02-712 Biological Modeling and Simulations: Homework 2

Thomas Zhang

September 2023

Question 1

a)

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{n+1,1} & c_{n+1,2} & \cdots & c_{n+1,n} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ g_0 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \\ g_{n+1} \end{bmatrix}$$

$c_{i,j}$ is the $j'th$ condition during the $i'th$ experiment. x_i is the contribution that environment condition c_i contributes to the overall growth rate. g_i is the growth rate of the $i'th$ experiment. The system is full rank since there are $n+1$ weights to learn and there are $n+1$ equations. Since the matrix is full rank, an iterative solving algorithm like Krylov-Subspace can be used.

b)

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{2n,1} & c_{2n,2} & \cdots & c_{2n,n} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ g_0 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \\ g_{2n} \end{bmatrix}$$

$c_{i,j}$ is the $j'th$ condition during the $i'th$ experiment. x_i is the contribution that environment condition c_i contributes to the overall growth rate. g_i is the growth rate of the $i'th$ experiment. The system is over determined $\forall \quad n > 1$ since there are $n+1$ weights to learn and there are $2n$ equations. Since the system is over determined, there are multiple solutions, so we can use least squares regression to find a solution that minimizes $\sum_i a_i x - g_i$ where a_i is one row of the matrix of conditions $[c_{1,1}, c1, 2, ...]$. This can be solved by solving the solution for $A^T A x = A^T g$.

c)

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} & c_{1,1}c_{1,1} & c_{1,1}c_{1,2} & \cdots & c_{1,n-1}c_{1,n} & c_{1,n}c_{1,n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ c_{2n,1} & c_{2n,2} & \cdots & c_{2n,n} & c_{2n,1}c_{2n,1} & c_{2n,1}c_{2n,2} & \cdots & c_{2n,n-1}c_{2n,n} & c_{2n,n}c_{2n,n} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{\frac{n^2-3n}{2}} \\ g_0 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \\ g_{2n} \end{bmatrix}$$

$c_{i,j}$ is the $j'th$ condition during the $i'th$ experiment. x_i is the contribution that environment condition c_i contributes to the overall growth rate. g_i is the growth rate of the $i'th$ experiment. The system is under determined since there are $1 + \frac{n^2-3n}{2}$ weights to learn and there are $2n$ equations. Since the system is under determined there are infinite solutions. So one approach that we can use to solve for x is to find the pseudo-inverse of A, which is denoted as $\bar{A}$ and then calculate $\bar{A}b = x$. The pseudo inverse of A is $Q_2\bar{\Sigma}Q_1^T$ where $\bar{\Sigma}$ the negative transpose of the singular value matrix Σ of A, Q_1 is the left singular matrix of A and Q_2 is the right singular matrix of A ($A = Q_1\Sigma Q_2^T$).

c)

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} & c_{1,1}c_{1,1} & c_{1,1}c_{1,2} & \cdots & c_{1,n-1}c_{1,n} & c_{1,n}c_{1,n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ c_{n^2,1} & c_{n^2,2} & \cdots & c_{n^2,n} & c_{n^2,1}c_{n^2,1} & c_{n^2,1}c_{n^2,2} & \cdots & c_{n^2,n-1}c_{n^2,n} & c_{n^2,n}c_{n^2,n} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{\frac{n^2-3n}{2}} \\ g_0 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \\ g_{n^2} \end{bmatrix}$$

$c_{i,j}$ is the $j'th$ condition during the $i'th$ experiment. x_i is the contribution that environment condition c_i contributes to the overall growth rate. g_i is the growth rate of the $i'th$ experiment. The system is over determined since there are $1 + \frac{n^2-3n}{2}$ weights to learn and there are n^2 equations and $n^2 > 1 + \frac{n^2-3n}{2}$. Since the system is over determined, there are multiple solutions, so we can use least squares regression to find a solution that minimizes $\sum_i a_i x - g_i$ where a_i is one row of the matrix of conditions $[c_{1,1}, c1, 2, ...]$. This can be solved by solving the solution for $A^T A x = A^T g$.

