Biomath 210 Homework 1

Huwenbo Shi (603-778-363) shihuwenbo@ucla.edu

October 5, 2015

Problem 1.1

By the definition of MM algorithm, x_{n+1} minimizes $g(x|x_n)$, therefore

$$g(x_{n+1}|x_n) \leqslant g(x_n|x_n) = f(x_n), \tag{1}$$

where the equality follows from the definition of surrogate function. Since

$$g(x|x_{n-1}) \geqslant f(x) \tag{2}$$

for all x in the domain of $f(x)$ by the definition of surrogate function, then the following inequality must hold

$$g(x_n|x_{n-1}) \geqslant f(x_n). \tag{3}$$

Therefore, we conclude that

$$g(x_{n+1}|x_n) \leqslant f(x_n) \leqslant g(x_n|x_{n-1}), \tag{4}$$

and that the sequence $g(x_{n+1}|x_n)$ decreases.

Problem 1.2

Without loss of generality, assume $g(\mathbf{z}|\mathbf{y})$ is a majorization of $f(\mathbf{z})$. For $\mathbf{y} = \mathbf{x}$, let $h(\mathbf{z}) = g(\mathbf{z}|\mathbf{x}) - f(\mathbf{z})$. By the definition of surrogate function, $h(\mathbf{z}) \geqslant 0$, because $g(\mathbf{z}|\mathbf{x}) \geqslant f(\mathbf{z})$ for all $\mathbf{z}$ in the domain of f. When $\mathbf{z} = \mathbf{x}$, $h(\mathbf{z})$ attains its minimum, 0, because

$$h(\mathbf{x}) = g(\mathbf{x}|\mathbf{x}) - f(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}) = 0. \tag{5}$$

At the point $\mathbf{x}$,

$$\nabla h(\mathbf{x}) = \nabla g(\mathbf{x}|\mathbf{x}) - \nabla f(\mathbf{x}) = 0, \tag{6}$$

from which we conclude

$$\nabla g(\mathbf{x}|\mathbf{x}) = \nabla f(\mathbf{x}). \tag{7}$$

Similar approach can be used to show $\nabla g(\mathbf{x}|\mathbf{x}) = \nabla f(\mathbf{x})$ when $g(\mathbf{z}|\mathbf{y})$ is a minorization of $f(\mathbf{z})$. In this case, $h(\mathbf{z})$ attains its maximum at the point $\mathbf{x}$.

Problem 1.3

Assume $g(\mathbf{z}|\mathbf{y})$ minorizes $f(\mathbf{z})$. For $\mathbf{y} = \mathbf{z}$, the function $h(\mathbf{z}) = f(\mathbf{z}) - g(\mathbf{x}|\mathbf{z})$ attains its minimum at the point $\mathbf{x}$, at which $h(\mathbf{x}) = 0$ and $\nabla h(\mathbf{x}) = 0$. For any vector $\mathbf{v}$ and sufficient small t, we have

$$\begin{aligned}
0 &\leqslant h(\mathbf{x} + t\mathbf{v}) - h(\mathbf{x}) \\
&= \int_0^1 [\nabla h(\mathbf{x} + rt\mathbf{v}) - \nabla h(\mathbf{x})]^* t\mathbf{v} \; dr \\
&= \int_0^1 [\nabla f(\mathbf{x} + rt\mathbf{v}) - \nabla f(\mathbf{x}) - \nabla g(\mathbf{x} + rt\mathbf{v}|\mathbf{x}) + \nabla g(\mathbf{x}|\mathbf{x})]^* t\mathbf{v} \; dr \\
&= \int_0^1 [\nabla f(\mathbf{x} + rt\mathbf{v}) - \nabla f(\mathbf{x}) \; dr - \int_0^1 \nabla g(\mathbf{x} + rt\mathbf{v}|\mathbf{x}) - \nabla g(\mathbf{x}|\mathbf{x})]^* t\mathbf{v} \; dr \\
&= t^2 \int_0^1 [s_f^2(\mathbf{x} + rt\mathbf{v}, \mathbf{x}) r\mathbf{v}]^* \mathbf{v} \; dr - t^2 \int_0^1 [s_g^2(\mathbf{x} + rt\mathbf{v}, \mathbf{x}) r\mathbf{v}]^* \mathbf{v} \; dr.
\end{aligned} \tag{8}$$

