负梯度方向优化方法总结

Atri

日期：May 26, 2023

1 无约束条件优化

求解一般形式的无约束条件优化问题：

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

使用迭代的方式进行搜索：

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k, \alpha_k>0$$

因此可以将问题大致分为两类：寻找步长 α_k 以及下降方向 $\boldsymbol{d}_k$.

1.1 梯度相关

最速下降方向即负梯度方向：

$$\boldsymbol{d}=-\nabla f(\boldsymbol{x}).$$

因此问题集中于步长的确定.

1.1.1 精确线搜索

构造函数 $h(\alpha)=f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)$. 若保证为下降方向，一种很自然的想法是令梯度为 0：

$$h^{'}(\alpha_k)=\nabla f(\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k)^\top\boldsymbol{d}_k=\nabla f(\boldsymbol{x}_{k+1})^\top\boldsymbol{d}_k=0.$$

因此可得该条件的几何意义：下一个迭代点 $\boldsymbol{x}_{k+1}$ 处的梯度与当前下降方向 $\boldsymbol{d}_k$ 正交.
考虑正定二次函数 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^\top Q\boldsymbol{x}+\boldsymbol{b}^\top\boldsymbol{x}+c$, 其在点 $\boldsymbol{x}_k$ 处沿 $\boldsymbol{d}_k$ 方向精确线搜索得到的步长为：

$$\alpha_k=-\frac{\boldsymbol{d}_k^\top\nabla f(\boldsymbol{x}_k)}{\boldsymbol{d}_k^\top Q\boldsymbol{d}_k}.$$

该结论需牢记，后面共轭梯度法会再用. 推导过程：
构造函数 $h(\alpha)=f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)=\frac{1}{2}(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)^\top Q(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)+\boldsymbol{b}^\top(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)+c$. 继续推导:

$$\begin{aligned}h(\alpha)=f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)&=\frac{1}{2}(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)^\top Q(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)+\boldsymbol{b}^\top(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)+c\\&=\frac{1}{2}\alpha^2\boldsymbol{d}_k^\top Q\boldsymbol{d}_k+\alpha\boldsymbol{d}_k^\top(Q\boldsymbol{x}_k+\boldsymbol{b})+f(\boldsymbol{x}_k)\\&=\frac{1}{2}\alpha^2\boldsymbol{d}_k^\top Q\boldsymbol{d}_k+\alpha\boldsymbol{d}_k^\top\nabla f(\boldsymbol{x}_k)+f(\boldsymbol{x}_k).\end{aligned}$$

这是个相对于 α 的正定二次函数, 可根据公式直接得出:

$$\alpha_k=-\frac{\boldsymbol{d}_k^\top\nabla f(\boldsymbol{x}_k)}{\boldsymbol{d}_k^\top Q\boldsymbol{d}_k}.$$

可以证明在正定二次函数下函数值为线性收敛：

$$\frac{f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*)}{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)} \le \left(1 - \frac{2}{1+\kappa}\right)^2.$$

证明过程：(为方便, 令 $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$).

使用负梯度为下降方向, 迭代公式:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{g}_k^\top Q \boldsymbol{g}_k} \boldsymbol{g}_k.$$

迭代点对应函数值为:

$$\begin{aligned} f(\boldsymbol{x}_{k+1}) &= f(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) \\ &= \frac{1}{2}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k)^\top Q (\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) + \boldsymbol{b}^\top (\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) + c \\ &= \frac{1}{2}\boldsymbol{x}_k^\top Q \boldsymbol{x}_k + \boldsymbol{b}^\top \boldsymbol{x}_k + c - \alpha_k \boldsymbol{g}_k^\top Q \boldsymbol{x}_k - \alpha_k \boldsymbol{b}^\top \boldsymbol{g}_k + \frac{1}{2}\alpha^2 \boldsymbol{g}_k^\top Q \boldsymbol{g}_k \\ &= f(\boldsymbol{x}_k) - \alpha_k \boldsymbol{g}_k^\top \boldsymbol{g}_k + \frac{1}{2}\alpha^2 \boldsymbol{g}_k^\top Q \boldsymbol{g}_k. \end{aligned}$$

