第十六讲：优化算法

寻找优化问题的非解析解

杨 林

大 纲

1. 无约束优化算法

1.1. 下降法

1.2. 牛顿法

2. 等式约束优化算法

大 纲

1. 无约束优化算法

1.1. 下降法

1.2. 牛顿法

2. 等式约束优化算法

1 无约束优化算法

$$\text{最小化} \qquad f(x)$$

1. f 凸, 二次连续可微（因此 $\mathbf{dom}\, f$ 是开集）
2. 我们假设最优值 $p^* = \inf_x f(x)$ 可达到（并且有限）

■ **迭代方法**

1. 生成点序列 $x^{(k)} \in \mathbf{dom}\, f, k = 0, 1, \cdots$, 其中
$$f\left(x^{(k)}\right) \to p^*$$
2. 可以解释为求解最优性条件的迭代方法
$$\nabla f(x^*) = 0$$

1 无约束优化算法

■ 下降方法

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)} \text{ 其中 } f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$$

1. 其他符号: $x^{+} = x + t\Delta x$, $x^{-} := x - t\Delta x$
2. Δx 是步径或搜索方向; t 是步进或步长
3. 从凸性来看, $f(x^{+}) < f(x)$ 意味着 $\nabla f(x)^{T}\Delta x < 0$ (即, Δx 是下降方向) ($f(x^{+}) - f(x) \geq \nabla f(x)^{T} t\Delta x$)

General descent method.

given a starting point $x \in \mathbf{dom}\, f$.
repeat
  1. Determine a descent direction Δx.
  2. *Line search.* Choose a step size $t > 0$.
  3. *Update.* $x := x + t\Delta x$.
until stopping criterion is satisfied.

---

1 无约束优化算法

■ **直线搜索方法**

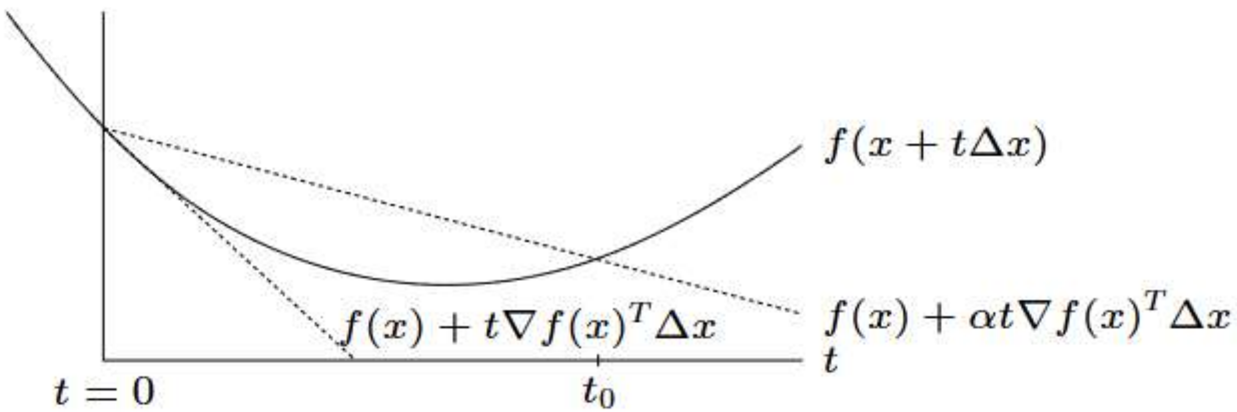
□ **精确直线搜索**: $t = \arg\min_{t>0} f(x + t\Delta x)$

□ **回溯直线搜索**（参数 $\alpha \in (0,1/2), \beta \in (0,1)$）

1. 从 $t = 1$ 开始,重复 $t := \beta t$ 直到

$$f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$$

2. 图形解释: 回溯直到 $t \le t_0$



1 无约束优化算法

□梯度下降方法

---

given a starting point $x \in \mathbf{dom}\, f$.
repeat

1. $\Delta x := -\nabla f(x)$.
2. *Line search.* Choose step size t via exact or backtracking line search.
3. *Update.* $x := x + t\Delta x$.

until stopping criterion is satisfied.

