

简单粗暴地推导Taylor公式的各种形式

泰勒(Taylor)公式，被某本数分书称作**一元函数微分学的顶峰**，我深以为然。Taylor将光滑函数在参考点用任意阶多项式近似的方案，往小了可以说解决数分/高数中各种求极限作业题，往大了说即便是最前沿的论文，这一函数近似也是家常便饭般在不同场合按需使用着，可以说是行走江湖必备神器。我之所以想到做这个整理，也正是在阅读优化领域的论文时发现自己对向量值函数的Taylor公式及其积分余项还不够熟练，所以干脆就追根溯源，把这个基础工具彻底整理一次。

本文结构：

- 一元函数的Taylor公式
 - 三种形式的Taylor公式
 - 积分余项
 - Lagrange余项（拉格朗日余项）
 - Peano余项（皮亚诺余项）
 - 推导
 - 总结
- 向量函数的Taylor公式
 - 7个Taylor公式
 - 推导
 - 总结
 - 应用：无约束优化的最优性条件

一元函数的Taylor公式

一元函数的Taylor公式大家都很熟悉：

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + R_n(x) \quad (1)$$

其中余项 $R_n(x)$ 有三种形式，分别为**积分余项**

$$R_n(x) = \frac{1}{n!} \int_{x_0}^x f^{(n+1)}(t)(x - t)^n dt \quad (2)$$

Lagrange余项

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}, \quad \xi \in (x_0, x) \quad (3)$$

和**Peano余项**

$$R_n(x) = o((x - x_0)^n) \quad (4)$$

这三种余项由上到下估计精度递减。其中一个小小的细节是Peano余项中没有出现 $f^{(n+1)}(\cdot)$ ，也就不要求 $f(\cdot)$ $n+1$ 阶可导。因为与其他两种余项公式的前提条件不同，因此我们证明的时候Peano余项就需要单独处理。（当然也很简单）

推导

在写这篇文章之前，我特意去学校图书馆放数学分析教材的区域翻了翻，发现积分余项这个东西被提及的次数莫名的少。起码架子上的几本工科数学分析在介绍完Peano余项和Lagrange余项后就都结束了。因此我合理地猜测，广大非数学系的学子们，可能有很多人是不知道Taylor公式的积分余项的。

我的另一点发现是，这些教材证明Lagrange余项的方法非常的**优雅**：先将 $R_n(x)$ 写成适当的形式，再连续使用 $n+1$ 次Cauchy中值定理得到结果。虽然也很简单易懂，但问题是真的**不好想到**，属于那种一看过程就懂，但事后完全想不起来的类型。

因此，针对这两点缺憾，我决心要用最直观的方法去构造证明，要用**最简单粗暴**的思路证明最完善的结论。希望读者们，看完这篇文章，公式记住了，证明看懂了，而且觉得思路顺畅自然，下次不会恍然。咳咳，言归正传，我们开始分析。

首先，比起证明，更重要的是理解Taylor公式本身。它是干什么的？答案言简意赅：它是在**某一个点周围，用 n 次多项式近似一个光滑函数**。如果我们把Taylor公式的多项式部分紧凑地合并，记作 $f_n(x)$ ，那么Taylor公式就是简单的两部分：

$$f(x) = f_n(x) + R_n(x) \quad (5)$$

其中 n 次多项式 $f_n(x)$ 用来近似原函数 $f(x)$ ， $R_n(x)$ 是近似后“剩下的”误差部分，所以叫余项。我们希望 $R_n(x)$ 越接近0越好，这说明我们近似地越精准。

那么在一个点周围，怎么样近似一个函数呢？这个答案也是显而易见的，就是该点的各阶导数值相等——0阶导数就是函数值，0阶导数相等意味着在这一点值一样；1阶导数是0阶导数的变化率，1阶导数相等意味着这一点函数值的变化快慢相等；2阶导数是1阶导数的变化率，2阶导数相等意味着函数1阶导数值的变化快慢相等，也就是这一点函数值变化快慢的变化快慢相等，以此类推……

这些导数值携带着函数的信息，越低阶的导数携带的信息越直接、重要，因此我们的近似思路就是从低到高尽可能让它们相等，至于具体相等到几阶，则根据实际需要的精度决定。这样 $f_n(x)$ 需要满足的条件就呼之欲出了，不多，就两点：

