最优化理论与方法第一次上机作业

郑灵超[*] 1601110040

2017 年 11 月 15 日

1 上机作业介绍

本次上机作业的任务是编写带步长 Newton 方法，稳定 Newton 方法和 More-Sorensen 方法的代码，并在几个测试函数上检验他们的性质。

2 算法介绍

2.1 带步长 Newton 方法

带步长的 Newton 方法是在传统的 Newton 方法的基础上，不再限制迭代步长为 1，而利用线搜索来寻找合适的步长，其算法为：

Algorithm 1 带步长 Newton 方法

Input:

函数信息，包括函数值 f，梯度值 g 和 Hesse 矩阵 G;　初始点 x_0;

Output:

最小值点 x，函数值 $f(x)$, 迭代次数 iter，函数调用次数 feva;

Procedure:

Step 1: 初始设置 $k = 0$, $\varepsilon_f > 0$;

Step 2: 计算 x_k 处的函数值，梯度和 Hesse 矩阵 f_k, g_k, G_k;

Step 3: 若 $k > 0$ 且 $|f_k - f_{k-1}| < \varepsilon_f$，循环终止，并输出 x_k, f_k;

Step 4: 计算 Newton 方法的下降方向 d_k: $G_k d_k = -g_k$;

Step 5: 利用线搜索计算步长 α_k;

Step 6: 计算 $x_{k+1} = x_k + \alpha_k d_k$;

Step 7: 令 $k = k + 1$, 回到 **Step 2**.

---

[*]Email: lczheng@pku.edu.cn

带步长 Newton 方法在传统的 Newton 方法的基础上引入了线搜索的步长，得到了比传统 Newton 方法更广泛的应用。但当 Hesse 矩阵不正定时，可能会出现下降方向无法计算的情况。

2.2 稳定 Newton 方法

稳定 Newton 方法利用了修正 Cholesky 分解 [1]，使得计算所得的下降方向始终满足 $g^T d < 0$，其算法[2]为：

Algorithm 2 稳定 Newton 方法

Input:

函数信息，包括函数值 f，梯度值 g 和 Hesse 矩阵 G;　初始点 x_0;

Output:

最小值点 x，函数值 $f(x)$, 迭代次数 iter，函数调用次数 feva;

Procedure:

Step 1: 初始设置 $k = 0, \varepsilon_f, \varepsilon_g > 0$;

Step 2: 计算 x_k 处的函数值，梯度和 Hesse 矩阵 f_k, g_k, G_k;

Step 3: 若 $k > 0$ 且 $|f_k - f_{k-1}| < \varepsilon_f$，循环终止，并输出 x_k, f_k;

Step 4: 计算 Hesse 矩阵 G_k 的修正 Cholesky 分解: $G_k + E_k = L_k D_k L_k^T$;

Step 5: 计算下降方向 d_k:

若 $\|g_k\| > \varepsilon_g, L_k D_k L_k^T d_k = -g_k$;

否则设 $D_{j,j} - E_{j,j}$ 当 $j = t$ 时取到最小值 φ,

若 $\varphi \geq 0$, 终止并输出; 否则求解 $L_k^T d_k = e_t$;

Step 6: 对 d_k 进行修正：若 $g_k^T d_k > 0$，取 $d_k = -d_k$;

Step 7: 利用线搜索计算步长 α_k;

Step 8: 计算 $x_{k+1} = x_k + \alpha_k d_k$;

Step 9: 令 $k = k + 1$, 回到 **Step 2**.

稳定 Newton 方法利用修正 Cholesky 分解解决了当 Hesse 矩阵有零特征值和负特征值而导致的下降方向无法计算的情况，得到了一个始终满足 $g^T d \leq 0$ 的下降方向 d，并在不定点处给出了二阶下降方向 d, 满足 $d^T G d < 0$.

2.3 More-Sorensen 方法

More-Sorensen 方法提出了二阶 Wolfe 准则:

$$x(\alpha) = x_k + \alpha^2 s_k + \alpha d_k, \quad (s_k, d_k)\text{是}x_k\text{处的下降对},$$

[1]修正 Cholesky 分解参见 [1] 的 Algorithm 3.3.2.

[2][1] 中 Algorithm 3.5.4

满足

$$f(x(\alpha)) \leq f(x) + \rho\alpha^2 \left[\nabla f(x)^T s + \frac{1}{2} d^T \nabla^2 f(x) d\right], \tag{1}$$

$$\nabla f(x(\alpha))^T x'(\alpha) \geq \sigma \left[\nabla f(x)^T d + 2\alpha \nabla f(x)^T s + \alpha d^T \nabla^2 f(x) d\right], \tag{2}$$

其中 $0 < \rho \leq \sigma < 1$.