---

1. 停止准则通常为 $\|\nabla f(x)\|_2 \leq \varepsilon$
2. 收敛结果:对于强凸函数 f,

$$f\left(x^{(k)}\right) - p^* \leq c^k \left(f\left(x^{(0)} - p\right)\right)$$

$c \in (0,1)$ 取决于 $x^{(0)}$ 和搜索类型

3. 非常简单,但通常非常慢;在实践中很少使用

1 无约束优化算法

□ $\mathbb{R}^2$ 中的二次问题

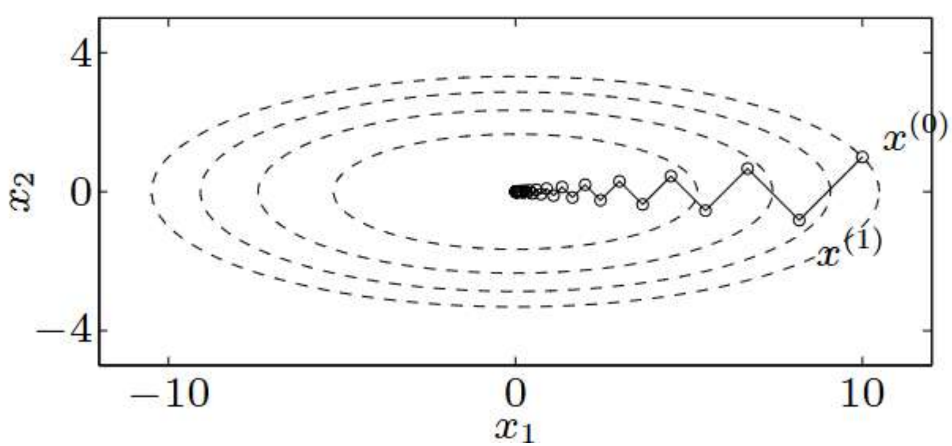
$$f(x) = (1/2)(x_1^2 + \gamma x_2^2)\ (\gamma > 0)$$

使用精确直线搜索,从 $x^{(0)} = (\gamma, 1)$ 开始:

$$x_1^{(k)} = \gamma\left(\frac{\gamma - 1}{\gamma + 1}\right)^k, x_2^{(k)} = \left(-\frac{\gamma - 1}{\gamma + 1}\right)^k$$

如果 $\gamma \gg 1$ 或 $\gamma \ll 1$, 则非常慢

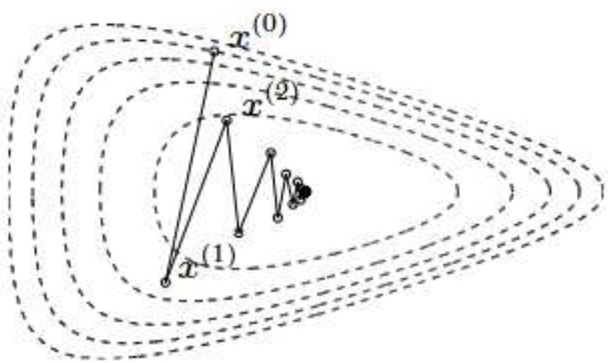
$\gamma = 10$ 的示例:



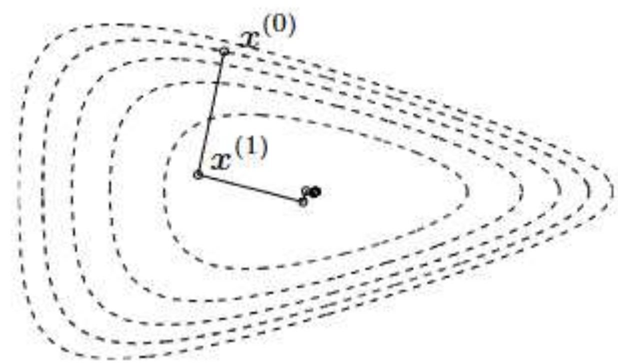
1 无约束优化算法

□ 非二次示例

$$f(x_1, x_2) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$$



回溯线搜索



精确线搜索

搜索类型严重影响算法收敛速度！

1 无约束优化算法

- **规范化最速下降方向（在 x 处，对于范数 $\|\cdot\|$）：**

$$\Delta x_{\mathrm{nsd}} = \arg\min\{\nabla f(x)^T v \mid \|v\| = 1\}$$

解释：对于小的 v，$f(x+v) \approx f(x) + \nabla f(x)^T v$

方向 Δx 是具有最负方向导数的单位范数步长

- **（未归一化）最速下降方向：**

$$\Delta x_{\mathrm{sd}} = \left\|\nabla f(x)^T\right\|_* \Delta x_{\mathrm{nsd}}$$