Dividing both sides of the inequality by t^2, and sending t to 0 gives

$$\begin{aligned}
0 &\leqslant \int_0^1 [s_f^2(\mathbf{x}, \mathbf{x}) r\mathbf{v}]^* \mathbf{v} \; dr - \int_0^1 [s_g^2(\mathbf{x}, \mathbf{x}) r\mathbf{v}]^* \mathbf{v} \; dr \\
&= \mathbf{v}^* \left(\int_0^1 s_f^2(\mathbf{x}, \mathbf{x}) r \; dr \right) \mathbf{v} - \mathbf{v}^* \left(\int_0^1 s_g^2(\mathbf{x}, \mathbf{x}) r \; dr \right) \mathbf{v} \\
&= \mathbf{v}^* \left(d^2 f(\mathbf{x}) - d^2 g(\mathbf{x}|\mathbf{x}) \right) \mathbf{v}.
\end{aligned} \tag{9}$$

Therefore, the difference matrix $d^2 f(\mathbf{x}) - d^2 g(\mathbf{x}|\mathbf{x})$ is positive semidefinite.

Problem 1.6

Minimizing $f(\theta)$ is equivalent to minimizing

$$\frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - \mu - \alpha_i - \beta_j)^2. \tag{10}$$

With the constraints $\sum_{i=1}^{I} \alpha_i = 0$ and $\sum_{j=1}^{J} \beta_j = 0$, we get the Lagragian

$$\mathcal{L}(\theta, \lambda) = \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - \mu - \alpha_i - \beta_j)^2 + \lambda_1 \sum_{i=1}^{I} \alpha_i + \lambda_2 \sum_{j=1}^{J} \beta_j. \tag{11}$$

The partial derivative of $\mathcal{L}$ with respect to each variables are

$$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - \mu - \alpha_i - \beta_j), \tag{12}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_i} = \sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \mu - \alpha_i - \beta_j) + \lambda_1, \tag{13}$$

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^{I}\sum_{k=1}^{K}(y_{ijk} - \mu - \alpha_i - \beta_j) + \lambda_2. \tag{14}$$

Set Equation (12) to 0, we get

$$0 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \mu - \alpha_i - \beta_j), \tag{15}$$

$$\mu = -\frac{1}{I}\sum_{i=1}^{I}\alpha_i - \frac{1}{J}\sum_{j=1}^{J}\beta_j + \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}. \tag{16}$$

Set Equation (13) to 0, we get

$$\begin{aligned} 0 &= \sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \mu - \alpha_i - \beta_j) + \lambda_1 \\ &= -JK\mu - JK\alpha_i - K\sum_{j=1}^{J}\beta_j + \sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk} + \lambda_1 \end{aligned} \tag{17}$$

Sum over i on both sides, we get

$$\begin{aligned} 0 &= -IJK\mu - JK\sum_{i=1}^{I}\alpha_i - IK\sum_{j=1}^{J}\beta_j + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk} + I\lambda_1 \\ &= \left(JK\sum_{i=1}^{I}\alpha_i + IK\sum_{j=1}^{J}\beta_j - \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}\right) \\ &- IJK\mu - JK\sum_{i=1}^{I}\alpha_i - IK\sum_{j=1}^{J}\beta_j + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk} + I\lambda_1 \\ &= I\lambda_1. \end{aligned} \tag{18}$$

Therefore, we get $\lambda_1 = 0$. Applying similar approach, we get $\lambda_2 = 0$ as well.

By substituting $\lambda_1 = 0$ and $\lambda_2 = 0$ into the partial derivatives and setting them to 0, we arrive at the following equations

$$-K\sum_{i=1}^{I}\alpha_i - K\sum_{j=1}^{J}\beta_j + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \mu) = 0 \tag{19}$$

$$-K\sum_{j=1}^{J}\beta_j + \sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \mu - \alpha_i) = 0 \tag{20}$$

$$-K\sum_{i=1}^{I}\alpha_i+\sum_{i=1}^{I}\sum_{k=1}^{K}(y_{ijk}-\mu-\beta_j)=0 \tag{21}$$

The zero Lagrange multipliers implies that solving the problem with the constraint is equivalent to solving the problem without the constraint. In other words, the solution to the unconstrained problem always satisfies the constraint. Substituting $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$ into the three equations above gives the following equations

$$\begin{aligned}
&-IJK\mu+\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}y_{ijk}=0\\
&-JK\alpha_i+\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk}-\mu)=0\\
&-IK\beta_j+\sum_{i=1}^{I}\sum_{k=1}^{K}(y_{ijk}-\mu)=0
\end{aligned} \tag{22}$$

Solve for μ, α_i, and β_j in Equation (22) results in the formula in (1.6).