Algorithm 3 More-Sorensen 方法

Input:

函数信息，包括函数值 f，梯度值 g 和 Hesse 矩阵 G;　初始点 x_0;

Output:

最小值点 x，函数值 $f(x)$, 迭代次数 iter，函数调用次数 feva;

Procedure:

Step 1: 初始设置 $k = 0$, $\varepsilon_f > 0$;

Step 2: 计算 x_k 处的函数值，梯度和 Hesse 矩阵 f_k, g_k, G_k;

Step 3: 若 $k > 0$ 且 $|f_k - f_{k-1}| < \varepsilon_f$，循环终止，并输出 x_k, f_k;

Step 4: 计算下降对 (s_k, d_k)；若下降对无法计算，终止并输出；

Step 5: 对 d_k 进行修正：若 $g_k^T d_k > 0$，取 $d_k = -d_k$;

Step 6: 计算步长 α_k;

Step 7: 计算 $x_{k+1} = x_k + \alpha_k d_k$;

Step 8: 令 $k = k + 1$, 回到 **Step 2**.

其中下降对的计算方法参见 [1] 的 162-163 页，而步长我们选取 $\alpha = \gamma^i$ 中满足二阶 Wolfe 条件(1)和 (2)的 i 最小的一个，其中 $0 < \gamma < 1$ 为参数。

3　程序说明

本次提交的代码包含以下文件：

- test.m:　运行文件，可以设置求解方法和求解问题。
- DampedNewton.m:　带步长 Newton 方法的函数。
- StableNewton.m:　稳定 Newton 方法的函数。
- ModCholesky.m　稳定 Newton 方法需要用到的修正 Cholesky 函数。
- Sorensen.m:　More Sorensen 方法的函数。
- DescentPair.m　More Sorensen 方法需要用到的计算下降对的函数。

设置的参数为

$$\varepsilon_f = 10^{-25}, \quad \varepsilon_g = 10^{-15}.$$

4 数值结果

4.1 Watson function

方法	n	f(x*)	迭代次数	函数调用次数
带步长 Newton 方法	6	2.3e-3	85	388
	9	1.3998e-6	109	500
	12	4.7224e-10	103	286
	15	/	/	/
	18	/	/	/
稳定 Newton 方法	6	2.3e-3	62	462
	9	1.3998e-6	9	250
	12	4.7224e-10	68	351
	15	8.448e-13	52	1325
	18	5.9097e-9	36	900
More-Sorensen 方法	6	2.3e-3	13	14
	9	1.3998e-6	14	15
	12	4.7224e-10	14	15
	15	2.7500e-8	12	2051
	18	1.0197e-7	11	2050

此处我们对于步长 Newton 和稳定 Newton 均采用强 Wolfe 准则的线搜索，其中稳定 Newton 在 n=9 时该方法失效，采用精确线搜索。

比较三个方法可以发现稳定 Newton 方法和 More-Sorensen 方法比起步长 Newton 方法更具有稳定性，能解决 Hesse 矩阵接近奇异和不定的情形。

4.2 Extended Powell singular function

方法	n	f(x*)	迭代次数	函数调用次数
带步长 Newton 方法	20	9.6476e-17	18	76
	40	1.9187e-17	19	80
	60	2.8781e-17	19	80
	80	3.8375e-17	19	80
	100	4.7968e-17	19	80
稳定 Newton 方法	20	9.6476e-17	18	76
	40	1.9187e-17	19	80
	60	2.8781e-17	19	80
	80	3.8375e-17	19	80
	100	4.7968e-17	19	80
More-Sorensen 方法	20	2.2507e-16	25	27
	40	9.2510e-17	27	28
	60	1.3877e-16	27	28
	80	1.8502e-16	27	28
	100	2.3128e-16	27	28

该问题三种方法都可以得到不错的结果，说明该问题本身的性质良好，而且该问题我们对步长 Newton 方法和稳定 Newton 方法都采用相同的非精确线搜索，可以发现这两种方法的计算结果完全一致，说明 Hesse 矩阵是一个正定矩阵。

4.3 Biggs EXP6 function

对于 $m = 13$ 的 BiggsEXP6 问题，我们的计算结果为：

- 步长 Newton 方法：无法收敛到正确的解。
- 稳定 Newton 方法：利用精确线搜索可以收敛，迭代次数 37，函数调用次数 925，$x = (1, 10, 1, 5, 4, 3)^T$,$f = 2.3190e - 18$.
- More-Sorensen 方法：可以收敛到正确解 $x = (4, 10, 3, 5, 1, 1)^T$, 迭代次数为 25，函数调用次数 39，$f = 1.4785e - 19$.

5 上机报告总结

本次上机报告，我们实现了三种梯度型下降方法，在此作一个总结：

1. 带步长 Newton 方法是普通 Newton 方法利用线搜索的一个推广，能解决 Hesse 矩阵正定情况的问题，但对较为奇异的 Hesse 矩阵和不定的 Hesse 矩阵比较乏力。
2. 稳定 Newton 方法利用了修正 Cholesky 分解，能够解决 Hesse 矩阵不正定情况下的问题。

3. More-Sorensen 方法利用了下降对的性质，能够解决 Hesse 矩阵不定的问题，但对于接近 0 的特征值没有较好的机算结果。

4. 改进的两种计算方法无论是在计算结果，还是在迭代效率上，都比起初始的 Newton 方法具有不小的提高。

参考文献

[1] Wenyu Sun and Ya-Xiang Yuan. Optimization theory and methods: nonlinear programming. Springer, 2006.