1. $f_n(x)$ 在 $x = x_0$ 点的 $0 \sim n$ 阶导数值与 $f(x)$ 均相等，即
$$f_n^{(k)}(x_0) = f^{(k)}(x_0), \forall k = 0, \dots, n$$
2. $f_n(x)$ 是一个不超过 n 次的多项式，i.e. $f_n^{(k)}(x) \equiv 0, \forall k \geq n+1$

多项式求导大家应该都会吧~ 容易验证公式(1)就是我们需要的 $f_n(x)$ ，各项系数就是 $f^{(k)}(x_0)$ 除去因多项式求导而产生的多余常数倍，可以说是**很简单粗暴**的构造了，完全符合我们的理念！记住了这个，Taylor公式的主体就搞定了——毕竟我们的目的是近似！余项只是完事儿之后附加的结果评估而已~

但作为新时代的学生，我们想必是有追求的。不仅要近似，还要明白我们的近似是否是精准的、合理的，因此要对余项进行分析。那么对 $R_n(x)$ 我们知道什么呢？我们只知道：

$$R_n(x) = f(x) - f_n(x) \quad (6)$$

因此从我们对 $f_n(x)$ 的两个条件，就可以对应于 $R_n(x)$ 的两个条件：

$$1. \quad R_n^{(k)}(x_0) = f^{(k)}(x_0) - f_n^{(k)}(x_0) = 0, \forall k = 0, \dots, n \quad (7.1)$$

$$2. \quad R_n^{(k)}(x) = f^{(k)}(x) - f_n^{(k)}(x) = f^{(k)}(x), \forall k \geq n+1 \quad (7.2)$$

公式(7.2)告诉我们，从 $k = n+1$ 阶导数起， n 次多项式 $f_n(x)$ 的影响就已经完全消失了，即

$$R_n^{(n+1)}(x) \equiv f^{(n+1)}(x)$$

那么自然地，只要反向一路积分回去，我们就可以得到

$$f(x) \text{ 或者 } R_n(x) = \underbrace{\int_{x_0}^x \cdots \int_{x_0}^{t_1}}_{n+1\uparrow} f^{(n+1)}(t_0) \underbrace{dt_0 \cdots dt_n}_{n+1\uparrow} + c_0 + c_1 x + \cdots + c_n x^n \quad (8)$$

其中多项式 $C(x) = c_0 + c_1 x + \cdots + c_n x^n$ 的选取调节(8)式的 $0 \sim n$ 阶导数值，如果选取 $C(x) = f_n(x)$ ，就得到了 $f(x)$ ，如果选取 $C(x) = 0$ ，就得到了满足条件(7.1)和(7.2)的余项 $R_n(x)$ 。

因此我们自然地得到了“简单粗暴型”余项公式

$$R_n(x) = \underbrace{\int_{x_0}^x \int_{x_0}^{t_n} \cdots \int_{x_0}^{t_1}}_{n+1\uparrow} f^{(n+1)}(t_0) \underbrace{dt_0 dt_1 \cdots dt_n}_{n+1\uparrow} \quad (9)$$

某种程度上其实我们已经完成了任务，写出了余项的精准表达式。只不过公式(9)显然不便于使用，所以我们还需要将(9)式化为常见的(2)和(3)式。

化简的方法有2种，一种常见的使用分部积分的方法我觉得思路不够直接，在这里我继续贯彻本文**简单粗暴**的理念，给出另一种证法：**不就是累次积分吗，我一层一层直接算下去不就好了！**从里向外看，注意到被积函数 $f^{(n+1)}(t_0)$ 只和 t_0 有关，容易想到如果将最内侧的两个积分换序，就可以消掉一个积分号。

具体而言，

$$\int_{x_0}^{t_2} \int_{x_0}^{t_1} f^{(n+1)}(t_0) dt_0 dt_1 = \int_{x_0}^{t_2} f^{(n+1)}(t_0) dt_0 \int_{t_0}^{t_2} dt_1 = \int_{x_0}^{t_2} f^{(n+1)}(t_0) (t_2 - t_0) dt_0$$

