

第三次大作业实验报告

王嘉禾 2193211079 计算机试验班 001

问题描述

10.15 等式约束熵极大化。考虑等式约束熵极大化问题

$$\begin{aligned} &\text{minimize} && f(x) = \sum_{i=1}^n x_i \log x_i \\ &\text{subject to} && Ax = b, \end{aligned}$$

其中 $\text{dom } f = \mathbf{R}_{++}^n$, $A \in \mathbf{R}^{p \times n}$, $p < n$ 。(一些相关分析见习题 10.9。)

生成一个 $n = 100$, $p = 30$ 的问题实例, 随机选择 A (验证其为满秩阵), 随机选择一个正向量作为 \hat{x} (例如, 其分量在区间 $[0, 1]$ 上均匀分布), 然后令 $b = A\hat{x}$ 。(于是, \hat{x} 可行。)

采用以下方法计算该问题的解。

- (a) **标准 Newton 方法。** 可以选用初始点 $x^{(0)} = \hat{x}$ 。
- (b) **不可行初始点 Newton 方法。** 可以选用初始点 $x^{(0)} = \hat{x}$ (和标准 Newton 方法比较), 也可以选用初始点 $x^{(0)} = \mathbf{1}$ 。
- (c) **对偶 Newton 方法,** 即将标准 Newton 方法应用于对偶问题。

证实三种方法求得相同的最优点 (和 Lagrange 乘子)。比较三种方法每步迭代的计算量, 假设利用了相应的结构。(但在你的实现中不需要利用结构计算 Newton 步径。)

解题思路

- (1) 可行初始点的牛顿方法: 提前保存好 KKT 矩阵, 每次迭代时计算梯度, 更新 KKT 矩阵, 获取牛顿下降方向 dnt, 后续的搜索和迭代过程与无约束的牛顿迭代方法完全相同。
- (2) 不可行初始点的牛顿方法: 与可行初始点的牛顿方法类似, 但求解下降方向和回溯直线搜索的过程不同, 需要额外定义一个残差向量 r, 在求解下降方向时使用 KKT 矩阵的逆矩阵与残差向量 r 的乘积, 和回溯直线搜索时使用残差范数进行比较
- (3) 对偶牛顿方法: 对偶函数的相关求解过程如下图所示

将原问题转化成对偶问题

设 $A_{con100} = [a_1, a_2, \dots, a_{100}]$, $a_i \in \mathbb{R}^{50}$

$$L(x, v) = \sum_{i=1}^{100} x_i \log x_i + v^T (Ax - b)$$

$$\text{令 } \frac{\partial L(x, v)}{\partial x_i} = 0, \text{ 得 } 1 + \log x_i + a_i^T v = 0$$

$$\Rightarrow x_i = e^{-a_i^T v - 1}$$

将 x 代入 $L(x, v)$ 得

$$\begin{aligned} g(v) &= \min_x L(x, v) = \sum_{i=1}^{100} e^{-a_i^T v - 1} (-a_i^T v - 1) \\ &\quad + v^T a_i e^{-a_i^T v - 1} - v^T b \\ &= -e^{-1} \sum_{i=1}^{100} e^{-a_i^T v} - v^T b \end{aligned}$$

因此原问题的对偶问题为

$$\min_{v \in \mathbb{R}^{50}} e^{-1} \sum_{i=1}^{100} e^{-a_i^T v} + v^T b$$

$$\text{令 } h(v) = e^{-1} \sum_{i=1}^{100} e^{-a_i^T v} + v^T b$$

$$\text{则 } \nabla h(v) = b - e^{-1} A e^{-A^T v} \quad \left(\frac{\partial h(v)}{\partial v_i} = b_i + e^{-1} \sum_{j=1}^{100} -a_{ij} e^{-a_j^T v} \right)$$

$$\nabla^2 h(v) = A \operatorname{diag}(e^{-A^T v}) A^T e^{-1}$$

对偶牛顿方法可分为两个阶段

阶段一：对 $h(v)$ 进行牛顿搜索，获得 h 的最小值所对应的 v^* ，这也是 $g(v)$ 的最大值所对应的 v^*

阶段二：对于 lagrange 函数 $L(x, v)$ ，固定 $v=v^*$ ，用牛顿搜索方法搜索最小值 p^* 和对应的 x^*

无约束牛顿搜索方法均与第二次大作业中问题三的解法完全相同

程序设定

初始时随机生成矩阵 A ，符合等式条件的初始点 x_{hat} ，通过 $A \cdot x_{\text{hat}}$ 构造出 b ，就可以等效于获得等式约束条件 $A \cdot x = b$ 的符合等式约束条件的初始点。对于三种方法，均取

