

数值分析 上机作业 3

李佳 2100010793

1 问题描述

对 Dennis-Schnabel 书中 Appendix B 中的例 2,3,5 进行数值求解, 其中例子分别为:

(2) Extended Powell Singular Function

a) $n = 4m, m \in \mathbb{N}_+$;

b) for $i = 1, 2, \dots, n/4$:

$$f_{4i-3}(x) = x_{4i-3} + 10x_{4i-2}$$

$$f_{4i-2}(x) = \sqrt{5}(x_{4i-1} - x_{4i})$$

$$f_{4i-1}(x) = (x_{4i-2} - 2x_{4i-1})^2$$

$$f_{4i}(x) = \sqrt{10}(x_{4i-3} - x_{4i})^2$$

c) 初值 $x_0 = (3, -1, 0, 1, \dots, 3, -1, 0, 1)$

d) 真解 $x_* = (0, 0, 0, 0, \dots, 0, 0, 0, 0)$

(3) Trigonometric Function

a) $n \in \mathbb{N}_+$;

b) for $i = 1, 2, \dots, n$:

$$f_i(x) = n - \sum_{j=1}^n \left(\cos x_j + i(1 - \cos x_i) - \sin x_i \right) = n + ni(1 - \cos x_i) - n \sin x_i - \sum_{j=1}^n \cos x_j$$

c) 初值 $x_0 = (1/n, 1/n, \dots, 1/n)$

(5) Wood Function(优化问题转化为非线性方程组求解问题)

a) $n = 4$;

b)

$$\begin{aligned} f(x) &= 100(x_1^2 - x_2)^2 + (1 - x_1)^2 + 90(x_3^2 - x_4)^2 + (1 - x_3)^2 \\ &\quad + 10.1((1 - x_2)^2 + (1 - x_4)^2) + 19.8(1 - x_2)(1 - x_4) \end{aligned}$$

c) 初值 $x_0 = (-3, -1, -3, -1)$

d) 极小值点 $x_* = (1, 1, 1, 1)$

2 算法描述

2.1 Newton 法

Newton 迭代法使用泰勒展开线性近似 ($F(x) \approx F(x_0) + J_F(x_0)(x - x_0)$, $J_F(x_0)$ 为 x_0 处 Jacobi 矩阵) 的方法, 求出一个近似零点, 再在该点附近线性近似得到下一个近似零点, 以此类推得到迭代序列. 具体来说, 迭代法为:

$$x_{k+1} = x_k - J_F(x_k)^{-1}F(x_k).$$

在 $F(x)$ 二阶连续可微且零点 x_* 处的 Jacobi 矩阵非奇异时, Newton 法有二阶局部收敛性;

该方法的缺点是每一步需要计算 Jacobi 矩阵, 并解一个线性方程组, 一般需要 $O(n^3)$ 的更新代价 (n 为方程个数), 因此不适合大规模计算.

2.2 改进的 Newton 法: Broyden 法

改进的 Newton 法: Broyden 法希望降低每一步求 $J_F(x_k)^{-1}$ 的代价, 认为在靠近零点时, 矩阵的变化很小, 希望通过秩一的更新来近似 $J_F(x_k)$, 这样再通过 Sherman-Morrison 公式即可得到 $J_F(x_k)^{-1}$.

具体来说, 记第 k 步近似的 Jacobi 矩阵为 A_k , 则希望寻找向量 v, w 使得 $A_{k-1} + vw^T$ 近似为 Jacobi 矩阵, 即要求 $F(x_k) - F(x_{k-1}) = (A_{k-1} + vw^T)(x_k - x_{k-1})$ 精确成立.

这样每一步的更新代价是 $O(n^2)$, 与原有的 Newton 法相比更具有实用性. Broyden 法在 F 满足一定条件时可达到局部超线性收敛.

另外, v, w 的取法不唯一. 以下给出两种常用的取法, 也是本报告中涉及的取法:

1. $w = Y_{k-1} := x_k - x_{k-1}$, 此时可计算出 $v = \frac{1}{\|Y_{k-1}\|_2^2} F(x_k)$ (该取法的好处: $\|Y_{k-1}\|_2 = 0$ 等价于迭代停止), 通过 Sherman-Morrison 公式计算得到

$$A_k^{-1} = A_{k-1}^{-1} - \frac{A_{k-1}^{-1}F(x_k)Y_{k-1}A_{k-1}^{-1}}{Y_{k-1}^T Y_{k-1} + Y_{k-1}^T A_{k-1}^{-1}F(x_k)}$$