这时候

$$R_n(x) = \underbrace{\int_{x_0}^x \int_{x_0}^{t_n} \cdots \int_{x_0}^{t_2}}_{n\uparrow} f^{(n+1)}(t_0) (t_2 - t_0) \underbrace{dt_0 dt_2 \cdots dt_n}_{n\uparrow}$$

被积函数中表达式不明的部分 $f^{(n+1)}(t_0)$ 仍然只和 t_0 有关，我们可以继续进行这样的操作。事实上，

$$\begin{aligned} \int_{x_0}^{t_i} \int_{x_0}^{t_{i-1}} f^{(n+1)}(t_0) (t_{i-1} - t_0)^k dt_0 dt_{i-1} &= \int_{x_0}^{t_i} f^{(n+1)}(t_0) dt_0 \int_{t_0}^{t_i} (t_{i-1} - t_0)^k dt_{i-1} \\ &= \frac{1}{k+1} \int_{x_0}^{t_i} f^{(n+1)}(t_0) (t_i - t_0)^{k+1} dt_0 \end{aligned}$$

重复以上计算 n 次，即得

$$\begin{aligned} R_n(x) &= \underbrace{\int_{x_0}^x \cdots \int_{x_0}^{t_3} \int_{x_0}^{t_2}}_{n\uparrow} f^{(n+1)}(t_0) (t_2 - t_0) \underbrace{dt_0 dt_2 \cdots dt_n}_{n\uparrow} \\ &= \frac{1}{n!} \int_{x_0}^x f^{(n+1)}(t_0) (x - t_0)^n dt_0 \quad \square. \end{aligned}$$

即积分余项公式(2)。

由积分余项公式(2)可以直接推出Lagrange余项公式(3)：

$$\begin{aligned}
R_n(x) &= \frac{1}{n!} \int_{x_0}^x f^{(n+1)}(t)(x-t)^n dt \\
&= \frac{1}{n!} f^{(n+1)}(\xi) \int_{x_0}^x (x-t)^n dt \\
&= \frac{1}{n!} f^{(n+1)}(\xi) \frac{1}{n+1} (x-x_0)^{n+1} \\
&= \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-x_0)^{n+1} \quad \square.
\end{aligned}$$

其中第二行使用了积分中值定理。

由Lagrange余项公式(3)可以直接得出比Peano余项(4)更强的结论 $R_n(x) = \mathcal{O}((x-x_0)^{n+1})$, 不过Peano余项在 $f(\cdot)$ 不 $n+1$ 阶可导时也成立, 所以我们不能借助所有出现过 $f^{(n+1)}(\cdot)$ 的已有结论, 而是直接从定义出发证明。

由无穷小的定义, (4)式即

$$\lim_{x \rightarrow x_0} \frac{R_n(x)}{(x-x_0)^n} = 0 \quad (10)$$

由(7.1)知, 上式左侧为 $\frac{0}{0}$ 型的极限, 故用洛必达法则有

$$\lim_{x \rightarrow x_0} \frac{R_n(x)}{(x-x_0)^n} = \lim_{x \rightarrow x_0} \frac{R'_n(x)}{n(x-x_0)^{n-1}}$$

由(7.1)知, 上式仍为 $\frac{0}{0}$ 型的极限。故我们反复使用洛必达法则, n 次后最终得到

$$\lim_{x \rightarrow x_0} \frac{R_n(x)}{(x-x_0)^n} = \lim_{x \rightarrow x_0} \frac{R'_n(x)}{n(x-x_0)^{n-1}} = \dots = \lim_{x \rightarrow x_0} \frac{R_n^{(n)}(x)}{n!} = 0 \quad \square.$$

总结

综上, 一元函数的Taylor公式部分我们就说完了。回顾一下整体的思路:

首先, 我们的目的是要用 n 次多项式 $f_n(x)$ 在 x_0 周围近似函数 $f(x)$;

为了近似 $f(x)$, 我们令 $f_n(x)$ 的 $0 \sim n$ 阶导数与 $f(x)$ 的 $0 \sim n$ 阶导数在 x_0 点均相等, 从而得到了各个系数的表达式;

剩余的部分就是余项 $R_n(x)$, 因为 $f_n(x)$ 从 $n+1$ 阶导数起恒为0, 因此 $R_n^{(n+1)}(x) \equiv f^{(n+1)}(x)$, 我们只需要一路积分回去就可以得出 $R_n(x)$ 的表达式;

