# Math 127 (Spring 2007)

Homework #1
30 Jan 2007

## Question 1

BLAST stands for Basic Local Alignment Search Tool. It is a family of programs developed by the National Center for Biotechnology Information (NCBI) for the purpose of comparing gene and protein sequences with those in public databases. BLAST was born in 1990 when five researchers – Stephen Altschul, Warren Gish, David Lipman (all from NCBI), Webb Miller (Pennsylvania State University), and Gene Myers (University of Arizona) – designed and implemented an algorithm that was able to search large genomic databases more than 50 times faster than traditional exhaustive methods with only a slight loss in accuracy. The programs in the BLAST family are based on that algorithm. Since then, it has been a popular tool used by researchers for a variety of biological enquires, such as discovering the function of a gene sequence in one animal by comparing it with sequences with known functions in other animals, or finding evolutionary relationships between different animals. The popularity of BLAST stems from its speed, easy access, and availability of specialized versions for different types of problems. The user is also allowed to adjust various parameters in order to fine-tune the search to suit the needs of his/her enquiry. Examples of the parameters include gap costs, statistical significance threshold and choice of database to search.

At `http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/tut1.html`, one can find a tutorial introducing the use of BLAST. The following example was extracted from this website. It is the amino acid sequence of the uncharacterized archaebacterial protein MJ0577 from the Methanococcus Jannaschii, and we wish to search for sequence relatives in the amino acid database. We click on the "Protein-protein BLAST (blastp)" tool on the BLAST homepage and paste the following sequence in FASTA format into the search box:

```
>gi|2501594|sp|Q57997|Y577_METJA PROTEIN MJ0577
MSVMYKKILYPTDFSETAEIALKHVKAFKTLKAEEVILLHVIDEREIKKRDIFS
LLLGVAGLNKSVEEFENELKNKLTEEAKNKMENIKKELEDVGFKVKDIIVVGIP
HEEIVKIAEDEGVDIIIMGSHGKTNLKEILLGSVTENVIKKSNKPVLVVKRKNS
```

Leaving all the other options unchanged, we click on the BLAST button and the FORMAT button on the following page. A colorful chart showing the 114 BLAST hits appears on the screen, with the hits ranked according to their alignment scores. The red hits are sequences in the database demonstrating incredibly close alignment with our input sequence. It is not surprising that we have one red hit, which is the original amino acid sequence MJ0577 in the database. Next, we have 12 pink hits with relatively good alignment. Studying these sequences reveals that most of them are stress response proteins. This tells us that the MJ0577 is likely to be a stress response protein. The chart also shows several green and blue hits which are lower on the alignment scale. Below the chart, we see our original amino acid sequence placed side-by-side with the hit sequences in the database. The individual acids which match up, as well as gaps in the alignments, are clearly shown. For instance, in our first pink hit, we have 43% exact matches, 68% positive matches, and 7% in gaps. The above example demonstrates the efficiency and usefulness of BLAST in biological enquiries.

## Question 2

The likelihood function of the given problem is

$$L(\theta) = (0.2+\theta)^{u_1}(0.3+\theta)^{u_2}(0.5-2\theta)^{u_3},$$

and the log-likelihood function is

$$l(\theta) = u_1 \log(0.2+\theta) + u_2 \log(0.3+\theta) + u_3 \log(0.5-2\theta).$$

The maximum likelihood estimate $\hat{\theta}$ is satisfies when

$$0 = l'(\hat{\theta}) = \frac{u_1}{0.2+\hat{\theta}} + \frac{u_2}{0.3+\hat{\theta}} - \frac{2u_3}{0.5-2\hat{\theta}} \tag{1}$$

Clearing the denominators gives

$$\begin{aligned} 0 &= u_1(0.3+\hat{\theta})(0.5-2\hat{\theta}) + u_2(0.2+\hat{\theta})(0.5-2\hat{\theta}) - 2u_3(0.2+\hat{\theta})(0.3+\hat{\theta}) \\ &= -2(u_1+u_2+u_3)\hat{\theta}^2 + (-0.1u_1+0.1u_2-u_3)\hat{\theta} + (0.15u_1+0.10u_2-0.12u_3) \end{aligned}$$