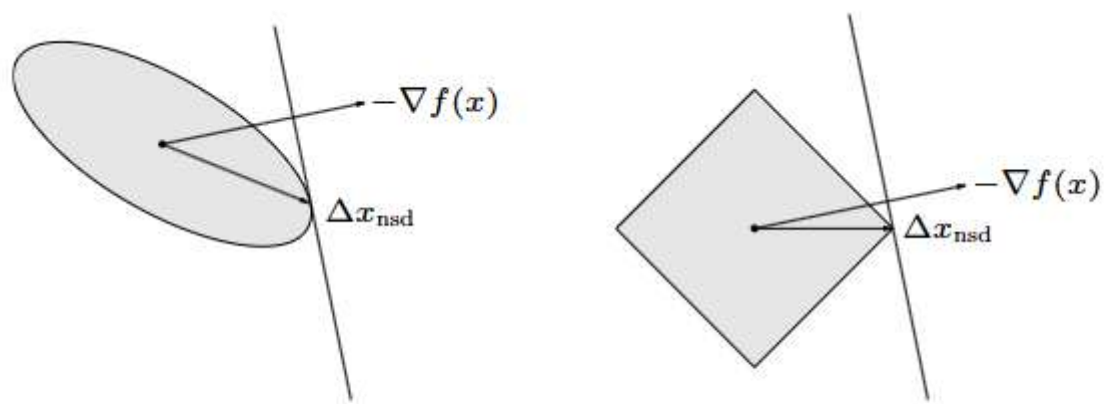
满足$\nabla f(x)^T \Delta x_{\mathrm{sd}} = -\left\|\nabla f(x)^T\right\|_*^2$

1 无约束优化算法

■ **例** 1:

1. 欧几里得范数: $\Delta x = -\nabla f(x)$
2. 二次范数: $\|x\|_P = \left(x^T P x\right)^{1/2}$ $(P \in S_{++}^n)$: $\Delta x_{\text{sd}} = -P^{-1}\nabla f(x)$
3. l_1-范数: $\Delta x_{\text{sd}} = -\left(\frac{\partial f(x)}{\partial x_i}\right)e_i$, 其中 $\left|\frac{\partial f(x)}{\partial x_i}\right| = \|\nabla f(x)\|_\infty$

□ 二次范数和 l_1-范数的单位球与归一化最速下降方向:

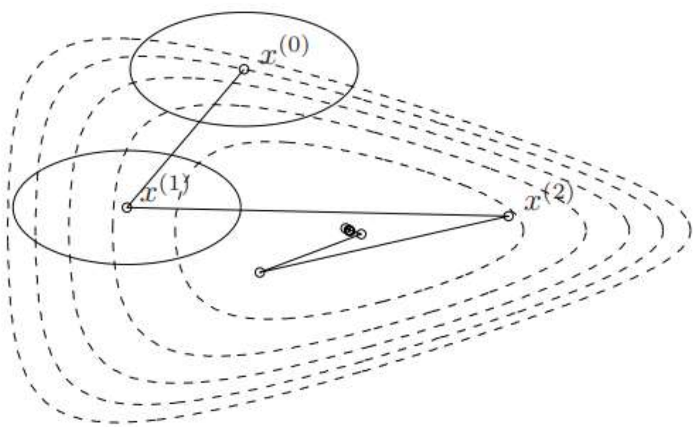


1 无约束优化算法

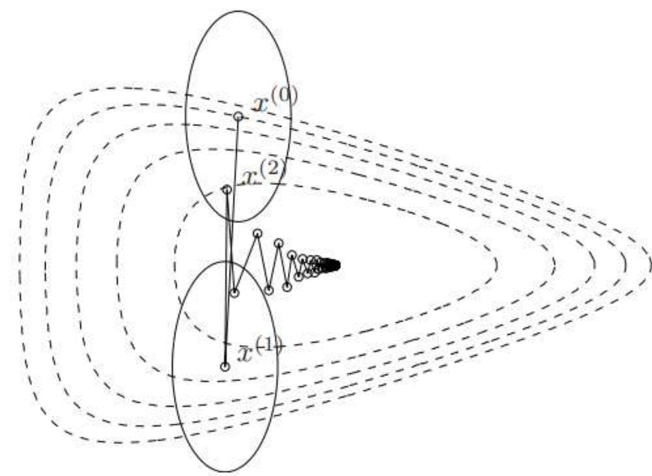
非二次示例（$\alpha = 0.1, \beta = 0.7$）

$$f(x_1, x_2) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$$

应用二次范数：$P_1 = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}, P_2 = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}$



