优化方法作业

计试 61 张🄵 2140506063

Week 4

9.25 周二

作业 1

证明非规范化最速下降方法对于精确直线搜索的收敛性

证明. 因为 $d_{k,sd} = ||\bigtriangledown f(x_k)||_* d_{k,nsd}$, 所以

$$||d_{k,sd}|| = ||\bigtriangledown f(x_k)|| \tag{1}$$

另外一个常用的结论是

$$\bigtriangledown f(x)^T d_{sd} = -||\bigtriangledown f(x)||_*^2 \tag{2}$$

此时, $\forall k \geq 0$,

$$\begin{aligned} f(x_{k+1}) &= \min_{t\geq 0}\{f(x_k + td_{k,sd})\} \\ &\leq \min_{t\geq 0}\{f(x_k) + \bigtriangledown f(x_k)^T td_{k,sd} + \frac{M}{2}||td_{k,sd}||_2^2\} \\ &\quad (\bigtriangledown^2 f(x_k) < MI) \\ &= f(x_k) + \min_{t\geq 0}\{-t * ||\bigtriangledown f(x_k)||_*^2 + t^2 * \frac{M}{2\gamma^2}||\bigtriangledown f(x_k)||_*^2\} \\ &\quad (Equation\ 2, Equation\ 1, ||d_{sd}||_2^2 \leq \frac{1}{\gamma^2}||d_{sd}||^2 = \frac{1}{\gamma^2}||\bigtriangledown f(x_k)||_*^2) \end{aligned} \tag{3}$$

因此

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq \min_{t\geq 0}\{-t * ||\bigtriangledown f(x_k)||_*^2 + t^2 * \frac{M}{2\gamma^2}||\bigtriangledown f(x_k)||_*^2\} \\ &= -\frac{\gamma^2}{2M}||\bigtriangledown f(x_k)||_*^2 \\ &\leq -\frac{1}{2M}||\bigtriangledown f(x_k)||^2 \end{aligned} \tag{4}$$

考虑函数值和最优值的偏差,

$$
\begin{aligned}
f(x_{k+1}) - p^* &= f(x_{k+1}) - f(x_k) + f(x_k) - p^* \\
&\leq f(x_k) - p^* - \frac{1}{2M}|| \bigtriangledown f(x_k)||_2^2 \\
&\quad (Equation\ 4) \\
&\leq f(x_k) - p^* - \frac{m}{M}(f(x_k) - p^*) \\
&\quad (f(x) - p^* \leq \frac{1}{2m}|| \bigtriangledown f(x)||^2) \\
&= (1 - \frac{m}{M})(f(x_k) - p^*)
\end{aligned} \tag{5}
$$

通过累乘可知

$$
f(x_k) - p^* \leq (1 - \frac{m}{M})^k (f(x_0) - p^*) \tag{6}
$$

固定 $x_0, 0 < m \leq M$, 有

$$
\lim_{k\to\infty} f(x_k) - p^* \leq (f(x_0) - p^*) \lim_{k\to\infty} (1 - \frac{m}{M})^k = 0
$$

注意到 $\forall k \geq 0, f(x_k) - p^* \geq 0$, 可知 $\lim_{k\to\infty} f(x_k) - p^* = 0$. 因此, 在非规范化最速下降中应用精确直线搜索, 最终能收敛到最优解.

□

证明非规范化最速下降方法对于回溯直线搜索的收敛性

证明. $\forall k \geq 0, 0 \leq t_k \leq \frac{\gamma^2}{M}$,

$$
\begin{aligned}
f(x_{k+1}) &\leq f(x_k) + \bigtriangledown f(x_k)^T t_k d_{k,sd} + \frac{M}{2}||t_k d_{k,sd}||_2^2 \\
&\quad (\bigtriangledown^2 f(x_k) < MI) \\
&\leq f(x_k) - t_k|| \bigtriangledown f(x_k)||_*^2 + \frac{M t_k^2}{2\gamma^2}||d_{k,sd}||_*^2 \\
&\quad (Equation\ 2, ||x||_* \geq \gamma||x||_2) \\
&= f(x_k) + (\frac{M t_k}{2\gamma^2} - 1) t_k|| \bigtriangledown f(x_k)||_*^2 \\
&\leq f(x_k) - \frac{1}{2} t_k|| \bigtriangledown f(x_k)||_*^2 \qquad (0 \leq t_k \leq \frac{\gamma^2}{M}) \\
&\leq f(x_k) - \alpha t_k|| \bigtriangledown f(x_k)||_*^2 \qquad (0 < \alpha < 0.5) \\
&= f(x_k) + \alpha f(x_k)^T t_k d_{k,sd} \\
&\quad (Equation\ 2)
\end{aligned} \tag{7}
$$