The solution to this quadratic equation is

$$\hat{\theta} = \frac{\mathbf{b}\cdot\mathbf{u} \pm \sqrt{\mathbf{u}^T\mathbf{A}\mathbf{u}}}{\mathbf{c}\cdot\mathbf{u}} \tag{2}$$

where

$$\mathbf{A} = \begin{bmatrix} 1.21 & 0.99 & 0.02 \\ 0.99 & 0.81 & 0.02 \\ 0.02 & 0.02 & 0.04 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -0.1 \\ 0.1 \\ 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$$

It remains for us to determine which of the above two possible solutions for $\hat{\theta}$ lies in the natural parameter space of the linear model (i.e. $\hat{\theta}$ satisfies $0 \le f_i(\hat{\theta}) \le 1$ for $i = 1,2,3$). We consider the hyperplane arrangement $\{f_i = 0\}_{i\in[3]}$ in $\mathbb{R}$, and note that the above inequalities imply that $\hat{\theta}$ must lie in the bounded region $-0.2 < \hat{\theta} < 0.25$. Varchenko's Formula tells us that equation (1) has precisely one root $\hat{\theta}_0$ lying in this bounded region and also another root $\hat{\theta}_1$ lying in the bounded region $-0.3 < \hat{\theta}_1 < 0.2$. Thus, $\hat{\theta}_1 < \hat{\theta}_0$ and they are the two roots in equation (2). Now, $\mathbf{c}\cdot\mathbf{u}$ is positive so

$$\frac{\mathbf{b}\cdot\mathbf{u} - \sqrt{\mathbf{u}^T\mathbf{A}\mathbf{u}}}{\mathbf{c}\cdot\mathbf{u}} < \frac{\mathbf{b}\cdot\mathbf{u} + \sqrt{\mathbf{u}^T\mathbf{A}\mathbf{u}}}{\mathbf{c}\cdot\mathbf{u}}$$

Therefore, we must have

$$\hat{\theta} = \frac{\mathbf{b}\cdot\mathbf{u} + \sqrt{\mathbf{u}^T\mathbf{A}\mathbf{u}}}{\mathbf{c}\cdot\mathbf{u}}$$

as the desired maximum likelihood estimate.

## Question 3

To prove Varchenko's Formula for $d = 2, m = 4$, we need to show that:

1. There is exactly one real solution in each bounded region of the hyperplane arrangement $\{f_i = 0\}_{i\in[4]}$.
2. There are no real solutions in each unbounded region of the hyperplane arrangement.
3. Each solution has multiplicity one.
4. The imaginary part of each solution is 0.

Proof of 1 and 2:

Given a region $R$ of the hyperplane arrangement, we want to find the number of distinct real solutions of the likelihood equations in (1.23) in $R$. Now, if any of the linear maps $f_i(\theta_1,\theta_2) = a_{i1}\theta_1 + a_{i2}\theta_2 + b_i$, $i \in [4]$ is negative in $R$, we may set $a'_{i1} = -a_{i1}$, $a'_{i2} = -a_{i2}$ and $b'_i = -b_i$, consider the likelihood equations (1.23) with these new coefficients and note that the equations remain unchanged. Thus, we may assume that all the maps $f_i$, $i \in [4]$ are positive in $R$, and define the log-likelihood function

$$l(\theta_1,\theta_2) = \sum_{i=1}^{4} u_i \log(a_{i1}\theta_1 + a_{i2}\theta_2 + b_i).$$

The solutions of the likelihood equations (1.23) are precisely the critical points of $l(\theta_1, \theta_2)$.

Suppose the given region $R$ is bounded. Since the $u_i$ are positive and the matrix $\mathbf{A} = (a_{ij})$ is of rank 2, Proposition 1.4 says that the log-likelihood function $l(\theta_1, \theta_2)$ is strictly concave on $R$. Hence, either $l(\theta_1, \theta_2)$ attains its unique maximum on the interior of $R$, or it tends to its supremum on the boundary $\partial R$ of $R$. Now, as $(\theta_1, \theta_2)$ tends to any point on $\partial R$, one or some of the linear maps $f_i$ will tend to zero while the other linear maps are bounded (since $R$ is bounded). Hence, $l(\theta_1, \theta_2) = \sum u_i \log(f_i)$ tends to $-\infty$ as $(\theta_1, \theta_2)$ approaches $\partial R$, so $l(\theta_1, \theta_2)$ cannot tend to its supremum on $\partial R$. Therefore, $l(\theta_1, \theta_2)$ has a unique critical point in $R$, implying that the likelihood equations (1.23) have exactly one real solution in $R$.

