区间 I 上二阶可导, 则 f 是 I 上的凸函数当且仅当 $f'' \geq 0$.

值得指出的是, 充分性也可以通过泰勒公式得到:

$$f(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(\xi)(x - a)^2 \geq f(a) + f'(a)(x - a).$$

例 1.5.1 证明约当 (Jordan) 不等式: $\sin x \geq \frac{2}{\pi}x, x \in [0, \pi/2]$.

证明. 因为 $(\sin x)'' = -\sin x \leq 0$, 所以 $\sin x$ 在 $[0, \pi/2]$ 上凹. 而通过 $(0, \sin 0)$ 和 $(\pi/2, \sin \pi/2)$ 的割线方程为 $y = 2x/\pi$, 从而根据凹函数的定义有 $\sin x \geq 2x/\pi (x \in [0, \pi/2])$. \square

例 1.5.2 证明: $\sin x \leq \frac{1}{2}x(\pi - x), x \in [0, \pi]$.

证明. 令 $f(x) = \sin x - \frac{1}{2}x(\pi - x)$, 则 $f''(x) = -\sin x + 1 \geq 0$, 因此 f 是 $[0, \pi]$ 上的凸函数, 所以 $f(x) \leq \max\{f(0), f(\pi)\} = 0$. \square

例 1.5.3 证明均值不等式.

证明. 因为 $(\ln x)'' < 0$, 所以 $\ln x$ 是凹函数. 由詹森不等式, 当 $a_j > 0$ 时成立

$$\ln\left(\frac{1}{n}a_1 + \cdots + \frac{1}{n}a_n\right) \geq \frac{1}{n}\ln a_1 + \cdots + \frac{1}{n}\ln a_n = \ln \sqrt[n]{a_1 \cdots a_n}.$$

故而几何平均小于等于算数平均. \square

例 1.5.4 (故地重游) 证明 Young 不等式.

证明. 根据 $\ln x$ 是凹性, 当 $a, b > 0$ 时有

$$\ln\left(\frac{1}{p}a^p + \frac{1}{q}b^q\right) \geq \frac{1}{p}\ln a^p + \frac{1}{q}\ln b^q = \ln ab.$$

从而成立 Young 不等式. \square

凸性区间与拐点 (inflection point)

- ▶ 设 f 在 D 上有定义, 区间 $I \subset D$. 若 f 是 I 上的凸/凹函数, 并且 I 不能再扩大, 则称 I 是 f 的一个凸/凹区间.
- ▶ 若 f 在 a 连续, 并且 a 两侧是两个凸性不同的严格凸性区间, 则称 $(a, f(a))$ 为函数图像的拐点, 或称 a 为 f 的拐点.

例 1.5.5 求函数 $f(x) = \sqrt[3]{x}$ 的凸性区间和拐点.

解. 根据 $f''(x) = -\frac{2}{9}x^{-5/3}$, 结合二阶导数判别法, 可知 $(-\infty, 0]$ 是凸区间、 $[0, +\infty)$ 是凹区间, 进而拐点是 $(0, 0)$. \square

当函数的光滑性较好时, 可以用下面的方法来搜索拐点.

拐点判定法

- ▶ 若 f 在 a 二阶可导且 a 是 f 的拐点, 则 $f''(a) = 0$;
- ▶ 若 f 在 a 三阶可导且 $f''(a) = 0, f'''(a) \neq 0$, 则 a 是 f 的拐点.

证明. 若 a 是拐点, 则 f' 在 a 两侧单调性相反, 因此 a 是 f' 的极值点, 从而 $f''(a) = 0$. 反之, 若 $f'''(a) \neq 0$, 则 f'' 在 a 两侧严格异号, 凸性不同, 所以 a 是拐点. \square

例 1.5.6 求函数 $f(x) = \sin x$ 的拐点.

