

# 等式约束熵极大化问题

考虑等式约束熵极大化问题:  $\min f(x) = \sum_{i=1}^m x_i \log x_i \text{ s.t. } Ax = b_i$

其中 $dom\ f = \mathbb{R}_{++}^n, A \in \mathbb{R}^{p \times n}, p < n$ 。

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。)

## A.标准Newton方法

### 问题分析

$$f(x) = \sum_{i=1}^m x_i \log x_i, \quad \text{s.t. } Ax = b,$$

其梯度和 Hessian 分别为：

$$\nabla f(x) = \log(x) - \mathbf{1}, \quad \nabla^2 f(x) = \text{diag}\left(\frac{1}{x}\right).$$

求解牛顿方向

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_x \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ 0 \end{bmatrix}$$

### 代码实现思路

- 1. **构造增广矩阵**：

$$\left[ \begin{array}{cc|c} \nabla^2 f(x) & A^T & -\nabla f(x) \\ A & 0 & 0 \end{array} \right].$$

- 2. **初等行变换**：将增广矩阵化为以下形式：

$$\left[ \begin{array}{cc|c} I & 0 & d_x \\ 0 & I & w \end{array} \right],$$

其中， $d_x$  和  $w$  分别为原问题和对偶问题的牛顿方向。

- 3. **提取结果**：
  - 提取  $d_x$  用于更新原变量；
  - 提取  $w$  作为中间结果（或用于进一步分析对偶变量）。

### 代码实现

见cpp文件standardNewton函数部分。

## B.不可行初始点Newton方法

### 问题分析

$$\min_x f(x) = \sum_{i=1}^m x_i \log x_i \quad \text{subject to} \quad Ax = b,$$

其中，目标函数的梯度和 Hessian 为：

$$\nabla f(x) = \log(x) - \mathbf{1}, \quad \nabla^2 f(x) = \text{diag}\left(\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_n}\right).$$

为了解这个问题，我们考虑对偶方法，并选择对偶变量的初始点  $v^0$ 。

通过构造增广矩阵并求解牛顿方向  $d_x$  和  $d_v$ ，满足以下线性系统：

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_v \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^T v \\ Ax - b \end{bmatrix}$$