Suppose the given region $R$ is unbounded. To show that the likelihood equations (1.23) do not have any solutions in $R$, it suffices to show that the strictly concave function $l(\theta_1, \theta_2) = \sum u_i \log(f_i)$ is unbounded in $R$ and thus does not have any critical points in $R$. To prove that, it suffices to show that there exists a subset $S \subseteq R$ in which one of the maps $f_i$ is unbounded while the other maps $f_j, i \neq j$ are bounded below. We first show that there is some $\epsilon > 0$ such that the set

$$S = \{(\theta_1, \theta_2) \in R | f_i(\theta_1, \theta_2) > \epsilon \quad \forall i \in [4]\}$$

is non-empty and unbounded. Indeed, since $R$ is unbounded, two of the one-dimensional faces of $R$ must be unbounded. WLOG, let these two faces lie on the lines $l_1$ and $l_2$ defined by the equations $f_1 = 0$ and $f_2 = 0$ respectively. Consider the line $l_3$ defined by $f_1 = f_2$. In the case when $l_1$ and $l_2$ are parallel, $l_3$ lies right in the middle between the two lines. In the case when $l_1$ and $l_2$ are not parallel, $l_3$ is an angle bisector of the two lines. In both cases, $l_3$ intersects $R$ and thus cannot coincide with any of the other lines $f_i = 0$, $i = 3, 4$. Now, pick $2\delta > 0$ to be the smallest distance of $l_3$ to any other line $f_i = 0$, $i \in [4]$ parallel to $l_3$. Next, let $S$ be the set of points on $l_3$ that is at least a distance of $\delta$ away from the other lines $f_i = 0$ not parallel to $l_3$. $S$ is non-empty and unbounded, because it is the result of subtracting from the unbounded $l_3$ large-enough bounded neighborhoods of points of intersection of $l_3$ with the other lines. Finally, some easy linear algebra shows that the minimum distance $\delta$ of $S$ from the lines $f_i = 0$ corresponds to some $\epsilon > 0$ such that $f_i(\theta_1, \theta_2) > \epsilon$ for all $i \in [4]$ and $(\theta_1, \theta_2) \in S$.

Next, we show that one of the maps $f_i$ is unbounded on $S$. Suppose on the contrary that all the maps $f_i$ are bounded on $S$. Then, S must lie in the intersection of the regions $\{(\theta_1, \theta_2) \in R| - M_i < f_i(\theta_1, \theta_2) < M_i\}$ where the $M_i$ are the bounds on the respective $f_i$. Each region is a strip parallel to the line $f_i = 0$. Now, the matrix $\mathbf{A}$ is of rank 2, so there are two lines $f_i = 0$ and $f_j = 0$, $i \neq j$ which are not parallel to each other. Hence, the intersection of the strips $-M_i < f_i < M_i$ and $-M_j < f_j < M_j$ is a parallelogram that is bounded. This contradicts the fact that the unbounded $S$ lies in that intersection. We have shown that the $f_i$ are bounded below on $S \subseteq R$ and one of the $f_i$ is unbounded on $S$. This completes the proof that the likelihood equations (1.23) do not have any solutions in $R$.

Proof of 3:

A solution $(\hat{\theta}_1, \hat{\theta}_2)$ to the likelihood equations (1.23) has multiplicity more than one if and only if the Hessian matrix (1.22) is zero at $(\hat{\theta}_1, \hat{\theta}_2)$. It was shown in Proposition 1.4 that the eigenvalues of the Hessian are strictly negative if all the $u_i$ are strictly positive. Hence, the Hessian cannot be zero, and the result follows.

Proof of 4:

First, we show that $\text{Im}(f_i(\hat{\theta}_1, \hat{\theta}_2)) = 0 \quad \forall i \in [4]$ if $(\hat{\theta}_1, \hat{\theta}_2)$ is a solution of the likelihood equations (1.23).