2. $w = F(x_k)$, 此时有 $v = \frac{1}{F(x_k)^T Y_{k-1}} F(x_k)$ (该取法的好处: A_k 恒为对称矩阵, 适合近似优化问题的 Hesse 矩阵), 类似地可以算出

$$A_k^{-1} = A_{k-1}^{-1} - \frac{A_{k-1}^{-1}F(x_k)F(x_k)^T A_{k-1}^{-1}}{F(x_k)^T Y_{k-1} + F(x_k)A_{k-1}^{-1}F(x_k)}.$$

2.3 为优化问题求解而改进的 Newton 法

由于优化问题中要求的是极小值, 而梯度为 0 的位置既有可能是极小值、也可能是极大值、也可能不是极值点, Hesse 矩阵正定时能保证其是极小值, 因此直接对梯度为 0 的方程使用 Newton 法或 Broyden 方法可能无法收敛到极小值点. 在

参考 Dennis-Schnabel 原书 (Numerical Methods For Unconstrained Optimization and Nonlinear Equations) 5.5 节后, 发现可以对 Newton 法改进以使得迭代序列收敛于极小值点:

令 $f(x)$ 为待优化的函数, 原本的 Newton 法为:

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \cdot \nabla f(x_k),$$

改进后为:

$$x_{k+1} = x_k - (\nabla^2 f(x_k) + \mu_k I)^{-1} \cdot \nabla f(x_k),$$

其中 $\mu_k \geq 0$ 使得 $\nabla^2 f(x_k) + \mu_k I$ 正定 (实际计算中, 仅取 μ_k 为一个正的常数).

(由于本上机报告是非线性方程数值解的上机作业, 因此不再考虑其它与牛顿法联系不大的优化方面的算法.)

3 计算结果

对例 2(Powell Function), 计算 $n = 4$ 的情形 (事实上, 该问题的 $4m$ 个变量的方程组, 等价于相同的 m 组 4 个变量的方程组, 因此只以 $n = 4$ 为例求解). 由于已知真解, 因此迭代停止的标准设置为误差的 2 范数不大于 10^{-8} . 采用 Newton 法和 Broyden 法 (取法 1) 计算, 均收敛于真解 $(0, 0, 0, 0)$, 且误差 2-范数随迭代步数的变化如图所示:

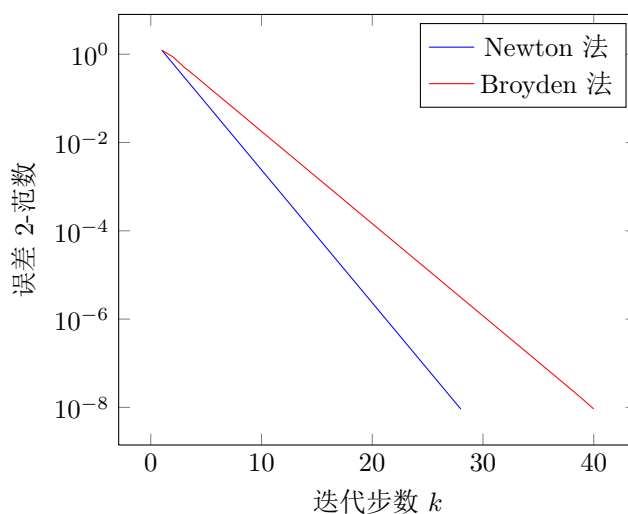


Figure 1: 例 2 使用 Newton 法、Broyden 法时误差 2-范数随迭代步数的变化图

对例 3(Trigonometric Function), 计算 $n = 10, 20, 30, 40, 50$ 的情形, 由于未知真解 (事实上, 有无穷多真解, 且计算发现存在非平凡的解, 即不是 $x_j = 2k\pi$ 的解), 因此迭代停止的标准设置为相邻两步的步长 2 范数不大于 10^{-8} , 即 $\|x_k - x_{k-1}\|_2 \leq 10^{-8}$. (当迭代法达到线性收敛时, $\|x_k - x_{k-1}\|_2$ 即可控制 $\|x_k - x_*\|_2$ 的大小, 因此考虑用这个量作为停机标准).

$n = 10, 20, 30, 40, 50$ 时, 以初值为 $(1/n, \dots, 1/n)$, 分别采用 Newton 法和 Broyden 法 (取法 1) 计算, 发现 Newton 法收敛, 但 Broyden 法不收敛. 在改变初值为 $(1/n^3, \dots, 1/n^3)$

时, Broyden 法可以收敛. 当 $n = 10$ 时, 相邻两步的步长 2-范数随迭代步数的变化如图所示; $n = 10, 20, 30, 40, 50$ 时两方法达到停机标准的迭代次数如表所示.