Problem 1.13

We maximize $L(\mathbf{p})$ with the constraints $p_A+p_B+p_O=1$ $\mathbf{p}$ and $\geqslant 0$. If we initialize the variables with non-negative values and keep the variables non-negative during the update, then we can consider only the equality constraint. Thus, we arrive at the Lagrangian

$$\begin{aligned}
\mathcal{L}(\mathbf{p},\lambda)=&\ x_A\ln(p_A^2+2p_Ap_O)+x_B\ln(p_B^2+2p_Bp_O)\\
&+x_{AB}\ln(2p_Ap_B)+x_O\ln p_O^2\\
&+\lambda(p_A+p_B+p_O)-\lambda.
\end{aligned} \tag{23}$$

Following (1.3) we find a minorization $g(\mathbf{p}|\mathbf{p}_n)$ for $\mathcal{L}(\mathbf{p},\lambda)$

$$\begin{aligned}
\mathcal{L}(\mathbf{p},\lambda)\geqslant&\ x_A\left[\frac{p_{n,A}^2}{p_{n,A}^2+2p_{n,A}p_{n,O}}\ln p_A^2+\frac{p_{n,A}p_{n,O}}{p_{n,A}^2+2p_{n,A}p_{n,O}}\ln p_Ap_O\right]\\
&+x_B\left[\frac{p_{n,B}^2}{p_{n,B}^2+2p_{n,B}p_{n,O}}\ln p_B^2+\frac{p_{n,B}p_{n,O}}{p_{n,B}^2+2p_{n,B}p_{n,O}}\ln p_Bp_O\right]\\
&+x_{AB}\left(\ln p_A+\ln p_B\right)+2x_O\ln p_O\\
&+\lambda(p_A+p_B+p_O)-\lambda+c_n,
\end{aligned} \tag{24}$$

where c_n is a variable dependent only on $\mathbf{p}$, x_A, x_B, x_{AB}, and x_O. Taking the partial derivative of $g(\mathbf{p}|\mathbf{p}_n)$ with respect to p_A, we get

$$\frac{\partial g}{\partial p_A}=\left(\frac{2x_Ap_{n,A}^2}{p_{n,A}^2+2p_{n,A}p_{n,O}}+\frac{x_Ap_{n,A}p_{n,O}}{p_{n,A}^2+2p_{n,A}p_{n,O}}+x_{AB}\right)\frac{1}{p_A}+\lambda. \tag{25}$$

For p_B we get

$$\frac{\partial g}{\partial p_B} = \left(\frac{2x_B p_{n,B}^2}{p_{n,B}^2 + 2p_{n,B}p_{n,O}} + \frac{x_B p_{n,B} p_{n,O}}{p_{n,B}^2 + 2p_{n,B}p_{n,O}} + x_{AB} \right) \frac{1}{p_B} + \lambda. \tag{26}$$

Lastly for p_O we get

$$\frac{\partial g}{\partial p_O} = \left(\frac{x_A p_{n,A} p_{n,O}}{p_{n,A}^2 + 2p_{n,A}p_{n,O}} + \frac{x_B p_{n,B} p_{n,O}}{p_{n,B}^2 + 2p_{n,B}p_{n,O}} + 2x_O \right) \frac{1}{p_O} + \lambda. \tag{27}$$