$$\varepsilon = 1e-8, \alpha = 0.3, \beta = 0.7$$

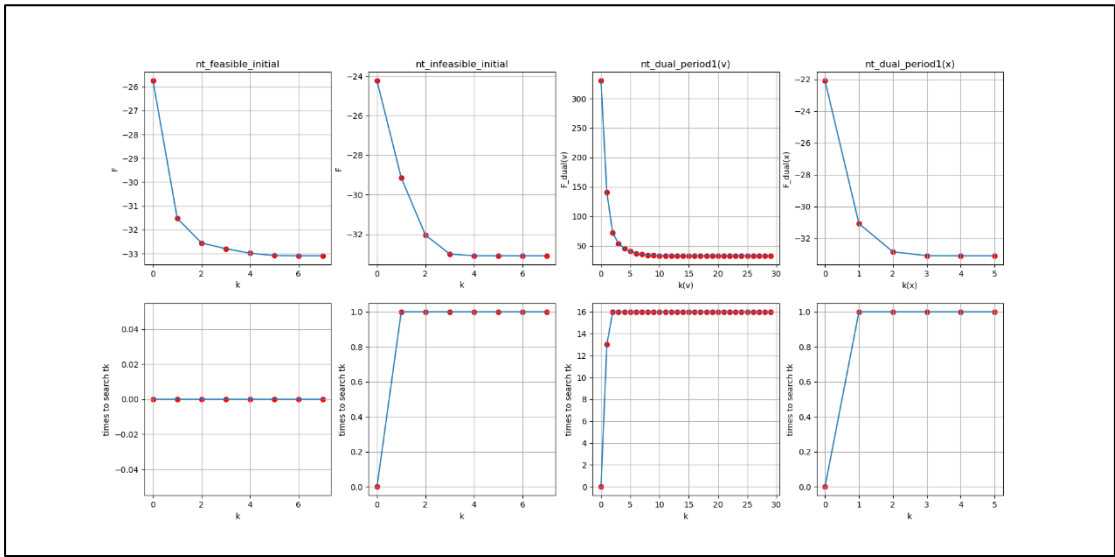
程序中以对偶残差的范数小于等于 ε 作为判断是否找到可行解的依据

代码实现

见 4_newton.py

实验结果

选取两组运行结果（左侧最终结果对应第一张图的迭代过程，右侧最终结果对应第二张图的迭代过程）



```

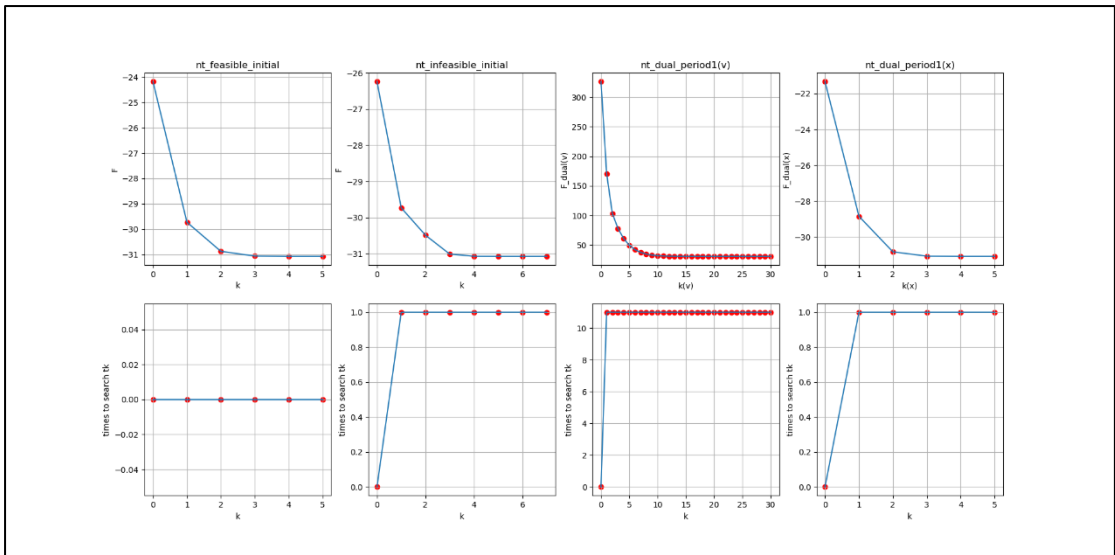
-----nt_feasible_initial-----
total iteration: 7
final p*: -33.09241110640459
-----nt_infeasible_initial-----
total iteration: 7
final p*: -33.0924111081274
-----nt_dual-----
total iteration: 34
final p*: -33.09241111969546

```

```

-----nt_feasible_initial-----
total iteration: 5
final p*: -31.07526850908804
-----nt_infeasible_initial-----
total iteration: 7
final p*: -31.07526850981922
-----nt_dual-----
total iteration: 35
final p*: -31.075268533830922