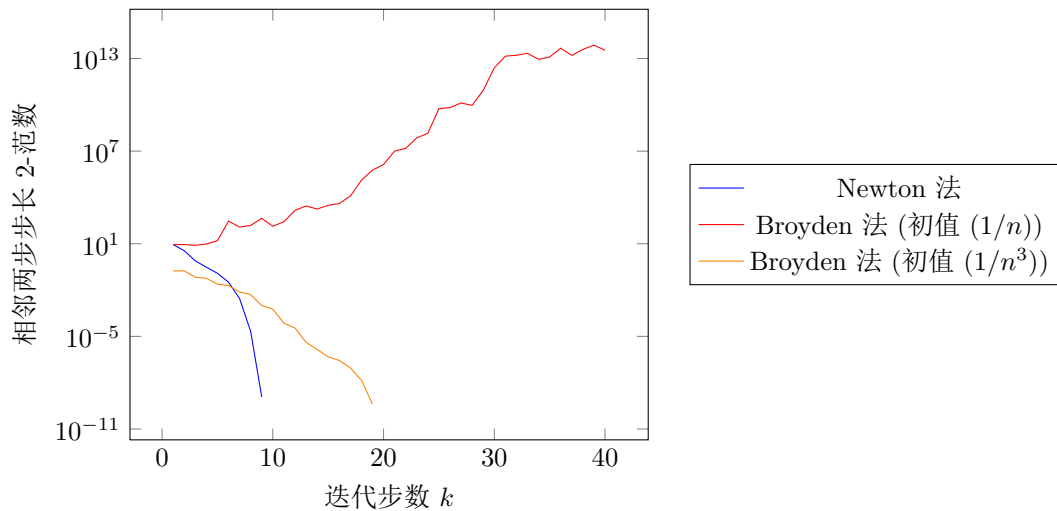


Figure 2: 例 3 在 $n = 10$ 时, 各方法相邻两步的步长 2-范数随迭代步数的变化

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
Newton 法 初值 $(1/n, \dots, 1/n)$	9	10	11	10	11
Broyden 法 初值 $(1/n^3, \dots, 1/n^3)$	19	24	30	36	43

Table 1: 例 3 各方法达到停机标准的迭代次数

例 5(Wood Function) 将优化问题转化为求解梯度为 0 的非线性方程组, 有可能不收敛于极小值点, 因此停机标准依然设置为相邻两步的步长 2 范数不大于 10^{-8} , 即 $\|x_k - x_{k-1}\|_2 \leq 10^{-8}$.

计算发现 Newton 法、Broyden 法 (取法 2) 均收敛于一个鞍点 $(-0.968, 0.947, -0.970, 0.951)$, Broyden 法 (取法 1) 可经过 13702 步的长时间迭代后收敛于极小值点 $(1, 1, 1, 1)$, 为优化问题改进的 Newton 法在取 μ_k 为正的常数时大多可以收敛于极小值点.

使用 Newton 法、Broyden 法 (取法 2)、改进的 Newton 法 (以 $\mu_k = 0.1$ 为例), 优化函数值 $f(x_k)$ 随迭代步数 k 的变化如图所示; 使用 Broyden 法 (取法 1) 时, 优化函数值 $f(x_k)$ 随迭代步数 k 的变化如图所示; 改进的 Newton 法在 μ_k 取不同的正的常数时达停机标准所需的迭代次数如表所示.

μ_k	0.01	0.02	0.05	0.06	0.07	0.08	0.09	0.10	0.20	0.30
迭代次数	45	41	41	41	35	58	38	42	44	45
μ_k	0.40	0.50	0.60	0.70	0.80	0.90	1.00	2.00	3.00	4.00
迭代次数	55	53	56	63	68	70	75	120	162	204