$$\begin{aligned}
\sum_{i=1}^{4} u_i \frac{\text{Im}(f_i)\bar{f}_i}{|f_i|^2} &= \sum_{i=1}^{4} u_i \frac{\text{Im}(f_i)}{f_i} \\
&= \frac{1}{2}\sum_{i=1}^{4} u_i \frac{f_i - \bar{f}_i}{f_i} \\
&= \frac{1}{2}\sum_{i=1}^{4} u_i \frac{(a_{i1}\hat{\theta}_1 + a_{i2}\hat{\theta}_2 + b_i) - (a_{i1}\bar{\hat{\theta}}_1 + a_{i2}\bar{\hat{\theta}}_2 + b_i)}{f_i} \\
&= \frac{1}{2}(\sum_{i=1}^{4} u_i \frac{a_{i1}}{f_i})(\hat{\theta}_1 - \bar{\hat{\theta}}_1) + \frac{1}{2}(\sum_{i=1}^{4} u_i \frac{a_{i2}}{f_i})(\hat{\theta}_2 - \bar{\hat{\theta}}_2) \\
&= 0
\end{aligned}$$

since $\sum u_i \frac{a_{i1}}{f_i} = \sum u_i \frac{a_{i2}}{f_i} = 0$ by the likelihood equations. Hence,

$$0 = \text{Im}(\sum_{i=1}^{4} u_i \frac{\text{Im}(f_i)\bar{f}_i}{|f_i|^2}) = -\sum_{i=1}^{4} u_i \frac{\text{Im}(f_i)^2}{|f_i|^2}.$$

Because $u_i > 0$ and $|f_i|^2 > 0$ for all $i \in [4]$, we must have $\text{Im}(f_i) = 0 \quad \forall i \in [4]$. Therefore,

$$0 = \text{Im}\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \text{Im}\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}\right) = \mathbf{A}\text{Im}\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$$

Since $\mathbf{A}$ is a rank 2 matrix, we get $\text{Im}(\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}) = 0$, so $(\hat{\theta}_1, \hat{\theta}_2)$ is a purely real solution.

## Question 4

By Proposition 1.9, the maximum likelihood estimate $\hat{\mathbf{p}}$ satisfies

$$\mathbf{A}\hat{\mathbf{p}} = \frac{1}{N}\mathbf{b} = \frac{1}{41}\begin{bmatrix} 12 \\ 28 \\ 42 \end{bmatrix}.$$

The solution of the above linear equations is given by

$$\hat{\mathbf{p}} = \frac{1}{41}\begin{bmatrix} -29 \\ 28 \\ 42 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x\begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + y\begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + z\begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{v}_0 + x\mathbf{v}_1 + y\mathbf{v}_2 + z\mathbf{v}_3 \tag{3}$$

where $x, y, z \in \mathbb{R}$. As $x, y, z$ varies over $\mathbb{R}$, we get the relatively open polytope $P_{\mathbf{A}}(b)$. Since $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ lie in the kernel of $\mathbf{A}$, and by equation (1.32), i.e. $\sum u_1 \log(\hat{p}_i) = 0$ for all $\mathbf{u} \in \ker(\mathbf{A})$, we have the relations

$$\hat{p}_1\hat{p}_4 = \hat{p}_2^2, \quad \hat{p}_1\hat{p}_5 = \hat{p}_2\hat{p}_3, \quad \hat{p}_1\hat{p}_6 = \hat{p}_3^2.$$

Substituting the formulas from equation (3) for each $\hat{p}_i$ in terms of $x, y, z$ into the above relations, we obtain 3 quadratic equations in 3 variables $x, y, z$ which can be solved.

Alternatively, we can use the fact from Theorem 1.10 that $\hat{\mathbf{p}}$ maximizes the strictly concave entropy function $H(\mathbf{p}) = -\sum p_i \log(p_i)$ over $P_{\mathbf{A}}(b)$. Substituting the formulas for each $p_i$ in terms of $x, y, z$, we get a function $H : \mathbb{R}^3 \to \mathbb{R}$ in terms of $x, y, z$ which we hope to maximize over $\mathbb{R}^3$. A simple hill-climbing algorithm gives us the optimal parameters

$$x = 0.176817$$
$$y = 0.242082$$
$$z = 0.331446$$

which corresponds to the estimate

$$\hat{(\mathbf{p})} = \begin{bmatrix} 0.0430 \\ 0.0872 \\ 0.1194 \\ 0.1768 \\ 0.2421 \\ 0.3314 \end{bmatrix}$$

By Birch's Theorem, this is the solution for the maximum likelihood estimate of our toric model.