```



实验结论

对于实验一：可见，三种方法求出的最优解几乎相等，误差在 $1e-8$ 的数量级，与设定相符。从迭代过程的示意图可看出，可行初始点和不可行初始点的牛顿搜索方法中误差的下降过程几乎完全相同，总迭代次数也相同，均在 7 次迭代时完成搜索，唯一的区别是不可行初始点的牛顿搜索在第一部迭代时找到了可行解。而对于对偶牛顿方法，阶段一搜索 v^* 的过程消耗了很多次迭代，使得搜索速度变慢；而阶段二搜索 x^* 的过程与前两种牛顿

方法相近。

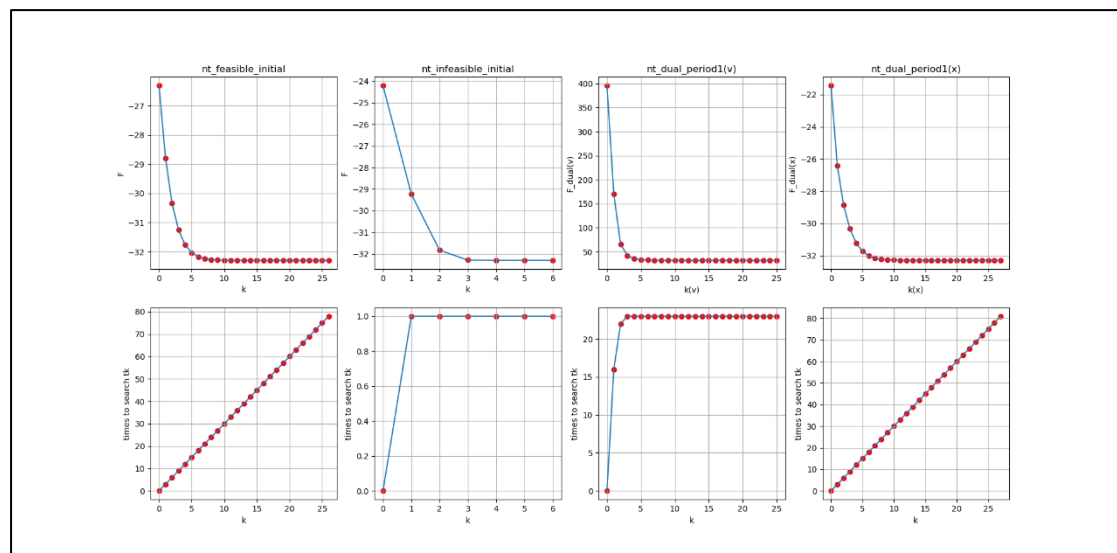
对于实验二：与实验一相似，三种方法求出的最优解的误差也在 $1e-8$ 的数量级，与设定相符。收敛速度也都很快，在 5 次和 7 次时完成迭代。而对偶牛顿方法同样在对偶函数 $g(v)$ 的搜索上消耗了巨大代价，进行了 30 次迭代，导致总迭代次数达到了 35 次。

由此可见，对于等式约束的极大熵问题，对偶牛顿方法代价较大，而可行初始点牛顿方法和不可行初始点的牛顿方法代价相似，不可行初始点的牛顿搜索会以极快的速度找到可行点。

拓展思考

在本次实验以及第二次大作业的问题三中，发现回溯直线搜索时无论 α 和 β 如何设置，回溯直线搜索时获得的步长 tk 均为 1，这从本实验的第二行图像中可看出，除对偶牛顿方法和非可行初始点的牛顿方法的第一二次迭代中出现对 tk 的搜索之外，之后的迭代过程中对 tk 的搜索次数均没有增长。

开始时猜测是程序实现过程出现故障导致迭代过程直接进入了二次收敛阶段。但思考后发现， $tk=1$ 是进入二次收敛阶段的必要条件，但不是充分条件，换言之，此时可能是阻尼牛顿阶段中的 tk 刚好为 1。其次，通过更多的实验发现，当 α 设置为一个大于 0.6 的值时，回溯直线搜索中对 tk 的搜索次数明显增加，以 $\alpha=0.8$ 为例



从函数值的下降曲线也可看出，出现了明显的阻尼牛顿阶段。但在这时，迭代次数明显增加，因为对于牛顿减少量的要求较高，导致迭代步长很小，这也是回溯直线搜索次数较多的原因。

结合问题三中实验二的结论：在 $[0.1 \ 0.4]$ 的区间中， α 的值对迭代次数的影响很小，猜测也是因为 α 的值很小，导致每次都采用了同样的 $tk=1$ 。在课件收敛性的证明中，将 α 限定在 $[0 \ 0.5]$ 中是为了凑出可行性结论，其中的很多放缩都十分夸张，每一步都考虑最坏情况，实际的情况可能比证明中的情况要更加优良，结论也更强。