代入步长, 化简得:

$$f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k) - \frac{1}{2}\frac{(\boldsymbol{g}_k^\top \boldsymbol{g}_k)^2}{\boldsymbol{g}_k^\top Q \boldsymbol{g}_k}.$$

根据 $g(\boldsymbol{x}) = Q\boldsymbol{x} + \boldsymbol{b} = 0$ 得 $\boldsymbol{x}^* = -Q^{-1}\boldsymbol{b}$ 代入得 $f(\boldsymbol{x}^*) = -\frac{1}{2}\boldsymbol{b}^\top Q^{-1}\boldsymbol{b} + c$. 因此

$$\begin{aligned} \frac{f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*)}{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)} &= \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) - \frac{1}{2}\frac{(\boldsymbol{g}_k^\top \boldsymbol{g}_k)^2}{\boldsymbol{g}_k^\top Q \boldsymbol{g}_k}}{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)} \\ &= 1 - \frac{\frac{1}{2}\frac{(\boldsymbol{g}_k^\top \boldsymbol{g}_k)^2}{\boldsymbol{g}_k^\top Q \boldsymbol{g}_k}}{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)} \\ &= 1 - \frac{\frac{1}{2}\frac{(\boldsymbol{g}_k^\top \boldsymbol{g}_k)^2}{\boldsymbol{g}_k^\top Q \boldsymbol{g}_k}}{\frac{1}{2}\boldsymbol{x}_k^\top Q \boldsymbol{x}_k + \boldsymbol{b}^\top \boldsymbol{x}_k + \frac{1}{2}\boldsymbol{b}^\top Q^{-1}\boldsymbol{b}} \\ &= 1 - \frac{\frac{(\boldsymbol{g}_k^\top \boldsymbol{g}_k)^2}{\boldsymbol{g}_k^\top Q \boldsymbol{g}_k}}{\boldsymbol{x}_k^\top Q \boldsymbol{x}_k + 2\boldsymbol{b}^\top \boldsymbol{x}_k + \boldsymbol{b}^\top Q^{-1}\boldsymbol{b}}. \end{aligned}$$

由于

$$\boldsymbol{x}_k^\top Q \boldsymbol{x}_k + 2\boldsymbol{b}^\top \boldsymbol{x}_k + \boldsymbol{b}^\top Q^{-1}\boldsymbol{b} = (Q\boldsymbol{x}_k + \boldsymbol{b})^\top Q^{-1}(Q\boldsymbol{x}_k + \boldsymbol{b}).$$

得

$$\begin{aligned} \frac{f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*)}{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)} &= 1 - \frac{\frac{(\boldsymbol{g}_k^\top \boldsymbol{g}_k)^2}{\boldsymbol{g}_k^\top Q \boldsymbol{g}_k}}{(Q\boldsymbol{x}_k + \boldsymbol{b})^\top Q^{-1}(Q\boldsymbol{x}_k + \boldsymbol{b})} \\ &= 1 - \frac{(\boldsymbol{g}_k^\top \boldsymbol{g}_k)^2}{(\boldsymbol{g}_k^\top Q \boldsymbol{g}_k)(\boldsymbol{g}_k^\top Q^{-1} \boldsymbol{g}_k)}. \end{aligned}$$

定理 1.1. 设 Q 为正定矩阵, 特征值为 $\lambda_1 \ge \cdots \ge \lambda_n > 0$, 则

$$\frac{(\boldsymbol{x}^\top \boldsymbol{x})^2}{(\boldsymbol{x}^\top Q \boldsymbol{x})(\boldsymbol{x}^\top Q^{-1} \boldsymbol{x})} \ge \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}.$$

将该式子代入上式, 得

$$\begin{aligned} \frac{f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*)}{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)} &= 1 - \frac{(\boldsymbol{g}_k^\top \boldsymbol{g}_k)^2}{(\boldsymbol{g}_k^\top Q \boldsymbol{g}_k)(\boldsymbol{g}_k^\top Q^{-1} \boldsymbol{g}_k)} \\ &\le 1 - \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2} \\ &\le \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2. \end{aligned}$$

设 $\kappa = \frac{\lambda_1}{\lambda_n}$ 为 Q 的条件数, 则

$$\frac{f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*)}{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)} \leq \left(1 - \frac{2}{1+\kappa}\right)^2.$$

1.1.2 固定步长的基本分析

可以发现，如果使用精确线搜索确定步长，那么程序的效率是没办法得到保证的，因此我们考虑直接提前确定固定的步长.