## Question 5

Let $\lambda_h$, $\lambda_t$ be the probabilities that the gambler's first coin comes up heads and tails respectively. Let $\rho_h$, $\rho_t$ be the corresponding probabilities for his second coin. Let $\theta$ be the probability he picks the first coin. Then, we have the vector of parameters $\pi = (\theta, \lambda_h, \lambda_t, \rho_h, \rho_t)$ satisfying $\lambda_h + \lambda_t = 1$ and $\rho_h + \rho_t = 1$, so there are a total of three free parameters with the parameter space

$$\Theta = \Delta_1 \times \Delta_1 \times \Delta_1.$$

The model is given by the polynomial map

$$\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^5, \quad \pi \longmapsto (f_i)_{0 \le i \le 4}, \quad f_i = \binom{4}{i}\theta\lambda_h^i\lambda_t^{4-i} + \binom{4}{i}(1-\theta)\rho_h^i\rho_t^{4-i}.$$

Our goal is to find estimates $\hat{\pi}$ which maximize the log-likelihood function

$$l_{\text{obs}}(\pi) = \sum_{i=0}^{4} u_i \log(f_i(\pi))$$

To test our hypothesis for some $\mathbf{u} = (u_0, \ldots, u_4)$, we compare the theoretical probability distribution $\mathbf{f}(\hat{\pi}) = (f_0(\hat{\pi}), \ldots, f_4(\hat{\pi}))$ with the observed distribution $\frac{1}{1000}\mathbf{u} = (\frac{1}{1000}u_0, \ldots, \frac{1}{1000}u_4)$. If the two distributions are close, we can conclude that our hypothesis explains the observed data well.

## Question 6

We define the hidden data by decomposing the observed data into the contributions made by each of the gambler's two coins, i.e. $u_i = u_{i1} + u_{i2}$ for $0 \le i \le 4$. Here, $u_{i1}$ is the number of times the gambler's first coin produced $i$ heads when picked and thrown 4 times. Similarly, $u_{i2}$ is defined for the second coin. The hidden model is given by

$$F : \mathbb{R}^3 \to \mathbb{R}^{2\times 5}, \quad \pi \longmapsto (f_{i1}, f_{i2})_{0 \le i \le 4}, \quad f_{i1} = \binom{4}{i}\theta\lambda_h^i\lambda_t^{4-i}, \quad f_{i2} = \binom{4}{i}(1-\theta)\rho_h^i\rho_t^{4-i}.$$

First, we solve the problem of maximizing the hidden log-likelihood function

$$\begin{aligned} l_{\text{hid}}(\pi) &= \sum_{i=0}^{4} u_{i1}\log(\binom{4}{i}\theta\lambda_h^i\lambda_t^{4-i}) + \sum_{i=0}^{4} u_{i2}\log(\binom{4}{i}(1-\theta)\rho_h^i\rho_t^{4-i}) \\ &= \alpha_1\log(\lambda_h) + \alpha_2\log(1-\lambda_h) + \beta_1\log(\rho_h) + \beta_2\log(1-\rho_h) \\ &\quad +\gamma_1\log(\theta) + \gamma_2\log(1-\theta) + c \end{aligned}$$

where

$$\begin{aligned} \alpha_1 &= \textstyle\sum_{i=0}^{4} u_{i1} i, \quad \alpha_2 = \sum_{i=0}^{4} u_{i1}(4-i), \\ \beta_1 &= \textstyle\sum_{i=0}^{4} u_{i2} i, \quad \beta_2 = \sum_{i=0}^{4} u_{i2}(4-i), \\ \gamma_1 &= \textstyle\sum_{i=0}^{4} u_{i1}, \quad \gamma_2 = \sum_{i=0}^{4} u_{i2} \\ c &= \textstyle\sum_{i=0}^{4} u_i \log(\binom{4}{i}). \end{aligned}$$

It is useful to note that $\alpha_1+\alpha_2 = 4\gamma_1$, $\beta_1+\beta_2=\gamma_2$, $\gamma_1+\gamma_2 = 1000$. Now, the function $g(x) = a\log(x) + b\log(1-x)$, where $a, b$ are constants, is maximized when

$$0 = g'(x) = \frac{a}{x} - \frac{b}{1-x} \quad \Longrightarrow \quad x = \frac{a}{a+b}.$$

Hence, given the hidden data $(u_{i1}), (u_{i2})$, the hidden log-likelihood function is maximized when

$$\lambda_h = \frac{\alpha_1}{\alpha_1+\alpha_2}, \quad \rho_h = \frac{\beta_1}{\beta_1+\beta_2}, \quad \theta = \frac{\gamma_1}{\gamma_1+\gamma_2} = \frac{\gamma_1}{1000}.$$