其中， $d_x$  是原问题的牛顿步长， $d_v$  是对偶问题的牛顿步长。通过求解这个系统，可以同时更新原始变量  $x$  和对偶变量  $v$ 。

### 代码实现

见cpp文件中infeasibleNewton函数部分。

### 实验结果

## C.对偶Newton方法

### 问题分析

#### 1. 对偶问题的形式

对于原始问题：

$$\min_x f(x), \quad \text{subject to} \quad Ax = b,$$

其中  $f(x) = \sum_i x_i \log x_i$ 。

可以写出拉格朗日对偶函数：

$$g(v) = \inf_x \{f(x) + v^\top (Ax - b)\}.$$

通过对  $x$  求极值点，可得到对偶问题的显式形式：

$$\max_v g(v) = b^\top v - \sum_{j=1}^n e^{-1-A_j^\top v},$$

其中  $A_j$  是矩阵  $A$  的列向量。

对偶变量  $v$  是约束  $Ax = b$  的拉格朗日乘子。

#### 2. 梯度和 Hessian 的计算

对  $g(v)$  求梯度：

$$\nabla g(v) = b - A \left[ e^{-1-A^\top v} \right],$$

其中  $e^{-1-A^\top v}$  表示逐元素指数计算。

对  $g(v)$  求 Hessian：

$$H_g(v) = A \text{diag}(e^{-1-A^\top v}) A^\top,$$

其中  $\text{diag}(e^{-1-A^\top v})$  是对角矩阵，其对角线元素是向量  $e^{-1-A^\top v}$  的元素。

这些表达式用于计算梯度和 Hessian，构成牛顿方向的更新。

### 3. 原始变量的回代

优化完成后，通过  $x$  和  $v$  的关系，回代求解原始问题最优解：

$$x = e^{-1-A^\top v}.$$

该关系来自对拉格朗日对偶函数中原始变量  $x$  求极值点的结果。

### 代码逻辑

1. 初始化对偶变量  $v$ ：
  - 初始化为零向量。
2. 计算梯度  $\nabla g(v)$  和 Hessian  $H_g(v)$ 。
3. 解线性系统：
  - 解  $H_g(v)\Delta v = -\nabla g(v)$ 。
4. 判断收敛条件：
  - 如果  $\Delta v$  的范数小于容差  $\text{tol}$ ，则终止迭代。
5. 更新变量：
  - 更新  $v \leftarrow v + \Delta v$ 。
6. 回代计算原始解  $x$ ：
  - 使用关系  $x = e^{-1-A^\top v}$ 。

### 代码实现

见cpp文件中的dualNewton函数部分。

### 实验结果

设定 $\alpha = 0.3, \beta = 0.8, \epsilon = 10^{-6}$

**随机生成矩阵A：见附属文件A.txt。**

Feasible point  $\hat{x}$ :

0.2607	0.235064	0.108208	0.555996	0.702413	0.74324	0.882926	0.584053	0.630977	0.339557	0.871641	0.184798	0.393275	0.431728	0.159265
0.513395	0.211547	0.372688	0.310698	0.31823	0.113636	0.301842	0.201649	0.233152	0.723081	0.530351	0.845204	0.84907	0.44743	0.462554
0.139367	0.342024	0.376146	0.148175	0.476463	0.784983	0.196565	0.121294	0.892348	0.796348	0.848121	0.422987	0.601021	0.738506	
0.117408	0.53136	0.842048	0.386616	0.441816	0.317494	0.338015	0.514913	0.229381	0.438043	0.68595	0.766693	0.113919	0.305378	
0.555307	0.608468	0.180609	0.391811	0.598205	0.713791	0.208883	0.438912	0.55304	0.567853	0.888382	0.393243	0.277357	0.179092	
0.507737	0.5546	0.19811	0.896703	0.365885	0.262666	0.865412	0.459879	0.574441	0.417355	0.574728	0.73208	0.583453	0.3771	0.264287
0.30997	0.680529	0.555557	0.641939	0.739605	0.219617	0.493718	0.495554	0.743045	0.764188	0.678566	0.617963	0.529199		

**$b = A * \hat{x}$ :**

-1.62277	1.75238	-2.62354	2.38162	-0.716927	4.21842	-2.71374	4.94208	-0.578789	-1.62751	0.954415	1.20528	0.735541	-2.3129	2.25533
2.55482	-0.774638	-1.84471	-2.32392	-3.39078	-1.91197	0.094906	6.01439	3.37301	-0.739257	-6.6817	2.07204	0.35901	2.36965	4.10899

**标准牛顿方法最优解：**

Optimal  $x$  (Standard Newton):

0.375651	0.37197	0.412313	0.383182	0.41823	0.632536	0.607331	0.543956	0.283283	0.302712	0.536753	0.276502	0.395212	0.488789	
0.333625	0.223168	0.196361	0.426739	0.342182	0.280131	0.33665	0.339266	0.436805	0.411129	0.309242	0.355495	0.90656	0.341033	
0.569653	0.490964	0.261432	0.246744	0.222169	0.349135	0.593101	0.421873	0.353062	0.213653	0.723313	0.681981	0.327862	0.409097	
0.554899	0.545989	0.246352	0.315799	0.58281	0.335216	0.408774	0.217954	0.404517	0.369471	0.417761	0.274748	0.340221	0.586324	
0.224591	0.364834	0.279618	0.359871	0.434788	0.507513	0.317098	0.698837	0.399674	0.552251	0.473379	0.503285	0.513707	0.279667	
0.296895	0.318071	0.647173	0.427255	0.32345	0.445101	0.53116	0.274745	0.651739	0.475222	0.574224	0.393474	0.333727	0.426775	0.62646