Table 2: 例 5 中改进的 Newton 法在 μ_k 取不同的正的常数时达停机标准所需的迭代次数

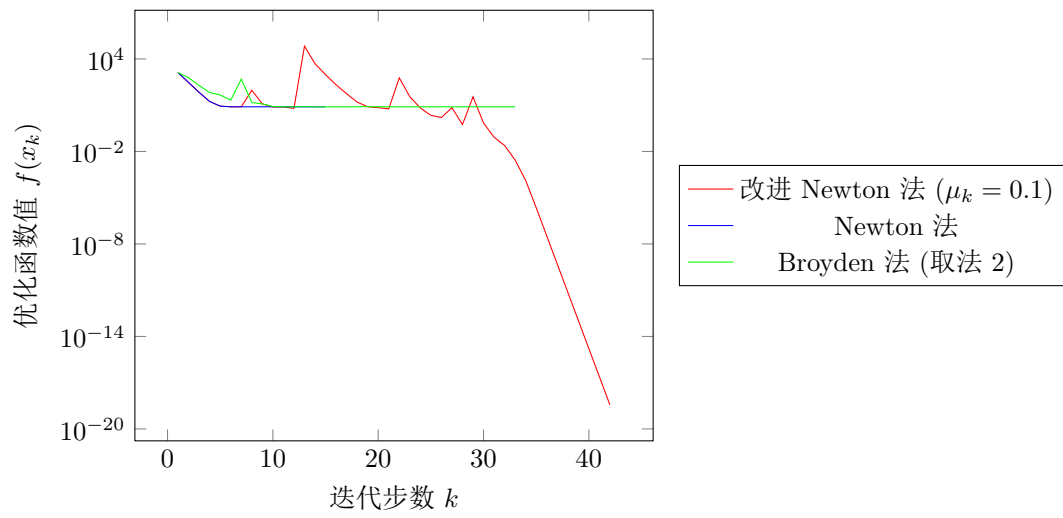


Figure 3: 例 5 使用 Newton 法、Broyden 法 (取法 2)、改进的 Newton 法 ($\mu_k = 0.1$), $f(x_k)$ 随迭代步数 k 的变化

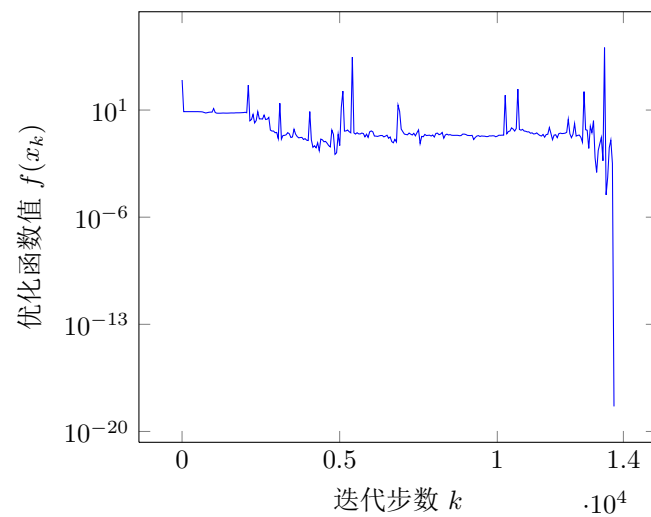


Figure 4: 例 5 使用 Broyden 法 (取法 1), $f(x_k)$ 随迭代步数 k 的变化

4 简明分析

对例 2(Powell Function), 由图可看出:

- 由于零点处 Jacobi 矩阵奇异, 因此 Newton 法无法达到二阶收敛, 但 Newton 法和 Broyden 法均能达到线性收敛;
- Newton 法的收敛速度快于 Broyden 法, 所需迭代次数更少.

对例 3(Trigonometric Function), 由图表可看出:

- 可能受附近多个零点的影响, 初值为 $(1/n, \dots, 1/n)$ 时, Broyden 法近似的 Jacobi 矩阵的误差让 x_k 不断向外, 继而导致 Jacobi 矩阵的近似变差, 从而又导致 x_k 不断偏离且步长越来越大, 导致无法收敛;
- 初值为 $(1/n^3, \dots, 1/n^3)$ 时, 离一个零点 $(0, \dots, 0)$ 的距离更近, 这使得 Broyden 法更容易收敛;
- 初值为 $(1/n, \dots, 1/n)$ 时, Newton 法均能收敛. 与前两条分析对比可知, 该例子中 Newton 法对初值的要求低于 Broyden 法.
- 在迭代法收敛时, 它们均至少线性收敛;
- 在 n 变化时, Newton 法达到停机要求所需的迭代次数几乎不变, Broyden 法 (初值为 $(1/n^3, \dots, 1/n^3)$) 所需迭代次数近似线性增长.

对例 5(Wood Function), 由图表可看出:

- Newton 法、Broyden 法 (取法 2) 均没能得到极小值点, 均收敛于鞍点, 这是用非线性方程组求解优化问题所不可克服的问题, 因为梯度为 0 不是极小值点的充分条件, 要确保得到极小值点, 必须利用 Hesse 矩阵的信息并利用优化方面的知识;
- Broyden 法 (取法 1) 虽然收敛于极小值点, 但由图可以看出其大部分时间在鞍点处徘徊, 到最后才快速到达极小值点, 猜测这可能是由于迭代过程有一定的不稳定性, 异常的波动恰好将点拉到了极小值点附近;
- 改进的 Newton 法在 μ_k 取为正的常数时, 大多数情况都能收敛于极小值点. 这可能是因为是在鞍点处, μ_k 项会产生扰动, 使点远离鞍点, 促使点向极小值点移动;
- 在 $\mu_k \leq 0.1$ 时, 迭代次数大多在 40 次上下, $\mu_k > 0.10$ 后, 迭代次数随 k 呈上升趋势, 且 $\mu_k > 1.00$ 后近似为线性增长. 这可能是因为 μ_k 较大时, 使每一步的步长减少, 导致迭代次数增多.