所以, $\forall 0 \le t_k \le \frac{\gamma^2}{M}$, 回溯停止条件成立. 因此, 回溯停止时的步长 $t \ge min\{1, \frac{\beta\gamma^2}{M}\}$, 即

$$\forall k \ge 0, t_k \ge min\{1, \frac{\beta\gamma^2}{M}\} \tag{8}$$

$$\begin{aligned} f(x_{k+1}) &\le f(x_k) + \alpha * t_k * \bigtriangledown f(x_k)^T d_{k,sd} \\ &\le f(x_k) + \alpha * min\{1, \frac{\beta\gamma^2}{M}\} * (-|| \bigtriangledown f(x_k)||_*^2) \\ &\quad (Equation\ 8, Equation\ 2) \\ &\le f(x_k) - \alpha\gamma^2 min\{1, \frac{\beta\gamma^2}{M}\} || \bigtriangledown f(x_k)||_2^2 \end{aligned} \tag{9}$$

因此

$$f(x_k) - f(x_{k+1}) \ge \alpha\gamma^2 min\{1, \frac{\beta\gamma^2}{M}\} || \bigtriangledown f(x_k)||_2^2 \tag{10}$$

Equation 10和Equation 4在形式上是类似的. 因此, 可以导出类似的结果:

$$f(x_{k+1}) - p^* \le (1 - 2m\alpha\gamma^2 min\{1, \frac{\beta\gamma^2}{M}\})(f(x_k) - p^*) \tag{11}$$

记 $c = 2m\alpha\gamma^2 min\{1, \frac{\beta\gamma^2}{M}\}$.

(这里 c 的记号好像有一点小错误，不过不影响结果)

此时, 亦有

$$f(x_k) - p^* \le (1 - c)^k (f(x_0) - p^*) \tag{12}$$

因为 $0 < 1-c < 1$, 故亦有 $\lim\limits_{k\to\infty} f(x_k) - p^* \le (f(x_0) - p^*) \lim\limits_{k\to\infty}(1-c)^k = 0$, $\lim\limits_{k\to\infty} f(x_k) - p^* = 0$. 因此, 在非规范化最速下降中应用回溯直线搜索, 最终能收敛到最优解.

$\square$

9.29 周六

作业 1

证明牛顿下降方法对于回溯直线搜索的收敛性

证明. 固定 α, β, 并取 $\eta = min\{1, 3(1-2\alpha)\}\frac{m^2}{L}$, $0 < \eta \le \frac{m^2}{L}$.

$\forall k \geq 0, 0 \leq t_k \leq \frac{m}{M}$,

$$
\begin{aligned}
f(x_{k+1}) \leq\ & f(x_k) + \bigtriangledown f(x_k)^T t_k d_{k,nt} + \frac{M}{2}||t_k d_{k,nt}||_2^2 \\
& (\bigtriangledown^2 f(x_k) < MI) \\
& = f(x_k) - t_k\lambda_k^2 + \frac{Mt_k^2}{2m} * d_{k,nt}^T mI d_{k,nt} \\
& (\bigtriangledown f(x_k)^T t_k d_{k,nt} = -\lambda_k^2) \\
& \leq f(x_k) - t_k\lambda_k^2 + \frac{Mt_k^2}{2m} * d_{k,nt}^T \bigtriangledown^2 f(x_k) d_{k,nt} \\
& (\bigtriangledown^2 f(x_k) > mI) \\
& = f(x_k) + \lambda_k^2(-t_k + \frac{Mt_k^2}{2m}) \qquad (\bigtriangledown f(x_k)^T t_k d_{k,nt} = -\lambda_k^2) \\
& \leq f(x_k) - \frac{1}{2}t_k\lambda_k^2 \qquad (0 \leq t_k \leq \frac{m}{M}) \\
& = f(x_k) - \alpha t_k \lambda_k^2 \qquad (0 < \alpha < 0.5)
\end{aligned}
\tag{13}
$$