对于满足不同的强弱条件的函数，我们需要确定的最优固定步长尚有区别. 对于梯度下降法迭代式 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha \nabla f(\boldsymbol{x}_k)$, 考虑当前梯度与最优点的线性关系：

$$\nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x}_k - \boldsymbol{x}^*) = \frac{1}{\alpha}(\boldsymbol{x}_k - \boldsymbol{x}_{k+1})^\top (\boldsymbol{x}_k - \boldsymbol{x}^*)$$

根据基本恒等式 (or 余弦定理)：$\boldsymbol{u}^\top \boldsymbol{v} = \frac{1}{2}(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - \|\boldsymbol{u} - \boldsymbol{v}\|^2)$, 计算上式：

$$\begin{aligned}
\nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x}_k - \boldsymbol{x}^*) &= \frac{1}{\alpha}(\boldsymbol{x}_k - \boldsymbol{x}_{k+1})^\top (\boldsymbol{x}_k - \boldsymbol{x}^*) \\
&= \frac{1}{2\alpha}(\|\boldsymbol{x}_k - \boldsymbol{x}_{k+1}\|^2 + \|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2) \\
&= \frac{1}{2\alpha}(\alpha^2 \|\nabla f(\boldsymbol{x}_k)\|^2 + \|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2) \\
&= \frac{\alpha}{2}\|\nabla f(\boldsymbol{x}_k)\|^2 + \frac{1}{2\alpha}(\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2)
\end{aligned}$$

根据右边括号内的形式，不难想到把该式对 k 从 0 到 $t-1$ 求和：

$$\begin{aligned}
\sum_{k=0}^{t-1} \nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x}_k - \boldsymbol{x}^*) &= \frac{\alpha}{2}\sum_{k=0}^{t-1}\|\nabla f(\boldsymbol{x}_k)\|^2 + \frac{1}{2\alpha}(\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}_t - \boldsymbol{x}^*\|^2) \\
&\leq \frac{\alpha}{2}\sum_{k=0}^{t-1}\|\nabla f(\boldsymbol{x}_k)\|^2 + \frac{1}{2\alpha}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2
\end{aligned}$$

分析到此为止，得出结论：该式上界只与 $\sum_{k=0}^{t-1}\|\nabla f(\boldsymbol{x}_k)\|^2$ 有关，加上常项.

在这里，我们讨论的三个特殊函数皆为凸函数. 假设引入凸函数的性质，例如其一阶性质：

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leq \nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x}_k - \boldsymbol{x}^*)$$

求和，且与上述基本分析结合，可得：

$$\sum_{k=0}^{t-1}(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)) \leq \frac{\alpha}{2}\sum_{k=0}^{t-1}\|\nabla f(\boldsymbol{x}_k)\|^2 + \frac{1}{2\alpha}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2.$$

1.1.3 L 连续 (梯度有界) 凸函数の固定步长

定义 1.1. 函数称为 L-连续当且仅当

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|.$$

若函数是凸的，则可以得出梯度有上界：$\|\nabla f(\boldsymbol{x})\| \leq L$.

因此在基本分析的基础上代入上式，有如下推导：

$$\sum_{k=0}^{t-1}(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)) \leq \frac{\alpha}{2}L^2 t + \frac{R^2}{2\alpha}.$$

根据函数的凸性：

$$f\left(\frac{1}{t}\sum_{k=0}^{t-1}\boldsymbol{x}_k\right) \leq \frac{1}{t}\sum_{k=0}^{t-1} f(\boldsymbol{x}_k).$$

因此得到：

$$
\begin{aligned}
f\left(\frac{1}{t}\sum_{k=0}^{t-1}\boldsymbol{x}_k\right) - f(\boldsymbol{x}^*) &\le \frac{1}{t}\sum_{k=0}^{t-1}(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)) \\
&\le \frac{\alpha}{2}L^2 + \frac{R^2}{2\alpha t} \\
&\le \frac{RL}{\sqrt{t}}.
\end{aligned}
$$

因此在这个条件下，目标函数值下降幅度为 $O(\frac{1}{\sqrt{t}})$.