解. 由 $f''(x) = -\sin x = 0$ 得 $x = n\pi (n \in \mathbb{Z})$, 而 $f'''(n\pi) = -\cos n\pi \neq 0$. 根据上述判别法, $x = n\pi (n \in \mathbb{Z})$ 是 f 的拐点. \square

1.6 函数的草图

函数的草图有助于直观理解函数的性质. 通常需要显示出函数的定义域、间断点、不可导点、单调区间、凸性区间、极值点、拐点、渐近线等具有指标意义的特性, 若能表现出奇偶性、周期性、零点、截距自然也是有益的. 我们以下例展示绘制草图的步骤.

例 1.6.1 绘制 $f(x) = \sqrt[3]{x^3 - x^2 - x + 1}$ 的草图.

解. 显然 f 是 \mathbb{R} 上的连续函数, 并且没有奇偶性、周期性.

1. 求导数

$$f'(x) = \frac{(x-1)(3x+1)}{3\sqrt[3]{(x-1)^4(x+1)^2}}, \quad f''(x) = -\frac{8}{9(x-1)^{\frac{4}{3}}(x+1)^{\frac{5}{3}}}.$$

此外, $x = \pm 1$ 是不可导点.

2. 判定单调区间和凸性区间

x	$(-\infty, -1)$	$(-1, -1/3)$	$(-1/3, 1)$	$(1, +\infty)$
$f'(x)$	+	+	-	+
$f''(x)$	+	-	-	-
$f(x)$	凸增	凹增	凹减	凹增

从中可以看出, $x = -1$ 是拐点, $x = -1/3$ 是极大值点, $x = 1$ 是极小值点.

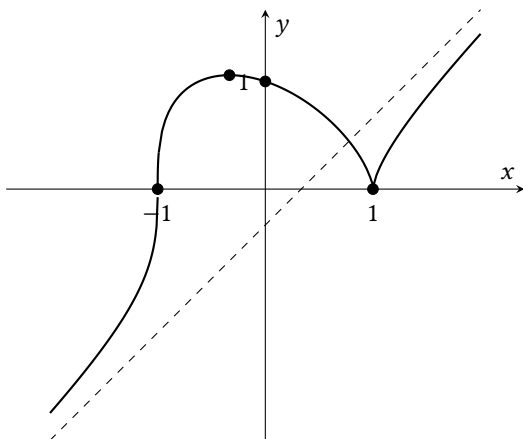
3. 求渐近线

利用 Taylor 公式容易发现, 当 $x \rightarrow \infty$ 时, 有

$$f(x) = x \cdot \sqrt[3]{1 - \frac{1}{x} - \frac{1}{x^2} + \frac{1}{x^3}} = x \left(1 - \frac{1}{3x} + o\left(\frac{1}{x}\right) \right) = x - \frac{1}{3} + o(1).$$

即渐近线为 $y = x - \frac{1}{3}$. 当然也可以用之前的标准方法求.

4. 作草图



□

3 课时/54 课时

1.7 牛顿迭代法

牛顿迭代法是一种快速高效求函数零点的方法，它是应用数学和计算数学最重要的算法之一。有据可查的最早案例是古希腊数学家赫伦在公元 60 年对于开方的算法

$$x_{n+1} = \frac{1}{2} \left(x_n + \frac{N}{x_n} \right).$$

用现代的观点，这可以看做求解函数 $f(x) = x^2 - N$ 的零点的一种方法。它的思想是朴素的：如果 x_n 大于 \sqrt{N} ，那么 N/x_n 小于 \sqrt{N} ，从而两者的平均比 x_n 更接近 \sqrt{N} 。显然，这与微积分并无多大联系，但迭代的思想是深刻的。

牛顿使用了微分的思想来求解多项式的零点。拉夫森 (Raphson) 给出了第一个真正意义上的迭代算法，但也仅限于多项式。1740 年，辛普森 (Thomas Simpson) 得到了现代意义的牛顿迭代法。19 世纪初，傅里叶 (Fourier) 将这一方法命名为牛顿迭代法。