Question 2

$$f(x) = \frac{4}{x^7} - \frac{3}{x^2} + 0.1$$

a)

$$\begin{aligned} f(x_{max}) = f(1) &= \frac{4}{1^7} - \frac{3}{1^2} + 0.1 = 1.1 \\ f(x_{mid}) = f(2) &= \frac{4}{2^7} - \frac{3}{2^2} + 0.1 = -0.61875 \\ f(x_{min}) = f(3) &= \frac{4}{3^7} - \frac{3}{3^2} + 0.1 = -0.23150434385 \end{aligned} \tag{1}$$

$$\begin{aligned} f(x_{max}) = f(1) &= \frac{4}{1^7} - \frac{3}{1^2} + 0.1 = 1.1 \\ f(x_{mid}) = f(1.5) &= \frac{4}{1.5^7} - \frac{3}{1.5^2} + 0.1 = -0.99922267947 \\ f(x_{min}) = f(2) &= \frac{4}{2^7} - \frac{3}{2^2} + 0.1 = -0.61875 \end{aligned} \tag{2}$$

$$\begin{aligned} f(x_{max}) = f(1) &= \frac{4}{1^7} - \frac{3}{1^2} + 0.1 = 1.1 \\ f(x_{mid}) = f(1.25) &= \frac{4}{1.25^7} - \frac{3}{1.25^2} + 0.1 = -0.9811392 \\ f(x_{min}) = f(1.5) &= \frac{4}{2^7} - \frac{3}{2^2} + 0.1 = -0.99922267947 \end{aligned} \tag{3}$$

The backwards error is 0.5 and the forwards error is 0.9811392 after the 3^{rd} bisection.

b)

$$\begin{aligned} f(x_{max}) = f(1) &= \frac{4}{1^7} - \frac{3}{1^2} + 0.1 = 1.1 \\ f(x_{mid}) = f(2.65226648352) &= \frac{4}{2^7} - \frac{3}{2^2} + 0.1 = -0.322135968884 \\ f(x_{min}) = f(3) &= \frac{4}{3^7} - \frac{3}{3^2} + 0.1 = -0.23150434385 \end{aligned} \tag{4}$$

$$\begin{aligned} f(x_{max}) = f(1) &= \frac{4}{1^7} - \frac{3}{1^2} + 0.1 = 1.1 \\ f(x_{mid}) = f(2.2780023652) &= \frac{4}{1.5^7} - \frac{3}{1.5^2} + 0.1 = -0.465547707306 \\ f(x_{min}) = f(2.65226648352) &= \frac{4}{2^7} - \frac{3}{2^2} + 0.1 = -0.322135968884 \end{aligned} \tag{5}$$

$$\begin{aligned} f(x_{max}) = f(1) &= \frac{4}{1^7} - \frac{3}{1^2} + 0.1 = 1.1 \\ f(x_{mid}) = f(1.89796216057) &= \frac{4}{1.25^7} - \frac{3}{1.25^2} + 0.1 = -0.687723865076 \\ f(x_{min}) = f(2.2780023652) &= \frac{4}{1.5^7} - \frac{3}{1.5^2} + 0.1 = -0.465547707306 \end{aligned} \tag{6}$$

The backwards error is 2.2780023652 and the forwards error is 0.687723865076

c)

$$f'(x) = -7(\frac{4}{x^8}) + 2(\frac{3}{x^3}) + 0 = \frac{21}{x^8} + \frac{6}{x^3}$$

$$f(1) = \frac{4}{1^7} - \frac{3}{1^2} + 0.1 = 1.1 \tag{7}$$

$$f(1.05) = \frac{4}{1.05^7} - \frac{3}{1.05^2} + 0.1 = 0.221636885146 \tag{8}$$

$$f(1.06609) = \frac{4}{1.06609^7} - \frac{3}{1.06609^2} + 0.1 = 0.0159992621509 \tag{9}$$

$$f(1.06745) = \frac{4}{1.06745^7} - \frac{3}{1.06745^2} + 0.1 = 0.000102031079136 \tag{10}$$

The estimate of backward error is $1.06609 - 1.06754 = 0.00145$ and the forward error is 0.000102

Question 3

a)

$$\begin{aligned} 2a + b + 3c &\leq 10 \\ a + b + 2c &\leq 7 \\ 3a + b + c &\leq 12 \\ a &\geq 0 \\ b &\geq 0 \\ c &\geq 0 \end{aligned}$$

$$\underset{a,b,c}{\text{argmax}} \quad f(a,b,c) = a + 2b + 5c$$

b)

$$\begin{aligned} 2a + b + 3c + k_1 &= 10 \\ a + b + 2c + k_2 &= 7 \\ 3a + b + c + k_3 &= 12 \\ a &\geq 0 \\ b &\geq 0 \\ c &\geq 0 \\ k_1 &\geq 0 \\ k_2 &\geq 0 \\ k_3 &\geq 0 \end{aligned}$$

$$\underset{a,b,c}{\text{argmin}} \quad g(a,b,c,k_1,k_2,k_3) = -f(a,b,c) = -a - 2b - 5c$$

c) Let $a = b = c = 0$

$$k_1 = 10 \quad k_2 = 7 \quad k_3 = 12 \quad g(a,b,c,k_1,k_2,k_3) = 0$$

Increase c since it causes $g(a,b,c)$ to drop the fastest.