Set the partial derivatives to 0, we get

$$\begin{aligned} -\lambda p_A &= \left(\frac{2x_A p_{n,A}^2}{p_{n,A}^2 + 2p_{n,A}p_{n,O}} + \frac{x_A p_{n,A} p_{n,O}}{p_{n,A}^2 + 2p_{n,A}p_{n,O}} + x_{AB} \right) \\ -\lambda p_B &= \left(\frac{2x_B p_{n,B}^2}{p_{n,B}^2 + 2p_{n,B}p_{n,O}} + \frac{x_B p_{n,B} p_{n,O}}{p_{n,B}^2 + 2p_{n,B}p_{n,O}} + x_{AB} \right) \\ -\lambda p_O &= \left(\frac{x_A p_{n,A} p_{n,O}}{p_{n,A}^2 + 2p_{n,A}p_{n,O}} + \frac{x_B p_{n,B} p_{n,O}}{p_{n,B}^2 + 2p_{n,B}p_{n,O}} + 2x_O \right). \end{aligned} \tag{28}$$

Summing all the equations above, we find

$$-\lambda = \frac{2(x_A + x_B + x_{AB} + x_O)}{p_A + p_B + p_C}. \tag{29}$$

Substitute $-\lambda$ into Equation (28), we find that the following update satisfies the constraints

$$\begin{aligned} p_{n+1,A} &= \frac{2x_{nA/A} + x_{nA/O} + x_{AB}}{2x} \\ p_{n+1,B} &= \frac{2x_{nB/B} + x_{nB/O} + x_{AB}}{2x} \\ p_{n+1,O} &= \frac{x_{nA/O} + x_{nB/O} + 2x_O}{2x}. \end{aligned} \tag{30}$$

This problem is similar to gene counting problems. In traditional gene counting problems, there are two alleles and three different genotypes. But in this problem, there are three alleles and four different blood types. Julia code and results for the first 5 iterations (starting with $p_A = p_B = p_C = 1/3$) is attached.

```
# data
xa = 186;
xb = 38;
xab = 13;
xo = 284;


# initialization
```

```
pa = 1/3;
pb = 1/3;
po = 1/3;

# update
for i=1:10
    x = xa+xb+xab+xo;
    xnaa = xa*(pa*pa)/(pa*pa+2*pa*po);
    xnao = xa*(2*pa*po)/(pa*pa+2*pa*po);
    xnbb = xb*(pb*pb)/(pb*pb+2*pb*po);
    xnbo = xb*(2*pb*po)/(pb*pb+2*pb*po);
    pa_tmp = (2*xnaa+xnao+xab)/(2*x);
    pb_tmp = (2*xnbb+xnbo+xab)/(2*x);
    po_tmp = (xnao+xnbo+2*xo)/(2*x);
    pa = pa_tmp;
    pb = pb_tmp;
    po = po_tmp;
    println(i, " ", pa, " ", pb, " ", po);
end
```

n	p_A	p_B	p_O
0	0.3333	0.3333	0.3333
1	0.2185	0.0505	0.7311
2	0.2142	0.0502	0.7357
3	0.2137	0.0501	0.7362
4	0.2136	0.0501	0.7363
5	0.2136	0.0501	0.7363

Problem 1.14

To find the optimal $\mathbf{p}$, we maximize $\ln f(\mathbf{f})$, with the constraint $\sum_{i=1}^{m} p_i = 1$ and $p_i \geqslant 0$. If we initialize $\mathbf{p}$ with non-negative values and each update does not make any p_i negative, then we can consider only the equality constraint. Thus, we arrive at the Lagrangian

$$\mathcal{L}(\mathbf{p}, \lambda) = \ln \left(\sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij} p_i p_j \right) + \lambda \sum_{i=1}^{m} p_i - \lambda. \tag{31}$$

By (1.3) we find a surrogate function $g(\mathbf{p}|\mathbf{p}_n)$ for $\mathcal{L}(\mathbf{p}, \lambda)$

$$\begin{aligned}
&\ln\left(\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}p_ip_j\right) + \lambda\sum_{i=1}^{m} p_i - \lambda \\
&\geqslant \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{w_{ij}p_{n,i}p_{n,j}}{f(\mathbf{p}_n)} \ln\left(\frac{f(\mathbf{p}_n)}{w_{ij}p_{n,i}p_{n,j}} w_i p_i p_j\right) + \lambda\sum_{i=1}^{m} p_i - \lambda \qquad (32) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{w_{ij}p_{n,i}p_{n,j}}{f(\mathbf{p}_n)} \ln p_i + \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{w_{ij}p_{n,i}p_{n,j}}{f(\mathbf{p}_n)} \ln p_j + \lambda\sum_{i=1}^{m} p_i - \lambda + c_n,
\end{aligned}$$