所以, $\forall 0 \leq t_k \leq \frac{m}{M} \leq 1$, 回溯停止条件成立, 即

$$
\forall k \geq 0, t_k \geq \frac{\beta m}{M} \tag{14}
$$

在阻尼牛顿阶段中, 有 $|| \bigtriangledown f(x_k)||_2 \geq \eta$. 此时有

$$
\begin{aligned}
f(x_{k+1}) - f(x_k) \leq\ & -\alpha t_k \lambda_k^2 \\
& \leq -\alpha\frac{\beta m}{M} \bigtriangledown f(x_k)^T \bigtriangledown^2 f(x_k)^{-1} \bigtriangledown f(x_k) \\
& (Equation\ 14, \lambda_k^2 = \bigtriangledown f(x_k)^T \bigtriangledown^2 f(x_k)^{-1} \bigtriangledown f(x_k)) \\
& \leq -\alpha\frac{\beta m}{M} \bigtriangledown f(x_k)^T \frac{I}{M} \bigtriangledown f(x_k) \qquad (\bigtriangledown^2 f(x_k) \leq MI) \\
& = -\alpha\frac{\beta m}{M^2}|| \bigtriangledown f(x_k)||_2^2 \\
& \leq -\alpha\frac{\beta m}{M^2}\eta^2 \qquad (|| \bigtriangledown f(x_k)||_2 \geq \eta)
\end{aligned}
\tag{15}
$$

记 $\gamma = \alpha\frac{\beta m}{M^2}\eta^2 > 0$, 有 $f(x_k) - f(x_{k+1}) \geq \gamma$, 即阻尼牛顿阶段每一步下降的量不小于一个正数.

在二次收敛阶段中, 我们进一步明确该阶段内步长的实际取值, 同时也将解释 η 的取法. 此时 $|| \bigtriangledown f(x_k)||_2 < \eta$. 固定 x_k, $d_{k,nt}$, 记 $\tilde{f}(t) = f(x_k + td_{k,nt})$, $t \geq 0$, 则 $\tilde{f}'(t) = \bigtriangledown^T f(x_k +$

$td_{k,nt})d_{k,nt}$, $\tilde{f}''(t) = d_{k,nt}^T \bigtriangledown^2 f(x_k + td_{k,nt})d_{k,nt}$. 以下, Equation 16对 $\tilde{f}''(t)$ 进行估计.

$$\begin{aligned}
|| \bigtriangledown^2 f(x_k + td_{k,nt}) - \bigtriangledown^2 f(x_k)||_2 &\leq L||td_{k,nt}||_2 \\
|d_{k,nt}^T(\bigtriangledown^2 f(x_k + td_{k,nt}) - \bigtriangledown^2 f(x_k))d_{k,nt}| &\leq L||d_{k,nt}^T||_2||td_{k,nt}||_2||d_{k,nt}||_2 \\
|\tilde{f}''(t) - f''(t)| &\leq Lt||d_{k,nt}||_2^3 \\
\tilde{f}''(t) &\leq f''(t) + Lt||d_{k,nt}||_2^3 \\
\tilde{f}''(t) &\leq \lambda_k^2 + Lt\lambda_k^3 m^{-3/2} \qquad (Equation\ 17)
\end{aligned} \tag{16}$$

其中

$$\begin{aligned}
d_{k,nt}^T \bigtriangledown^2 f(x_k)d_{k,nt} &= \lambda_k^2 \\
d_{k,nt}^T mI d_{k,nt} &\leq \lambda_k^2 \\
m||d_{k,nt}||_2^2 &\leq \lambda_k^2 \\
||d_{k,nt}||_2 &\leq \lambda_k m^{-1/2}
\end{aligned} \tag{17}$$