利用累次积分 (或其他方法) 不断消去积分号, 就得到了积分余项公式;

而Lagrange余项可以用积分余项直接导出, Peano余项可以用定义+洛必达法则直接得到。

向量函数的Taylor公式

一般意义下的多元函数/向量函数的Taylor公式必然是复杂的, 因为Taylor公式中出现了 $f(x)$ 的 $0 \sim n$ 阶导数。而一个多元函数0阶导数是它本身, 1阶导数是梯度向量, 2阶导数是Hessian矩阵.....这意味着如果要再往后写, 用矩阵的语言都已经无法再简洁描述一个多元函数的高阶导数了, 我们只能用更高级的张量记号或者回到最朴素的各分量多重角标求和。这样的所得到的公式既不便记忆, 也不便使用。

所幸在实际应用中, 对于一个向量函数, 我们几乎不需要去计算它超过2阶的导数。比如在优化中, 如果对一个光滑函数进行近似, 一般二阶近似已经足够了, 大多数都还是一阶方法。因此, 对于向量函数, 我们只写出它2阶连续可导时能够得到的几个Taylor公式, 这些最重要也最常用。

这几个公式如下：

- 0阶带积分余项的泰勒公式

$$f(x) = f(x_0) + \left\langle \int_0^1 \nabla f(x_0 + \theta(x - x_0)) d\theta, x - x_0 \right\rangle \quad (11.1)$$

- 1阶带积分余项的泰勒公式

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \left\langle \int_0^1 \nabla^2 f(x_0 + \theta(x - x_0))(1 - \theta) d\theta (x - x_0), x - x_0 \right\rangle \quad (11.2)$$

- 0阶带Lagrange余项的泰勒公式

$$f(x) = f(x_0) + \langle \nabla f(\xi), x - x_0 \rangle \quad (11.3)$$

- 1阶带Lagrange余项的泰勒公式

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2} \langle \nabla^2 f(\xi)(x - x_0), x - x_0 \rangle \quad (11.4)$$

- 0阶带Peano余项的泰勒公式

$$f(x) = f(x_0) + o(1) \quad (11.5)$$

- 1阶带Peano余项的泰勒公式

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + o(\|x - x_0\|) \quad (11.6)$$

- 2阶带Peano余项的泰勒公式

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2} \langle \nabla^2 f(x_0)(x - x_0), x - x_0 \rangle + o(\|x - x_0\|^2) \quad (11.7)$$

其中 $\langle x, y \rangle$ 表示 x 和 y 的内积，简而言之对两个列向量 $\langle x, y \rangle = x^T y$ ； $\|x\|$ 表示 x 的范数， $\|x\| = \sqrt{\langle x, x \rangle}$ 。

推导

对于向量版本的公式(11.1)-(11.7)，我们一个自然的想法就是要利用标量版本的公式(1)-(4)。毕竟都是Taylor公式，只是对象不同，原理应该类似。

那么如何建立起向量函数和标量函数的桥梁呢？这里就用到了一个常见小技巧，在学习方向导数、优化证明、变分法等多个地方都见到过的一个**非常常用的小技巧**，那就是令

$$g(\alpha) = f(x_0 + \alpha s) \quad (12)$$

其中 x_0 是一个初始点， s 是一个方向，不妨假设 $\|s\| = 1$ 。那么当 $\alpha \rightarrow 0$ 时， $g(\alpha)$ 就表示 $f(x_0)$ 在 s 方向上做微小的变动。这个几何意义使我们可以利用标量函数 g 在向量函数 f 的局部对其进行各种便利的分析和近似处理。而Taylor公式就是局部近似。

那么我们对 $g(\alpha)$ 在0处使用Taylor公式，

$$g(\alpha) = g(0) + g'(0) \alpha + \frac{g''(0)}{2!} \alpha^2 + \dots$$

emmm，看起来需要先解决求导的问题。根据多元函数的求导链式法则，我们不难计算出

$$g'(\alpha) = \frac{\partial}{\partial \alpha} f(x_0 + \alpha s) = \langle \nabla f(x_0 + \alpha s), s \rangle \quad (13.1)$$

$$g''(\alpha) = \frac{\partial}{\partial \alpha} \langle \nabla f(x_0 + \alpha s), s \rangle = \langle \nabla^2 f(x_0 + \alpha s) \cdot s, s \rangle \quad (13.2)$$

