第五章 最优化方法（5）
最速下降法、牛顿法、阻尼牛顿法

无约束最优化问题

之前求解一个无约束规划的最优化问题，通常采用 “一阶导数为0 + 二阶充分条件” 来求解。即：

无约束极小化

$$\min f(x), \quad x \in \mathbb{R}^n$$

解析方法的套路是：

1. 求梯度方程：$\nabla f(x) = 0$，解出所有**候选点**（驻点）。
2. 求 Hessian：$\nabla^2 f(x)$，在每个候选点判断是否正定/半正定，从而判断局部极小/鞍点/极大。

这其实是在 **完全解析地** 求解一个非线性方程组 $\nabla f(x) = 0$。
然而这种方法在实际问题中，存在以下问题：

1. 高维、复杂函数时根本解不出 $\nabla f(x) = 0$；
2. 驻点太多，解析地枚举检查不现实；
3. Hessian 计算和正定性判断在高维下代价很高。

这种 **解析法** 往往不适用于实际计算。在实际计算中，往往采用 **迭代法** 逐渐逼近答案，这就是为什么有 **最速下降法 / 牛顿法 / 阻尼牛顿法** 这些迭代方法。

最速下降法

算法思想

- 在当前点 $x^{(k)}$，函数下降最快的方向就是 **负梯度方向** $-\nabla f(x^{(k)})$。
- 然后沿这个方向做一维搜索，找一个“最好”的步长 α_k。

迭代公式

通常给定一个初始的迭代点 $x^{(1)}$。

- **求梯度** $\nabla f(x)$
  对函数求一阶导 / 梯度 $\nabla f(x)$.

- **搜索方向** $d^{(k)}$:
  梯度 $\nabla f(x)$ 代入迭代点 $x^{(k)}$，得到 $\nabla f(x^{(k)})$。
  取 **梯度相反方向** 作为 **搜索方向**:

$$d^{(k)} = -\nabla f(x^{(k)})$$

- **步长** α_k:

$$\alpha_k = \arg\min_{\alpha \geq 0} f\big(x^{(k)} + \alpha d^{(k)}\big)$$

  将 $x^{(k)} + \alpha d^{(k)}$ 代入到最优化问题的函数，然后通过求极小值点得到此步的 α_k。
  具体的:
  - 令 $\varphi(\alpha) = f\big(x^{(k)} + \alpha d^{(k)}\big)$，对 α 求导 $\varphi'(\alpha)$，令 $\varphi'(\alpha) = 0$ 解出 α_k。
  - 很多题是二次函数，α_k 很好算。
- **更新迭代点** $x^{(k+1)}$:

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$$

  将求得的步长 α_k 用于 $x^{(k+1)}$ 的更新。

一般到梯度为 0 的时候停止迭代，$|\nabla f(x^{(k)})| \leq \varepsilon$。

例题

> 已知下列最优化问题和初始点:
>
> $$f(x) = x_1^2 + 3x_2^2, \qquad x^{(1)} = (2, 1)^T$$
>
> 用最速下降法 **迭代一次** 得到的答案。

解答:

- **求梯度**

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)^T = (2x_1, 6x_2)^T.$$

  在初始点 $x^{(1)} = (2, 1)^T$ 处:

$$\nabla f(x^{(1)}) = (4, 6)^T.$$

- **搜索方向**

  最速下降法的方向是负梯度方向:

$$d^{(1)} = -\nabla f(x^{(1)}) = -(4, 6)^T = (-4, -6)^T.$$

- **求步长** α_1

  令一维函数

$$\varphi(\alpha) = f\big(x^{(1)} + \alpha d^{(1)}\big),$$

  先写出沿直线上的点:

$$x^{(1)} + \alpha d^{(1)} = (2, 1)^T + \alpha(-4, -6)^T = (2 - 4\alpha, 1 - 6\alpha)^T.$$

代入最优化问题函数 f:

$$\begin{aligned}\varphi(\alpha) &= (2-4\alpha)^2 + 3(1-6\alpha)^2 \\ &= 124\alpha^2 - 52\alpha + 7.\end{aligned}$$

对 α 求导，令导数为 0:

$$\varphi'(\alpha) = 248\alpha - 52 = 0 \quad \Rightarrow \quad \alpha_1 = \frac{52}{248} = \frac{13}{62}.$$