采用二次范数 $\|\cdot\|_{P_1}$ 的最速下降法.
椭圆表示球体 $\left\{x \mid \|x - x^{(k)}\|_{P_1} \le 1\right\}$
在 $x^{(0)}$ 和 $x^{(1)}$ 处的边界



采用二次范数 $\|\cdot\|_{P_2}$ 的最速下降法.
椭圆表示球体 $\left\{x \mid \|x - x^{(k)}\|_{P_2} \le 1\right\}$
在 $x^{(0)}$ 和 $x^{(1)}$ 处的边界

1 无约束优化算法

■ 牛顿法

牛顿步径: $\Delta x_{\mathrm{nt}} = -\nabla^2 f(x)^{-1}\nabla f(x)$

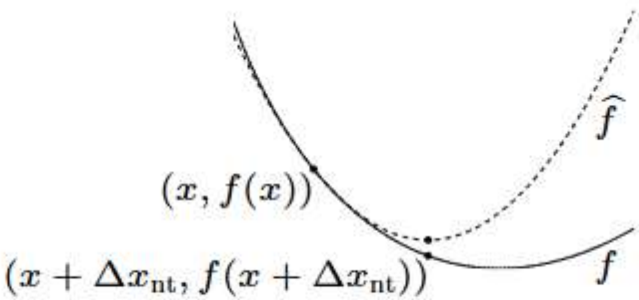
□ 解释

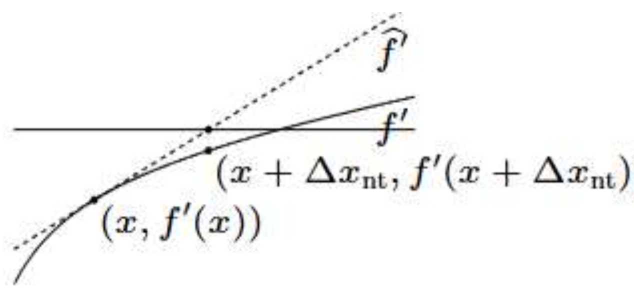
1. $x + \Delta x_{\mathrm{nt}}$ 最大限度地减少二阶近似

$$\hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$$

2. $x + \Delta x_{\mathrm{nt}}$ 极小化二阶近似

$$\nabla f(x+v) \approx \nabla \hat{f}(x+v) = \nabla f(x) + \nabla^2 f(x) v = 0$$

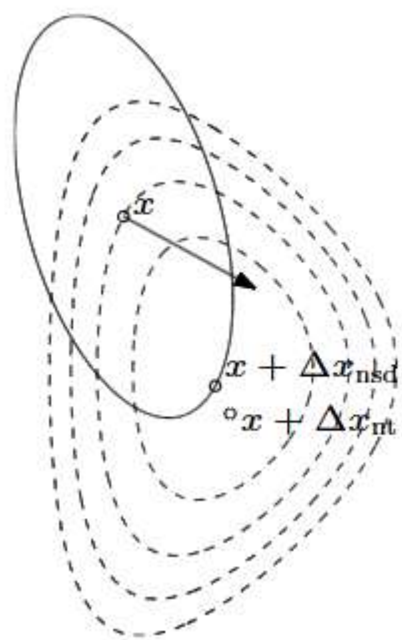




1 无约束优化算法

Δx_{nt} 是 x 处沿局部 Hessian 范数的最速下降方向

$$\|u\|_{\nabla^2 f(x)} = \left(u^T \nabla^2 f(x) u\right)^{1/2}$$



虚线是函数 f 的等高线;椭圆是 $\{x + v \mid v^T \nabla^2 f(x) v = 1\}$

大 纲

1. 无约束优化算法

1.1. 下降法

1.2. 牛顿法

2. 等式约束优化算法

2 等式约束优化算法*

$$
\begin{array}{ll}
\text{最大化} & f(x) \\
\text{约束条件} & Ax = b
\end{array}
$$

1. f 凸, 两次连续可微 (因此 $\mathbf{dom}\, f$ 是开集)
2. $A \in \mathbb{R}^{p \times n}$, $\mathbf{rank}\, A = p < n$
3. 我们假设存在一个最优解 x^* 最优值, 并用 $p^* = \inf_x \{f(x) \mid Ax = b\} = f(x^*)$