0.341634 0.244148 0.42674 0.357338 0.374846 0.331367 0.238504 0.329985 0.403972 0.476826 0.498167 0.591668 0.396593 0.281307 0.521784

迭代次数： k = 5

最优值： p\* = -34.4719

不可行初始点牛顿方法最优解：

Optimal x (Infeasible Newton):  
0.375652 0.371968 0.412313 0.38318 0.41823 0.632537 0.607331 0.543957 0.283283 0.302712 0.536752 0.276502 0.395213 0.488789 0.333626 0.223169 0.196361 0.426739 0.342181 0.280131 0.336649 0.339267 0.436803 0.411129 0.309242 0.355493 0.906559 0.341034 0.569653 0.490966 0.26143 0.246743 0.222169 0.349135 0.593102 0.421874 0.353063 0.213653 0.723311 0.681982 0.327862 0.409096 0.554899 0.545992 0.246352 0.315798 0.58281 0.335215 0.408773 0.217953 0.404517 0.369473 0.417761 0.274747 0.340222 0.586325 0.22459 0.364835 0.279618 0.359871 0.434789 0.507512 0.317098 0.698837 0.399673 0.552253 0.473377 0.503287 0.513709 0.279667 0.296896 0.318071 0.647172 0.427256 0.32345 0.445102 0.53116 0.274744 0.651741 0.475222 0.574224 0.393473 0.333728 0.426776 0.626461 0.341633 0.244147 0.426739 0.357338 0.374847 0.331367 0.238517 0.329985 0.403973 0.476826 0.498168 0.591668 0.396594 0.281307 0.521781

迭代次数： k = 6

最优值： p\* = -34.4719

对偶牛顿方法最优解：

Optimal x (Dual Newton):  
0.375652 0.371968 0.412313 0.38318 0.41823 0.632537 0.607331 0.543957 0.283283 0.302712 0.536752 0.276502 0.395213 0.488789 0.333626 0.223169 0.196361 0.426739 0.342181 0.280131 0.336649 0.339267 0.436803 0.411129 0.309242 0.355493 0.906559 0.341034 0.569653 0.490966 0.26143 0.246743 0.222169 0.349135 0.593102 0.421874 0.353063 0.213653 0.723311 0.681982 0.327862 0.409096 0.554899 0.545992 0.246352 0.315798 0.58281 0.335215 0.408773 0.217953 0.404517 0.369473 0.417761 0.274747 0.340222 0.586325 0.22459 0.364835 0.279618 0.359871 0.434789 0.507512 0.317098 0.698837 0.399673 0.552253 0.473377 0.503287 0.513709 0.279667 0.296896 0.318071 0.647172 0.427256 0.32345 0.445102 0.53116 0.274744 0.651741 0.475222 0.574224 0.393473 0.333728 0.426776 0.626461 0.341633 0.244147 0.426739 0.357338 0.374847 0.331367 0.238517 0.329985 0.403973 0.476826 0.498168 0.591668 0.396594 0.281307 0.521781

迭代次数： k = 4

最优值： p\* = -34.4719

总结

该问题的最优值 $p^* \approx -34.4719$ ，三种不同方法均成功求得结果。

三种方法求得的 $x^*$ 及 $v^*$ ，在误差范围内，证实了三种方法求得相同的最优点和拉格朗日乘子。

比较三种方法每步迭代的计算量：

标准 Newton 方法：需要对高维增广矩阵进行高斯消元以得到牛顿方向，并进行回溯搜索。

不可行初始点 Newton 方法：同样需要高维矩阵消元，但额外需要计算残差。

对偶 Newton 方法：每次迭代中需要消元的矩阵较小，若仅需计算最优值 $p^*$ ，甚至不需进行回代，因此计算量最小。