where c_n is a variable dependent only on w_{ij} and $\mathbf{p}_n$. Take the partial derivative of g with respect to p_k, we get

$$\begin{aligned}
\frac{\partial g}{\partial p_k} &= \frac{1}{p_k}\sum_{j=1}^{m} \frac{w_{kj}p_{n,k}p_{n,j}}{f(\mathbf{p}_n)} + \frac{1}{p_k}\sum_{i=1}^{m} \frac{w_{ik}p_{n,i}p_{n,k}}{f(\mathbf{p}_n)} + \lambda \\
&= \frac{2}{p_k}\sum_{j=1}^{m} \frac{w_{kj}p_{nk}p_{nj}}{f(\mathbf{p}_n)} + \lambda,
\end{aligned} \qquad (33)$$

where the last equality follows because $w_{ij} = w_{ji}$. Set the partial derivative $\frac{\partial g}{\partial p_k}$ to 0, we get

$$\begin{aligned}
0 &= \frac{2}{p_k}\sum_{j=1}^{m} \frac{w_{kj}p_{nk}p_{nj}}{f(\mathbf{p}_n)} + \lambda \\
\lambda p_k &= -2\sum_{j=1}^{m} \frac{w_{kj}p_{nk}p_{nj}}{f(\mathbf{p}_n)} \\
\sum_{j=1}^{m} \frac{w_{kj}p_{nk}p_{nj}}{f(\mathbf{p}_n)} &= \frac{-\lambda p_k}{2}.
\end{aligned} \qquad (34)$$

Sum over k on both sides, we get

$$\begin{aligned}
\lambda\sum_{k=1}^{m} p_k &= -2 \\
\lambda &= \frac{-2}{\sum_{k=1}^{m} p_k}.
\end{aligned} \qquad (35)$$

Substitute λ into Equation (34), and change indexing from k to i, we get

$$\frac{\sum_{j=1}^{m} w_{ij}p_{n,i}p_{n,j}}{\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}p_{n,i}p_{n,j}} = \frac{p_i}{\sum_{i=1}^{m} p_i}. \qquad (36)$$

The update

$$p_{n+1,i} = \frac{\sum_{j=1}^{m} w_{ij}p_{n,i}p_{n,j}}{\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}p_{n,i}p_{n,j}} = \frac{p_{n,i}\sum_{j=1}^{m} w_{ij}p_{n,j}}{f(\mathbf{p}_n)} \qquad (37)$$

clearly satisfies the constraint $\sum_{i=1}^{m} p_i = 1$, $p_i \geqslant 0$, and makes $\lambda = -2$. The update of the MM algorithm also guarantees to increase fitness without imposing extra assumptions.

Problem 1.15

The likelihood of the data is

$$L(\theta) = \prod_{i=1}^{m} \frac{c_{x_i}\theta^{x_i}}{q(\theta)} = \prod_{i=1}^{m} \frac{c_{x_i}\theta^{x_i}}{\sum_{k=0}^{\infty} c_k\theta^k} \tag{38}$$

Taking the logarithm, we get

$$\begin{aligned} \ln L(\theta) &= \sum_{i=1}^{m} \left[\ln\left(c_{x_i}\theta^{x_i}\right) - \ln\left(\sum_{k=0}^{\infty} c_k\theta^k\right)\right] \\ &= -m\ln q(\theta) + \sum_{i=1}^{m} x_i \ln\theta + c, \end{aligned} \tag{39}$$

where c is a constant dependent only on the coefficients c_k. Because of the log-concavity of $q(\theta)$, we have the inequality

$$\ln q(\theta) \leqslant \ln q(\theta_n) + \frac{q'(\theta_n)}{q(\theta_n)}(\theta - \theta_n) \tag{40}$$

which implies the minorization

$$g(\theta|\theta_n) = -m\ln q(\theta_n) - m\frac{q'(\theta_n)}{q(\theta_n)}(\theta - \theta_n) + \sum_{i=1}^{m} x_i \ln\theta + c. \tag{41}$$

The derivative of g with respect to θ is

$$\frac{dg}{d\theta} = -m\frac{q'(\theta_n)}{q(\theta_n)} + \frac{1}{\theta}\sum_{i=1}^{m} x_i. \tag{42}$$

Set the derivative to 0, and solve for θ gives the update

$$\theta_{n+1} = \frac{\bar{x}q(\theta_n)}{q'(\theta_n)} \tag{43}$$

Because the update keeps the parameter positive, one can start with any positive θ_0 to keep θ_{n+1} positive.