$$\begin{aligned} k_1 &= 10 - 3c \\ k_2 &= 7 - 2c \\ k_3 &= 12 - c \end{aligned}$$

Set LHS to 0 for each of the constraints to find how much c can be increased before some other variable gets set to 0.

$$\begin{aligned} 0 &= 10 - 3c & c &= \frac{10}{3} \\ 0 &= 7 - 2c & c &= \frac{7}{2} \\ 0 &= 12 - c & c &= 12 \end{aligned}$$

$c = \frac{10}{3}$ is the minimum value of c before some other variable becomes 0, so we set $c = \frac{10}{3}$ which makes $k_1 = 0$. Next we rearrange the constraint with k_1 so that c is expressed in terms of variables that have been set to 0 and substitute the expression into the other constraints where there is a c.

$$\begin{aligned} c &= \frac{1}{3}(10 - 2a - b - k_1) \\ k_2 &= 7 - \frac{2}{3}(10 - 2a - b - k_1) - a - b \\ &= \frac{1}{3} + \frac{1}{3}a - \frac{1}{3}b = \frac{1}{3} \\ k_3 &= 12 - \frac{1}{3}(10 - 2a - b - k_1) - 3a - b \\ &= \frac{26}{3} - \frac{7}{3}a - \frac{2}{3}b = \frac{26}{3} \\ g(a,b,c,k_1,k_2,k_3) &= -a - 2b - \frac{5}{3}(10 - 2a - b - k_1) \\ &= \frac{7}{3}a - \frac{1}{3}b + \frac{5}{3}k_1 - \frac{50}{3} = -\frac{50}{3} \end{aligned}$$

g decreases fastest with b so b is the next variable to be increased. As before find the minimum amount b can be increased before some other parameter becomes 0.

$$0 = \frac{1}{3}(10 - 2a - b - k_1) \qquad b = 10 - 2a - k_1 = 10$$
$$0 = \frac{1}{3} + \frac{1}{3}a - \frac{1}{3}b \qquad b = 1 + a = 1$$
$$0 = \frac{26}{3} - \frac{7}{3}a - \frac{2}{3}b \qquad b = \frac{1}{2}(26 - 7a) = 13$$

$b = 1$ is the minimum. So we set $b = 1$ which makes $k_2 = 0$.

$$\begin{aligned}
b &= 1 + a - 3k_2 \\
c &= \frac{1}{3}(10 - 2a - (1 + a - 3k_2) - k_1) \\
&= \frac{1}{3}(9 - 3a - k_1 + 3k_2) = 3 \\
k_3 &= \frac{26}{3} - \frac{7}{3}a - \frac{2}{3}(1 + a - 3k_2) \\
&= 8 - 3a + 2k_2 = 8 \\
g(a, b, c, k_1, k_2, k_3) &= \frac{7}{3}a - \frac{1}{3}(1 + a - 3k_2) + \frac{5}{3}k_1 - \frac{50}{3} \\
&= 2a + \frac{5}{3}k_1 + k_2 - 17 = -17
\end{aligned}$$

Now no variable can decrease g so the solution is:

$$a = 0 \quad b = 1 \quad c = 3 \quad k_1 = 0 \quad k_2 = 0 \quad k_3 = 8$$

d) The optimal solution is to perform single-cell RNA-seq 0 times, $a = 0$, perform bulk RNA-seq once, $b = 1$, and perform spatial RNA-seq 3 times $c = 3$. The utility of this solution is 17

e) This solution would use \$10,000, take 7 days of time, and require 4 grams of tissue.