由凸优化问题的最有条件与凸问题KKT条件知: $x^* \in \mathbf{dom}\, f$ 是上述优化问题的最优解的充要条件是, 存在 $v^* \in \mathbb{R}^p$ 满足

$$
Ax^* = b, \qquad \nabla f(x^*) + A^T v^* = 0
$$

第一组方程称为**原可行方程**, 第二组方程称为**对偶可行方程**

2 等式约束优化算法*

■ **方法一：消除等式约束**

我们首先确顶矩阵 $F \in \mathbb{R}^{n\times(n-p)}$ 和向量 $\hat{x} \in \mathbb{R}^n$, 以参数化可行集

$$\{x \mid Ax = b\} = \left\{Fz + \hat{x} \mid z \in \mathbb{R}^{(n-p)}\right\}$$

这里 $\hat{x}$ 可以选用 $Ax = b$ 的任何特殊解, 矩阵 F 满足 $\mathcal{R}(F) = \mathcal{N}(A)$, 其中 $\mathcal{R}(F) = \left\{Fz \mid z \in \mathbb{R}^{(n-p)}\right\}, \mathcal{N}(A) = \{x \mid Ax = 0\}$, 分别称为 F 的**值域(像)**和 A 的**零空间(核)、**

消除等式约束后的优化问题为

$$\text{最小化} \qquad \tilde{f}(z) = f(Fz + \hat{x})$$

利用它的解 z^* 可以缺点等式约束问题的解 $x^* = Fz^* + \hat{x}$

- 关于 F 的选择(仿射不变性):矩阵 F 满足 $\mathcal{R}(F) = \mathcal{N}(A)$, 并且 $T \in \mathbb{R}^{(n-p)\times(n-p)}$ 是非奇异的, 则 $\tilde{F} = FT$ 也是合适的, 即对 z 做坐标变换 $z' = Tz$

2 等式约束优化算法*

■ 方法二：用对偶方法求解

优化问题的对偶函数为

$$g(\nu) = -b^T\nu + \inf_x\left(f(x) + \nu^T Ax\right) = -b^T\nu - f^*\left(-A^T\nu\right)$$

这里 f^* 是 f 的共轭，因此对偶问题为

$$\text{最大化} \qquad -b^T\nu - f^*\left(-A^T\nu\right)$$

既然根据定义存在最优解，该问题是严格可行的，因此 Slater 条件成立，即强对偶性成立，即存在 ν^* 满足 $g(\nu^*) = p^*$

2 等式约束优化算法*

■ 等式约束的牛顿法

为了导出等式约束问题在可行点 x 处的 Newton 方向 Δx_{nt}，我们将目标函数换成其在 x 附件的二阶 Taylor 近似

最小化 $\hat{f}(x+v)=f(x)+\nabla f(x)^T v+(1/2)v^T\nabla^2 f(x)v$

约束条件 $A(x+v)=b$

该问题的变量为 v. 由原可行方程和对偶可行方程得, Newton 方向 Δx_{nt} 由以下方程决定

$$\begin{bmatrix}\nabla^2 f(x) & A^T \\ A & 0\end{bmatrix}\begin{bmatrix}\Delta x_{nt} \\ \omega\end{bmatrix}=\begin{bmatrix}-\nabla f(x) \\ 0\end{bmatrix}$$

其中 ω 是该二次问题的最优对偶变量， Newton方向 Δx_{nt} 只在**矩阵** $\begin{bmatrix}\nabla^2 f(x) & A^T \\ A & 0\end{bmatrix}$ **非奇异**的点有定义.

2 等式约束优化算法*

■ 等式约束的牛顿法

我们将等式约束的 Newton 减量定义为

$$\lambda(x) = \left(\Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}}\right)^{1/2}$$

这和之前的无约束条件 Newton 法完全一样，我们可以进行同样的解释. 我们可以设计出算法:

等式约束优化的 Newton 方法

给定 满足 $Ax = b$ 的初始点 $x \in \mathbf{dom}\ f$,误差阈值 $\epsilon > 0$

重复进行

1. 计算 Newton 方向和 Newton 减量 $\Delta x_{nt}, \lambda(x)$
2. **停止准则.** 如果$\lambda^2/2 \le \epsilon$ 则**退出**
3. **直线搜索.** 通过回溯直线搜索确定步长 t
4. **修改.** $x \coloneqq x + t\Delta x_{nt}$

谢 谢 ！