牛顿的想法非常漂亮。假设函数 f 在预知范围内有零点 x_* 。首先猜测 x_0 是 x_* 的一个近似值，记 $x_* = x_0 + \epsilon_0$ ，则有

$$f(x_*) = f(x_0 + \epsilon_0) = f(x_0) + f'(x_0)\epsilon_0 + \frac{f''(\xi_0)}{2}\epsilon_0^2.$$

如果估计尚算准确，那么 ϵ_0 应是很小的数，忽略高阶项可得

$$0 = f(x_*) \approx f(x_0) + f'(x_0)\epsilon_0.$$

因此，在 $f'(x_0) \neq 0$ 的条件下，可以得到误差项的近似值

$$\epsilon_0 \approx -\frac{f(x_0)}{f'(x_0)}.$$

进而可得 x_* 的一个理应更佳的近似

$$x_* \approx x_1 := x_0 - \frac{f(x_0)}{f'(x_0)}.$$

再令 $x_* = x_1 + \epsilon_1$ ，通过同样的方法，可得

$$x_* \approx x_2 := x_1 - \frac{f(x_1)}{f'(x_1)}.$$

如果上述步骤可以一直进行下去，直觉上 $\{x_n\}$ 将收敛于 x_* 。

虽然牛顿并未明确写出迭代公式并且处理的也并非一般的函数，但本质思想已经蕴含其中。因此下述公式被称为**牛顿迭代公式**。

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

牛顿迭代法的本质是将函数近似为线性函数——用切线的零点代替函数的零点. 事实上, 函数 f 的图像在 $(x_n, f(x_n))$ 处的切线是

$$y = f'(x_n)(x - x_n) + f(x_n),$$

它与 x 轴的交点横坐标恰为 $x_n - f(x_n)/f'(x_n)$. 因此牛顿迭代法也称为**牛顿切线法**.

现在我们来考察牛顿法的收敛性. 一般而言, 迭代数列的行为是非常复杂的, 但在单调有界或压缩数列这两种简单情形下, 易得其收敛性.

牛顿迭代法全局收敛性

设 f 在 $[a, b]$ 上二阶连续可导且 $x_* \in (a, b)$ 是 f 的零点. 若 $f' > 0, f'' > 0$, 则以 $x_0 \in (x_*, b)$ 为初值的牛顿迭代数列 $\{x_n\}$ 单调递减收敛于 x_* . 若 $|f'| \geq L > 0, |f''| \leq U$, 则有误差估计

$$|x_n - x_*| \leq \frac{|f(x_n)|}{L}, \quad |x_{n+1} - x_*| \leq \frac{U}{2L} |x_n - x_*|^2.$$

证明. 根据凸性可知 $x_n > x_*$, 再由 $f' > 0$ 可得 $x_{n+1} < x_n$. 因此 $\{x_n\}$ 单调递减有下界, 设其收敛于 x_∞ . 于是, 对递推式求极限可知

$$x_\infty = x_\infty - \frac{f(x_\infty)}{f'(x_\infty)}.$$

因此 $f(x_\infty) = 0$. 再根据零点的唯一性, 知 $x_\infty = x_*$.

当 $|f'| \geq L > 0$ 时, 根据拉格朗日中值定理可得

$$|f(x_n)| = |f(x_n) - f(x_*)| = |f'(\eta_n)(x_n - x_*)| \geq L|x_n - x_*|.$$

从而

$$|x_n - x_*| \leq \frac{|f(x_n)|}{L}.$$

如果还成立 $|f''| \leq U$, 则根据泰勒公式

$$0 = f(x_*) = f(x_n) + f'(x_n)(x_* - x_n) + \frac{1}{2}f''(\xi_n)(x_* - x_n)^2$$

可得

$$|x_{n+1} - x_*| = \frac{|f''(\xi_n)|}{2|f'(x_n)|} (x_n - x_*)^2 \leq \frac{U}{2L} |x_n - x_*|^2.$$