1.1.4 L 光滑 (二次上界) 凸函数の固定步长

老规矩，先丢出定义。不给定义的话连怎么和基本分析结合都不知道

定义 1.2. 设 f 是连续可微函数, f 称为 L-光滑的, 如果 f 满足

$$
f(\boldsymbol{y}) \le f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}) + \frac{L}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2.
$$

其一个比较重要的充要条件是, f 的梯度是 Lipschitz 的：

$$
\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \le L\|\boldsymbol{x}-\boldsymbol{y}\|.
$$

证明. 由中值定理 (或许不是中值定理？)

$$
\begin{aligned}
f(\boldsymbol{x}+\boldsymbol{p}) &= f(\boldsymbol{x}) + \int_0^1 \nabla f(\boldsymbol{x}+t\boldsymbol{p})^\top \boldsymbol{p}\mathrm{d}t \\
&= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top \boldsymbol{p} + \int_0^1 (\nabla f(\boldsymbol{x}+t\boldsymbol{p}) - \nabla f(\boldsymbol{x}))^\top \boldsymbol{p}\mathrm{d}t
\end{aligned}
$$

令 $\boldsymbol{p} = \boldsymbol{y}-\boldsymbol{x}$, 稍微转换上式，继续推导：

$$
\begin{aligned}
&|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x})| \\
=&\left|\int_0^1 (\nabla f(\boldsymbol{x}+t(\boldsymbol{y}-\boldsymbol{x})) - \nabla f(\boldsymbol{x}))^\top(\boldsymbol{y}-\boldsymbol{x})\mathrm{d}t\right| \\
\le&\int_0^1 \|\nabla f(\boldsymbol{x}+t(\boldsymbol{y}-\boldsymbol{x})) - \nabla f(\boldsymbol{x})\|\|\boldsymbol{y}-\boldsymbol{x}\|\mathrm{d}t \\
\le&\int_0^1 Lt\|\boldsymbol{y}-\boldsymbol{x}\|^2\mathrm{d}t \\
=&\frac{L}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2
\end{aligned}
$$

因此可得定义. □

接下来我们先考虑单步下降，令固定步长 $\alpha = \frac{1}{L}$, 得到迭代式：

$$
\boldsymbol{x}_{k+1} - \boldsymbol{x}_k = -\frac{1}{L}\nabla f(\boldsymbol{x}_k)
$$

根据光滑性，进行推导：

$$
\begin{aligned}
f(\boldsymbol{x}_{k+1}) &\le f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^\top(\boldsymbol{x}_{k+1}-\boldsymbol{x}_k) + \frac{L}{2}\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|^2 \\
&= f(\boldsymbol{x}_k) - \frac{1}{L}\|\nabla f(\boldsymbol{x}_k)\|^2 + \frac{1}{2L}\|\nabla f(\boldsymbol{x}_k)\|^2 \\
&= f(\boldsymbol{x}_k) - \frac{1}{2L}\|\nabla f(\boldsymbol{x}_k)\|^2
\end{aligned}
$$

换种形式：
$$\frac{1}{2L}\|\nabla f(\boldsymbol{x}_k)\|^2 \le f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1})$$
将两边进行累加求和：
$$\frac{1}{2L}\sum_{k=0}^{t-1}\|\nabla f(\boldsymbol{x}_k)\|^2 \le f(\boldsymbol{x}_0) - f(\boldsymbol{x}_t)$$
代入基本分析：
$$\sum_{k=0}^{t-1}(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)) \le (f(\boldsymbol{x}_0) - f(\boldsymbol{x}_t)) + \frac{L}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2$$
移项：
$$\sum_{k=0}^{t}(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)) \le \frac{L}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2$$
不难得出：
$$f(\boldsymbol{x}_t) - f(\boldsymbol{x}^*) \le \frac{L}{2t}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2.$$
因此，收敛幅度为 $O(\frac{1}{t})$.