对Equation 16两侧积分两次

$$\begin{aligned}
\tilde{f}''(t) &\leq \lambda_k^2 + Lt\lambda_k^3 m^{-3/2} \\
\int_0^t \tilde{f}''(t)dt &\leq \int_0^t (\lambda_k^2 + Lt\lambda_k^3 m^{-3/2})dt \\
\tilde{f}'(t) - \tilde{f}'(0) &\leq t\lambda_k^2 + \frac{Lt^2\lambda_k^3}{2m^{3/2}} \\
\tilde{f}'(t) &\leq -\lambda_k^2 + t\lambda_k^2 + \frac{Lt^2\lambda_k^3}{2m^{3/2}} \\
\int_0^t \tilde{f}'(t) &\leq \int_0^t (-\lambda_k^2 + t\lambda_k^2 + \frac{Lt^2\lambda_k^3}{2m^{3/2}})dt \\
\tilde{f}(t) - \tilde{f}(0) &\leq -t\lambda_k^2 + \frac{t^2}{2}\lambda_k^2 + \frac{Lt^3\lambda_k^3}{6m^{3/2}} \\
\tilde{f}(t) &\leq f(x_k) - t\lambda_k^2 + \frac{t^2}{2}\lambda_k^2 + \frac{Lt^3\lambda_k^3}{6m^{3/2}}
\end{aligned} \tag{18}$$

特殊地, 在Equation 18中取 $t = 1$, 有

$$\begin{aligned}
f(x_k + d_{k,nt}) &\leq f(x_k) + \lambda_k^2(-\frac{1}{2} + \frac{L\lambda_k}{6m^{3/2}}) \\
f(x_k + d_{k,nt}) &\leq f(x_k) - \alpha\lambda_k^2 \qquad (Equation\ 20)
\end{aligned} \tag{19}$$

其中

$$\begin{aligned}\lambda_k^2 &= \bigtriangledown f(x_k)^T \bigtriangledown^2 f(x_k)^{-1} \bigtriangledown f(x_k) \\ &\leq \bigtriangledown f(x_k)^T \frac{I}{m} \bigtriangledown f(x_k) \\ &= \frac{1}{m} || \bigtriangledown f(x_k)||_2^2 \\ &\leq \frac{\eta^2}{m} \leq 3(1-2\alpha)\frac{m^{3/2}}{L} \\ that\ makes &-\frac{1}{2} + \frac{L\lambda_k}{6m^{3/2}} \leq -\alpha\end{aligned} \tag{20}$$

Equation 20说明, 在二次收敛阶段，回溯直线搜索的步长取 1 时即满足停止条件, 所以每一步的步长必定为 1. 此时

$$\begin{aligned}\bigtriangledown f(x_{k+1}) &= \bigtriangledown f(x_k + d_{k,nt}) - \bigtriangledown f(x_k) - \bigtriangledown^2 f(x_k) d_{k,nt} \\ &= \int_0^1 (\bigtriangledown^2 f(x_k + t d_{k,nt}) d_{k,nt}) dt - \int_0^1 (\bigtriangledown^2 f(x_k) d_{k,nt}) dt \\ &= \int_0^1 (\bigtriangledown^2 (f(x_k + t d_{k,nt}) - \bigtriangledown^2 f(x_k)) d_{k,nt}) dt \\ &\leq \int_0^1 (L||t d_{k,nt}||_2 d_{k,nt}) dt \\ &= L||d_{k,nt}||_2 d_{k,nt} \int_0^1 t dt \\ &= \frac{L}{2} ||d_{k,nt}||_2 d_{k,nt} \\ || \bigtriangledown f(x_{k+1})||_2 &\leq \frac{L}{2} ||d_{k,nt}||_2^2 \\ &\leq \frac{L}{2m^2} || \bigtriangledown f(x_k)||_2^2 \qquad (Similar\ to\ Equation\ 17)\end{aligned} \tag{21}$$

Equation 21说明了牛顿下降过程的二次收敛性.

根据Equation 21和Equation 15, $\forall \epsilon > 0$, 在不超过 $k = k_1 + k_2 = \frac{f(x_0) - p^*}{\gamma} + log_2(log_2(\frac{2m^2}{L\epsilon}))$ 的步数后, $|| \bigtriangledown f(x_k)|| < \epsilon$. 因此, 在牛顿下降中应用回溯直线搜索, 最终能收敛到最优解.

□