特别地，当 $\alpha = 0$ 时

$$g'(0) = \langle \nabla f(x_0 + 0s), s \rangle = \langle \nabla f(x_0), s \rangle \quad (14.1)$$

$$g''(0) = \langle \nabla^2 f(x_0 + 0s) \cdot s, s \rangle = \langle \nabla^2 f(x_0) \cdot s, s \rangle \quad (14.2)$$

完成了准备工作之后，只剩下代入公式了。这时如果使用带积分余项的Taylor公式(1)+(2)，可以得到

- $n = 0$ 时

$$\begin{aligned} f(x_0 + s) &= g(1) = g(0) + R_0(1) \\ &= g(0) + \int_0^1 g'(t) dt \\ &= f(x_0) + \int_0^1 \langle \nabla f(x_0 + ts), s \rangle dt \\ &= f(x_0) + \left\langle \int_0^1 \nabla f(x_0 + ts) dt, s \right\rangle \end{aligned}$$

- $n = 1$ 时

$$\begin{aligned} f(x_0 + s) &= g(1) = g(0) + g'(0) + R_1(1) \\ &= g(0) + g'(0) + \int_0^1 g''(t)(1-t) dt \\ &= f(x_0) + \langle \nabla f(x_0), s \rangle + \int_0^1 \langle \nabla^2 f(x_0 + ts) \cdot s, s \rangle (1-t) dt \\ &= f(x_0) + \langle \nabla f(x_0), s \rangle + \left\langle \int_0^1 \nabla^2 f(x_0 + ts)(1-t) dt \cdot s, s \right\rangle \end{aligned}$$

最后进行变量代 (gai) 换 (ming) $t = \theta, s = x - x_0$ 即可得(11.1)-(11.2).

公式(11.3)-(11.7)的推导模式相同，此处略过。

总结

多元函数部分的要点就2个：

1. 把公式(11.1)-(11.7)记住。由于积分型余项最精准，其实只需要记住(11.1)和(11.2)就可以轻松得到后面5个。—(然而其实这两个是最难的。)—
2. 记住技巧(12)。基本上实际中做展开都是以(12)的形式完成的（参见应用部分的证明），一方面用某个抽象的方向 s 替代 $x - x_0$ ，另一方面可以调节步长 α ，非常常用。

应用：无约束优化的最优性条件

对于优化问题(P)

$$\min_x f(x) \quad (P)$$

我们称 x^* 是(P)的一个**局部极小点**，如果 $\exists x^*$ 的邻域 U , s.t. $\forall x \in U, f(x^*) \leq f(x)$. 若此时还成立 $\forall x \in U - \{x^*\}, f(x^*) < f(x)$, 则称 x^* 是(P)的一个**严格局部极小点**。

通过使用向量函数的Taylor公式，我们可以证明以下3个定理：

定理1.(一阶必要条件) 若 f 连续可导， x^* 是(P)的一个局部极小点，则有

$$\nabla f(x^*) = 0 \quad (15)$$

定理2.(二阶必要条件) 若 f 二阶连续可导， x^* 是(P)的一个局部极小点，则有

$$\begin{aligned} \nabla f(x^*) &= 0 \\ \nabla^2 f(x^*) &\geq 0 \end{aligned} \quad (16)$$

定理3.(二阶充分条件) 若 f 二阶连续可导, 且点 x^* 满足

$$\begin{aligned}\nabla f(x^*) &= 0 \\ \nabla^2 f(x^*) &> 0\end{aligned}\tag{17}$$

则 x^* 是(P)的一个严格局部极小点.