- **更新迭代点**

$$\begin{aligned}x^{(2)} &= x^{(1)} + \alpha_1 d^{(1)} \\ &= (2,1)^T + \frac{13}{62}(-4,-6)^T \\ &= \left(\frac{36}{31}, -\frac{8}{31}\right)^T\end{aligned}$$

用最速下降法从 $x^{(1)} = (2,1)^T$ 出发迭代一次的结果为

$$\boxed{x^{(2)} = \left(\frac{36}{31}, -\frac{8}{31}\right)^T}$$

牛顿法

算法思想

- 把 $f(x)$ 在当前点做 **二阶泰勒展开**，用一个二次函数近似它。
- 然后直接求这个近似二次函数的极小点，得到一个“牛顿方向”。

迭代公式

通常给定一个初始的迭代点 $x^{(1)}$。

- **求梯度 和 Hessian** $\nabla f(x)$，$\nabla^2 f(x)$

  求函数一阶导得到 $\nabla f(x)$;

  求函数二阶导得到 $\nabla^2 f(x)$.

- **牛顿方向** $d^{(k)}$

  将当前迭代点代入 梯度 和 Hessian，得到 $\nabla f(x^{(k)})$，$\nabla^2 f(x^{(k)})$

  计算当前 **牛顿方向** 作为搜索方向:

$$d^{(k)} = -[\nabla^2 f(x^{(k)})]^{-1} \nabla f(x^{(k)})$$

- 牛顿法中的步长 $\alpha = 1$（步长固定为 1）
- **更新迭代点** $x^{(k+1)}$:

$$x^{(k+1)} = x^{(k)} + d^{(k)}$$

一般到梯度 为 0 的时候停止迭代，$||\nabla f(x^{(k)})|| \leq \varepsilon$。

例题

> 已知下列最优化问题和初始点：
>
> $$f(x) = x_1^2 + 3x_2^2, \qquad x^{(1)} = (2,1)^T$$
>
> 用牛顿法求解答案。

解答：

- **求梯度 $\nabla f(x)$ 和 Hessian $\nabla^2 f(x)$**

  梯度：

  $$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)^T = (2x_1, 6x_2)^T$$

  Hessian：

  $$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}$$

- **代入当前迭代点 $x^{(1)}$**

  $$\nabla f(x^{(1)}) = (4,6)^T$$

  Hessian 不依赖点，因此仍为：

  $$\nabla^2 f(x^{(1)}) = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}$$

- **计算牛顿方向 $d^{(1)}$**

  $$d^{(1)} = -[\nabla^2 f(x^{(1)})]^{-1} \nabla f(x^{(1)})$$

  求 Hessian 的逆矩阵：

  $$[\nabla^2 f(x^{(1)})]^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/6 \end{pmatrix}$$

  计算方向：

  $$d^{(1)} = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/6 \end{pmatrix} \begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

- **更新迭代点**

  $$x^{(2)} = x^{(1)} + d^{(1)} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

此时：

$$\nabla f(x^{(2)}) = (0,0)^T \Rightarrow ||\nabla f(x^{(2)})|| = 0 \le \varepsilon$$

停止迭代，经过一次牛顿迭代法得到解：

$$\boxed{x^{(2)} = (0,0)^T}$$

阻尼牛顿法

算法思想

牛顿法虽然具有很快的收敛速度，但是当：

- Hessian 不是正定矩阵；
- 迭代步长过大导致更新点偏离下降方向；
- 初始点距离极小点较远时，

牛顿法有可能 **震荡、发散或下降幅度不稳定。**

为了解决这一问题，对牛顿方向再进行 **线性搜索（Line Search）**，通过求步长 (\alpha_k)，让更新方向更加稳健——这就是 **阻尼牛顿法**（Damped Newton Method）。