To summarize, the EM-algorithm for this problem is

**Input:** The observed data $\mathbf{u} = (u_0, \ldots, u_4)$.

**Output:** A proposed maximum $\hat{\pi} \in \Theta$ of the log-likelihood function $l_{\text{obs}}(\pi)$.

**Step 0:** Pick initial parameters $\theta, \lambda_h, \rho_h$. Our vector of parameters is now

$$\pi = (\theta, \lambda_h, 1-\lambda_h, \rho_h, 1-\rho_h).$$

**E-Step:** Define the hidden data

$$u_{i1} = u_i \frac{\binom{4}{i}\theta\lambda_h^i\lambda_t^{4-i}}{\binom{4}{i}\theta\lambda_h^i\lambda_t^{4-i} + \binom{4}{i}(1-\theta)\rho_h^i\rho_t^{4-i}}$$
$$u_{i2} = u_i \frac{\binom{4}{i}(1-\theta)\rho_h^i\rho_t^{4-i}}{\binom{4}{i}\theta\lambda_h^i\lambda_t^{4-i} + \binom{4}{i}(1-\theta)\rho_h^i\rho_t^{4-i}}$$

**M-Step:** Calculate the maximum likelihood estimate for the hidden model

$$\lambda_h^* = \frac{\sum_{i=0}^4 u_{i1} i}{4\sum_{i=0}^4 u_{i1}}, \quad \rho_h^* = \frac{\sum_{i=0}^4 u_{i2} i}{4\sum_{i=0}^4 u_{i2}}, \quad \theta^* = \frac{\sum_{i=0}^4 u_{i1}}{1000}$$

**Step 3:** If $l_{\text{obs}}(\pi^*) - l_{\text{obs}}(\pi) > \epsilon$, then set $\pi := \pi^*$ and go back to E-Step.

**Step 4:** Output the parameter vector $\hat{\pi} := \pi^*$ and the corresponding probability distribution $\hat{\mathbf{p}} = \mathbf{f}(\hat{\pi})$.

We ran the algorithm on the observed data $\mathbf{u} = (150, 150, 150, 350, 200)$ and tried different starting parameters. The listing below shows some of the parameters used and the estimates obtained.

$$\begin{array}{rll}
(\theta, \lambda_h, \rho_h) = (0.5, 0.5, 0.5) & \mathbf{f}(\hat{\pi}) \approx \frac{1}{1000}(33, 177, 358, 323, 109) & l_{\text{obs}}(\hat{\pi}) = -1765.5 \\
(\theta, \lambda_h, \rho_h) = (0.8, 0.8, 0.8) & \mathbf{f}(\hat{\pi}) \approx \frac{1}{1000}(33, 177, 358, 323, 109) & l_{\text{obs}}(\hat{\pi}) = -1765.5 \\
(\theta, \lambda_h, \rho_h) = (0.1, 0.1, 0.1) & \mathbf{f}(\hat{\pi}) \approx \frac{1}{1000}(33, 177, 358, 323, 109) & l_{\text{obs}}(\hat{\pi}) = -1765.5 \\
(\theta, \lambda_h, \rho_h) = (0.3, 0.6, 0.6) & \mathbf{f}(\hat{\pi}) \approx \frac{1}{1000}(33, 177, 358, 323, 109) & l_{\text{obs}}(\hat{\pi}) = -1765.5 \\
(\theta, \lambda_h, \rho_h) = (0.3, 0.599, 0.6) & \mathbf{f}(\hat{\pi}) \approx \frac{1}{1000}(154, 140, 181, 304, 221) & l_{\text{obs}}(\hat{\pi}) = -1550.6 \\
(\theta, \lambda_h, \rho_h) = (0.25, 0.65, 0.35) & \mathbf{f}(\hat{\pi}) \approx \frac{1}{1000}(154, 140, 181, 304, 221) & l_{\text{obs}}(\hat{\pi}) = -1550.6 \\
(\theta, \lambda_h, \rho_h) = (0.90, 0.15, 0.45) & \mathbf{f}(\hat{\pi}) \approx \frac{1}{1000}(154, 140, 181, 304, 221) & l_{\text{obs}}(\hat{\pi}) = -1550.6
\end{array}$$

It seems that when $\lambda_h = \rho_h$, the estimate with $l_{\text{obs}}(\hat{\pi}) = -1765.5$ is obtained; and for every other case, the estimate with $l_{\text{obs}}(\hat{\pi}) = -1550.6$ is obtained. A trial running the algorithm with 1000 different random initial parameters could not produce any result better than the latter estimate.