1.1.5 强凸 (二次下界) 函数の固定步长

定义 1.3. 令 f 为可微函数, 如果存在 $\mu > 0$, 使得
$$f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top(\boldsymbol{y} - \boldsymbol{x}) + \frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2.$$
则 f 为强凸函数.

其充要条件：$h(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{\mu}{2}\|\boldsymbol{x}\|^2$ 为凸函数.
完全可以对比一下 L-光滑函数, 在定义上把 $\ge$ 改成了 $\le$:
$$f(\boldsymbol{y}) \le f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) + \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2.$$
其凸函数的充要条件把二次项挪到了前面：
$$h(\boldsymbol{x}) = \frac{L}{2}\|\boldsymbol{x}\|^2 - f(\boldsymbol{x}).$$
直接将定义和基本分析结合一块：
$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \le \frac{1}{2\alpha}(\alpha^2\|\nabla f(\boldsymbol{x}_k)\|^2 + \|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2) - \frac{\mu}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2.$$
整理可得
$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2 \le 2\alpha(f(\boldsymbol{x}^*) - f(\boldsymbol{x}_k)) + \alpha^2\|\nabla f(\boldsymbol{x}_k)\|^2 + (1 - \mu\alpha)\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2.$$
可以认为, 每次迭代, $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2$ 都缩减了一个常数, 及噪声项.
下面直接给出定理:

定理 1.2. *设 $f : \mathbb{R}^n \to \mathbb{R}$ 为可微函数, 全局极小点为 $\boldsymbol{x}^*$, 并且 f 是 L-光滑的和 μ-强凸的. 选择 $\alpha = \frac{1}{L}$, 则对任意给定初始点 $\boldsymbol{x}_0$, 梯度下降法满足:*
$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2 \le \left(1 - \frac{\mu}{L}\right)\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2$$
$$f(\boldsymbol{x}_t) - f(\boldsymbol{x}^*) \le \frac{L}{2}\left(1 - \frac{\mu}{L}\right)^t\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2.$$
即, 距离平方线性收敛, 目标函数误差对 t 指数下降.

令 $R=\|\boldsymbol{x}_0-\boldsymbol{x}^*\|$, 要搜索到满足 $f(\boldsymbol{x}_t)-f(\boldsymbol{x}^*)<\epsilon$ 的解, 可以令

$$\frac{L}{2}\left(1-\frac{\mu}{L}\right)^t\|\boldsymbol{x}_0-\boldsymbol{x}^*\|^2<\epsilon,$$

由此可得

$$t>\frac{L}{\mu}\log\left(\frac{R^2L}{2\epsilon}\right).$$

因此, 只需 $O(\log\frac{1}{\epsilon})$ 次迭代.

1.1.6 二次函数上的固定步长

正定二次函数为光滑和强凸函数, 因此令步长为 $\alpha=\frac{1}{L}=\frac{1}{\lambda_1}$, 其收敛性为

$$f(\boldsymbol{x}_{t+1})-f(\boldsymbol{x}^*)\leq\left(1-\frac{\mu}{L}\right)f(\boldsymbol{x}_t)-f(\boldsymbol{x}^*)=\left(1-\frac{1}{\kappa}\right)f(\boldsymbol{x}_t)-f(\boldsymbol{x}^*).$$

故与精确线搜索一样, 固定步长的梯度法在二次函数上依然能够实现线性收敛.

1.1.7 总结

梯度法在

- 凸 + L-连续上收敛性: $O(\frac{1}{\sqrt{t}})$.
- 凸 + L-光滑上收敛性: $O(\frac{1}{t})$.
- L-光滑 + μ-强凸: $O(\exp(-t))$.

最大单调步长: $f(\boldsymbol{x})$ 是 L-光滑的, 若固定步长 $\alpha<\frac{2}{L}$. 则能保证 $f(\boldsymbol{x}_{k+1})<f(\boldsymbol{x}_k)$.