Question 4

a)

$$\begin{bmatrix} \frac{f(A,T,x+dx,y)-f(A,T,x-dx,y)}{2dx} \\ \\ \frac{f(A,T,x,y+dy)-f(A,T,x,y-dy)}{2dy} \end{bmatrix}$$

b)

$$\begin{bmatrix} \frac{d^2f}{dx^2} & \frac{d^2f}{dxdy} \\ \\ \frac{d^2f}{dydx} & \frac{d^2f}{dy^2} \end{bmatrix}$$

$$\begin{aligned}
\frac{df}{dx^2} &= \frac{f(A,T,x+2dx,y) - 2f(A,T,x,y) + f(A,T,x-2dx,y)}{4dx^2} \\
\frac{df}{dxdy} &= \frac{f(A,T,x+dx,y+dy) - f(A,T,x-dx,y+dy) - f(A,T,x+dx,y-dy) + f(A,T,x-dx,y-dy)}{4dxdy} \\
\frac{df}{dydx} &= \frac{f(A,T,x+dx,y+dy) - f(A,T,x-dx,y+dy) - f(A,T,x+dx,y-dy) + f(A,T,x-dx,y-dy)}{4dxdy} \\
\frac{df}{dy^2} &= \frac{f(A,T,x,y+2dy) - 2f(A,T,x,y) + f(A,T,x,y-2dy)}{4dy^2}
\end{aligned}$$

c) Newton Rathson Algorithm

Data: $\Delta x, \Delta y$ = increment size of 'finer' grid; dx, dy = step size for finite derivatives
$M, N \leftarrow \text{SHAPE}(A)$
$m, n \leftarrow \text{SHAPE}(T)$
best_score $\leftarrow \infty$
best_x, best_y $\leftarrow$ 0, 0
for $i \leftarrow 0$ **to** $(N - n)\Delta x$ **do**
  for $j \leftarrow 0$ **to** $(M - m)\Delta y$ **do**
    $x \leftarrow i \times dx$
    $y \leftarrow j \times dy$
    for $k \leftarrow 0$ **to** 10 **do**
      Grad $\leftarrow$ GRADIENT(A, T, x, y, dx, dy)
      Hess $\leftarrow$ HESSIAN(A, T, x, y, dx, dy)
      $x \leftarrow x + \text{Hess}^{-1}\text{Grad}[0]$
      $y \leftarrow y + \text{Hess}^{-1}\text{Grad}[1]$
    end
    if SCORE$(A, T, x, y) <$ best_score **then**
      $best\_score \leftarrow \text{SCORE}(A, T, x, y)$
      $best\_x \leftarrow x$
      $best\_y \leftarrow y$
    end
  end
end
return$(best_x, best_y)$

d, e) For test case 1 the best coordinates were $(4.499997017072263, 3.5)$ For test case 2 the best coordinates were $(4.876557112637574, 0.3019367812834329)$

Question 5

a)

$$a_i \overrightarrow{x} - b_i = c_i - d_i \quad \forall i \in \{1, 2, ..., m\}$$

$$\underset{x}{argmin} \quad f(c, d) = \sum_i c_i + d_i$$

$$c_i \geq 0 \quad \forall i \in \{1, 2, ..., m\}$$

$$d_i \geq 0 \quad \forall i \in \{1, 2, ..., m\}$$

m is the number of rows of the matrix A. a_i is the $i'th$ row of the matrix A. b_i is the $i'th$ element in vector b. An algorithm that can be used to solve this linear program is an iterative algorithm like Krylov-Subspace if the program is full rank. If the matrix is over determined then we can use least-squares optimization. Lastly if the matrix is under determined we can just solve for the pseudo-inverse and calculate $\bar{A}b$.

b)

$$x_i = c_i - d_i \quad \forall i \in \{1, 2, ..., m\}$$

$$\underset{x}{argmin} \quad f(x, c, d) = \sum_i (a_i \overrightarrow{x}_i - b_i)^2 + \lambda(c_i + d_i)$$

$$c_i \geq 0 \quad \forall i \in \{1, 2, ..., m\}$$

$$d_i \geq 0 \quad \forall i \in \{1, 2, ..., m\}$$

m is the number of rows of the matrix A. a_i is the $i'th$ row of the matrix A. b_i is the $i'th$ element in vector b. One algorithm that we can use to solve for $\overrightarrow{x}$ is the conjugate gradient algorithm which generally works with non-linear programs.

c) The least squares regressions optimization equation is $\|A\overrightarrow{x} - \overrightarrow{b}\|_2^2$. This is a convex function. The L1 regularization term is an absolute value, which is also a convex function. The objective function for part b, is the sum of two convex functions, and the sum of two convex functions is also convex. Thus the objective function for b which is:

$$f(x, c, d) = \sum_i (a_i \overrightarrow{x}_i - b_i)^2 + \lambda(c_i + d_i)$$

is also convex.