如果记 $\bar{x}_n = \frac{U}{2L} x_n, \bar{\xi} = \frac{U}{2L} \xi$, 那么

$$|\bar{x}_{n+1} - \bar{x}_*| \leq |\bar{x}_n - \bar{x}_*|^2.$$

因此, 基本上可以这样说: 如果第 n 步精确到了小数点后第 K 位, 那么第 $n+1$ 步就会精确到 $2K$ 位. \square

牛顿迭代法局部收敛性

设 f 在 $[a, b]$ 上二阶连续可导且 $x_* \in (a, b)$ 是 f 的零点. 若 $f'(x_*) \neq 0$, 则存在 $\delta > 0$, 以任意 $x_0 \in U^\circ(x_*; \delta)$ 为初值的牛顿迭代数列 $\{x_n\}$ 均

收敛于 x_* , 且有误差估计

$$\lim_{n \rightarrow \infty} \frac{(x_{n+1} - x_*)}{(x_n - x_*)^2} = \frac{f''(x_*)}{2f'(x_*)}.$$

证明. 在 x_* 的某邻域上 f' 非零, 进而可定义函数

$$\varphi(x) = x - \frac{f(x)}{f'(x)}.$$

注意到

$$\varphi(x_*) = x_*, \quad \varphi'(x_*) = \frac{f(x_*)f''(x_*)}{[f'(x_*)]^2} = 0.$$

因此给定 $\theta \in (0, 1)$, 存在 $\delta > 0$, 使得在更小的邻域 $U(x_*, \delta)$ 上 $|\varphi'| \leq \theta$. 根据拉格朗日中值定理, 有

$$\varphi^n(U(x_*, \delta)) \subset U(x_*, \theta^n \delta).$$

进而, 对任意 $x_0 \in U(x_*, \delta)$, 可定义牛顿迭代数列 $x_{n+1} = \varphi(x_n)$, 且

$$x_n \in U(x_*, \theta^n \delta).$$

故而 $\lim_{n \rightarrow \infty} x_n = x_*$.

为了得到误差估计, 再次利用泰勒公式

$$0 = f(x_*) = f(x_n) + f'(x_n)(x_* - x_n) + \frac{1}{2}f''(\xi_n)(x_* - x_n)^2$$

可得

$$\frac{(x_{n+1} - x_*)}{(x_n - x_*)^2} = \frac{f''(\xi_n)}{2f'(x_n)}.$$

根据导数的连续性, 取极限可得结论. \square

值得指出的是, 作为一个算法, 上述迭代公式事实上是无法执行的, 因为很多情况下 f/f' 是超越函数, 无法求得精确值, 因此还需借助其它手段. 牛顿正是借助了泰勒公式, 求出了天文学中开普勒方程 $x - \epsilon \sin x = M$ 的近似解.

近似牛顿法

设 f 在 $[a, b]$ 上 n 阶连续可微, T_n 是 f 在 $c \in (a, b)$ 点的 n 阶泰勒多项式. 若 $|f'| \geq L > 0$, $|f^{(n)}| \leq M_n$, 且 x_* 和 ξ_n 分别是 f 和 T_n 在 $[a, b]$ 上的零点, 则

$$|\xi_n - x_*| \leq \frac{|f(\xi_n)|}{L} \leq \frac{M_n \cdot |\xi_n - c|^n}{L \cdot n!}.$$

证明. 根据拉格朗日余项的泰勒公式, 有

$$|f(x) - T_n(x)| \leq \frac{M_n}{n!} |x - c|^n.$$

由于 $T_n(\xi_n) = 0$, 所以

$$|f(\xi_n)| \leq \frac{M_n}{n!} |\xi_n - c|^n.$$

利用

$$|f(\xi_n)| = |f(\xi_n) - f(x_*)| \geq L|\xi_n - x_*|$$

即得.

□

牛顿法也可以用来处理复变函数, 进而产生美妙的牛顿分形 (Newton fractal), 有兴趣的读者可以查阅相关资料.