迭代公式

通常给定一个初始迭代点 $x^{(1)}$。

- **求梯度 $\nabla f(x)$ 和 Hessian $\nabla^2 f(x)$**

  求函数一阶导得到梯度；
  求二阶导得到 Hessian。

- **牛顿方向 $d^{(k)}$**

  将当前迭代点代入求得：

  $$d^{(k)} = -[\nabla^2 f(x^{(k)})]^{-1} \nabla f(x^{(k)})$$

- **精确线搜索求步长 α_k**

  阻尼牛顿法不像普通牛顿法固定 $\alpha = 1$，而是通过线搜索确定步长：

  $$\alpha_k = \arg\min_{\alpha \geq 0} f(x^{(k)} + \alpha d^{(k)})$$

  确定步长做法：

  1. 构造单变量函数 $\varphi(\alpha) = f\big(x^{(k)} + \alpha d^{(k)}\big)$；
  2. 对 α 求导，解出最优步长 α_k。

- **更新迭代点**

  $$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$$

  停止准则一般为 $|\nabla f(x^{(k)})| \leq \varepsilon$.

例题

> 已知下列最优化问题和初始点：
>
> $$f(x) = x_1^2 + 3x_2^2, \qquad x^{(1)} = (2, 1)^T$$
>
> 用**阻尼牛顿法迭代两次。**

解答：

第一次迭代

- **梯度与 Hessian**

$$\nabla f(x) = (2x_1, 6x_2)^T$$

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}$$

- **代入 $x^{(1)} = (2,1)^T$**

$$\nabla f(x^{(1)}) = (4,6)^T$$

- **牛顿方向 $d^{(1)}$**

$$d^{(1)} = -[\nabla^2 f(x^{(1)})]^{-1} \nabla f(x^{(1)}) = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/6 \end{pmatrix} \begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

- **求步长 α_1**

  构造

$$\varphi(\alpha) = f(x^{(1)} + \alpha d^{(1)}) = f(2-2\alpha, 1-\alpha)$$

  代入：

$$\varphi(\alpha) = (2-2\alpha)^2 + 3(1-\alpha)^2 = 16\alpha^2 - 14\alpha + 7$$

  求导并令其为零：

$$\varphi'(\alpha) = 32\alpha - 14 = 0 \Rightarrow \alpha_1 = \frac{14}{32} = \frac{7}{16}$$

- **更新得到 $x^{(2)}$**

$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = (2,1)^T + \frac{7}{16}(-2,-1)^T = \left(\frac{9}{8}, \frac{9}{16}\right)^T$$

第二次迭代

- **在 $x^{(2)}$ 处的梯度**

$$\nabla f(x^{(2)}) = \left(2 \cdot \tfrac{9}{8}, 6 \cdot \tfrac{9}{16}\right)^T = \left(\frac{9}{4}, \frac{27}{8}\right)^T$$

  Hessian 仍为 $\nabla^2 f(x^{(2)} = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}$。

- **牛顿方向 $d^{(2)}$**

$$d^{(2)} = -[\nabla^2 f(x^{(2)})]^{-1} \nabla f(x^{(2)}) = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/6 \end{pmatrix} \begin{pmatrix} 9/4 \\ 27/8 \end{pmatrix} = \begin{pmatrix} -9/8 \\ -9/16 \end{pmatrix}$$

  注意到这里 $d^{(2)} = -x^{(2)}$。

- **求 α_2**

  令

$$x^{(2)} + \alpha d^{(2)} = x^{(2)} + \alpha(-x^{(2)}) = (1-\alpha)x^{(2)}$$

  则

$$\varphi(\alpha) = f(x^{(2)} + \alpha d^{(2)}) = f\big((1-\alpha)x^{(2)}\big) = (1-\alpha)^2 f(x^{(2)})$$

显然当 $(1-\alpha)^2$ 最小即 $\alpha = 1$ 时，$\varphi(\alpha)$ 取得极小：

$$\alpha_2 = 1$$

- **更新得到** $x^{(3)}$

$$x^{(3)} = x^{(2)} + \alpha_2 d^{(2)} = x^{(2)} + 1 \cdot (-x^{(2)}) = (0,0)^T$$

此时：

$$\nabla f(x^{(2)}) = (0,0)^T \Rightarrow ||\nabla f(x^{(2)})|| = 0 \leq \varepsilon$$

达到极小点。用阻尼牛顿法从 $x^{(1)} = (2,1)^T$ 出发**迭代两次**，得到解：

$$\boxed{x^{(3)} = (0,0)^T}$$