其中矩阵 $A > 0$ 表示 A 正定, $A \geq 0$ 表示 A 半正定

定理1证明: 因为 x^* 是(P)的一个局部极小点, 故

$\exists r > 0, s. t. \forall x \in \{x^* + \alpha s \mid \|s\| = 1, 0 \leq \alpha < r\}$, 有 $f(x^*) \leq f(x)$

由于

$$f(x) = f(x^*) + \langle \nabla f(x^*), \alpha s \rangle + o(\|\alpha s\|) \geq f(x^*)$$

即

$$\alpha \langle \nabla f(x^*), s \rangle + o(\alpha) \geq 0$$

上式分别除以 α 再令 $\alpha \rightarrow 0$, 得

$$\lim_{\alpha \rightarrow 0} \frac{\alpha \langle \nabla f(x^*), s \rangle}{\alpha} = \langle \nabla f(x^*), s \rangle \geq 0, \quad \forall \|s\| = 1\tag{18}$$

在(18)中取 $s = -\frac{\nabla f(x^*)}{\|\nabla f(x^*)\|}$, 得到

$$-\|\nabla f(x^*)\| \geq 0$$

推出

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

定理2证明: 因为 x^* 是(P)的一个局部极小点, 故

$\exists r > 0, s. t. \forall x \in \{x^* + \alpha s \mid \|s\| = 1, 0 \leq \alpha < r\}$, 有 $f(x^*) \leq f(x)$

由于

$$f(x) = f(x^*) + \langle \nabla f(x^*), \alpha s \rangle + \frac{1}{2} \langle \nabla^2 f(x^*) \alpha s, \alpha s \rangle + o(\|\alpha s\|^2) \geq f(x^*)$$

即

$$\alpha \langle \nabla f(x^*), s \rangle + \frac{1}{2} \alpha^2 \langle \nabla^2 f(x^*) s, s \rangle + o(\alpha^2) \geq 0$$

由定理1, $\nabla f(x^*) = 0$, 上式分别除以 α^2 再令 $\alpha \rightarrow 0$, 得

$$\lim_{\alpha \rightarrow 0} \frac{\frac{1}{2} \alpha^2 \langle \nabla^2 f(x^*) s, s \rangle}{\alpha^2} = \frac{1}{2} \langle \nabla^2 f(x^*) s, s \rangle \geq 0, \quad \forall \|s\| = 1\tag{19}$$

(19)即

$$\nabla^2 f(x^*) \geq 0 \quad \square.$$

上面两个证明的思路是极其相似的, 只需要对邻域中的任意一点 x 使用Taylor公式展开. 利用 x^* 极小值点的性质和无穷小的定义直接计算即可. x 展开的阶数不同, 则定理结果就不同. 接下来证明定理3, 方法仍然是类似的, 只不过需要把思路“逆转过来”。

定理3证明: 要证 x^* 是(P)的一个严格局部极小点, 即要

$\exists r > 0, s.t. \forall x \in \{x^* + \alpha s \mid \|s\| = 1, 0 < \alpha < r\},$ 有 $f(x^*) < f(x)$

由于

$$f(x) = f(x^*) + \langle \nabla f(x^*), \alpha s \rangle + \frac{1}{2} \langle \nabla^2 f(x^*) \alpha s, \alpha s \rangle + o(\|\alpha s\|^2)$$

即

$$\begin{aligned} f(x) - f(x^*) &\geq \alpha \langle \nabla f(x^*), s \rangle + \frac{1}{2} \alpha^2 \langle \nabla^2 f(x^*) s, s \rangle + o(\alpha^2) \\ &= \frac{1}{2} \alpha^2 \langle \nabla^2 f(x^*) s, s \rangle + o(\alpha^2) \end{aligned}$$

由于 $\nabla^2 f(x^*)$ 正定, 故 $\nabla^2 f(x^*)$ 的最小特征值 $\lambda_{\min} > 0$, 即

$$\langle \nabla^2 f(x^*) s, s \rangle \geq \lambda_{\min} \|s\|^2 = \lambda_{\min}$$

而 $\frac{o(\alpha^2)}{\alpha^2} = 0$ 可以推出 $\exists r > 0, s.t. \forall 0 < \alpha < r,$ 有 $o(\alpha^2) < \frac{\lambda_{\min}}{4} \alpha^2$, 此时

$$\begin{aligned} f(x) - f(x^*) &\geq \frac{1}{2} \alpha^2 \langle \nabla^2 f(x^*) s, s \rangle + o(\alpha^2) \\ &> \frac{\lambda_{\min}}{2} \alpha^2 - \frac{\lambda_{\min}}{4} \alpha^2 \\ &= \frac{\lambda_{\min}}{4} \alpha^2 \\ &> 0 \end{aligned}$$

□.