Problem 1.17

First, we find the density of X,

$$\begin{aligned}
\Pr(X = x|\lambda) &= \frac{\lambda^x}{x!}e^{-\lambda} \\
\Pr(\Lambda = \lambda|\alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} \\
\Pr(X = x) &= \int \Pr(X = x|\lambda)\Pr(\Lambda = \lambda|\alpha, \beta)d\lambda \\
&= \int e^{-\lambda}\frac{\lambda^x}{x!}\frac{\beta^\alpha\lambda^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta\lambda}d\lambda \\
&= \frac{\beta^\alpha}{x!\Gamma(\alpha)}\int e^{-\lambda(1+\beta)}\lambda^{[(x+\alpha)-1]}d\lambda
\end{aligned} \tag{44}$$

Let $u = \lambda(1 + \beta)$, then $\lambda = u/(1 + \beta)$, and $d\lambda = du/(1 + \beta)$. Then the integral

$$\begin{aligned}
\int e^{-\lambda(1+\beta)}\lambda^{[(x+\alpha)-1]}d\lambda &= \frac{1}{(1+\beta)^{x+\alpha}}\int e^{-u}u^{[(x+\alpha)-1]}d\lambda \\
&= \frac{\Gamma(x+\alpha)}{(1+\beta)^{x+\alpha}},
\end{aligned} \tag{45}$$

And we conclude with

$$\Pr(X = x) = \frac{\beta^\alpha\Gamma(x+\alpha)}{\Gamma(\alpha)(1+\beta)^{x+\alpha}x!}. \tag{46}$$

To find mean of X,

$$\mathrm{E}[X] = \frac{\beta^\alpha}{\Gamma(\alpha)}\sum_{x=0}^{\infty}\frac{\Gamma(x+\alpha)}{(1+\beta)^{x+\alpha}(x-1)!} \tag{47}$$

The rest of this problem is unfinished.

Problem 1.18

First, notice that

$$\begin{aligned}
&3[(x - x_n)^2 + (y - y_n)^2 + (z - z_n)^2] \\
&= \|(1, 1, -1)\|^2\|(x - x_n, y - y_n, z - z_n)\|^2.
\end{aligned} \tag{48}$$

By Cauchy-Schwarz inequality

$$\begin{aligned}
&\|(1, 1, -1)\|^2\|(x - x_n, y - y_n, z - z_n)\|^2 \\
&\geqslant \|(1, 1, -1)^*(x - x_n, y - y_n, z - z_n)\|^2 \\
&= [(x - x_n) + (y - y_n) - (z - z_n)]^2.
\end{aligned} \tag{49}$$

Therefore,

$$\begin{aligned}
&3[(x-x_n)^2+(y-y_n)^2+(z-z_n)^2]-(x_n+y_n-z_n)^2+2(x_n+y_n-z_n)(x+y-z)\\
&\geqslant [(x-x_n)+(y-y_n)-(z-z_n)]^2-(x_n+y_n-z_n)^2+2(x_n+y_n-z_n)(x+y-z)\\
&=(x-2x_n+y-2y_n-z+2z_n)(x+y-z)+2(x_n+y_n-z_n)(x+y-z)\\
&=[(x+y-z)-2(x_n+y_n-z_n)](x+y-z)]+2(x_n+y_n-z_n)(x+y-z)\\
&=(x+y-z)^2-2(x_n+y_n-z_n)(x+y-z)+2(x_n+y_n-z_n)(x+y-z)\\
&=(x+y-z)^2.
\end{aligned} \tag{50}$$

To show tangency property, plug in $x=x_n,\ y=y_n,\ z=z_n$ on both sides. On both sides we get $(x_n+y_n-z_n)^2$. Therefore, the right hand side of the inequality is a majorization of the left hand side.