Yen-Chun Chen
chenyenc@oregonstate.edu

Q1.

(a) $f(x) = \frac{2x}{\lambda^2} \exp\left(-\left(\frac{x}{\lambda}\right)^2\right)$

Since the samples independent to each others

$$f(x_1, x_2, \ldots, x_n | \lambda) = \prod_{i=1}^{n} \frac{2x_i}{\lambda^2} \exp\left[-\left(\frac{x_i}{\lambda}\right)^2\right]$$

Take natural log of $f(x_1, x_2, \ldots, x_n)$ for simpler calculation.

$$L(\lambda) = \log f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \log\left(\frac{2x_i}{\lambda^2} \exp\left(-\frac{x_i^2}{\lambda^2}\right)\right)$$

$$= \sum \log \frac{2x_i}{\lambda^2} + \log \exp\left(-\frac{x_i^2}{\lambda^2}\right) = \sum\left(\log 2x_i + \log \lambda^{-2} - \frac{x_i^2}{\lambda^2}\right)$$

$$= \sum\left(\log 2 + \log x_i - 2\log\lambda - \frac{x_i^2}{\lambda^2}\right) = n \cdot \log 2 + \sum_{i=1}^{n} \log x_i - 2n\log\lambda - \sum_{i=1}^{n}\left(\frac{x_i^2}{\lambda^2}\right)$$

Take the derivative with respect to λ to find the maximum

$$0 = \frac{\partial L(\lambda)}{\partial \lambda} \Rightarrow 0 + 0 - 2n\frac{1}{\lambda} - (-2)\sum_{i=1}^{n} \frac{x_i^2}{\lambda^3} = 0$$

$$\Rightarrow \frac{-\cancel{2}n}{\cancel{\lambda}} + \cancel{2}\frac{1}{\lambda^{\cancel{3}}}\sum_{i=1}^{n} x_i^2 = 0 \Rightarrow \cancel{n}\lambda^2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

$$\Rightarrow \hat{\lambda} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

#

(b) $n = 3$, $x_1 = 1.35$, $x_2 = 2.76$, $x_3 = 3.12$

$$\hat{\lambda} = \sqrt{\frac{1}{3}\left[(1.35)^2 + (2.76)^2 + (3.12)^2\right]}$$

$$= 2.528 \quad \#$$

Q2.

(a)

$$f(x_1, \dots x_n \mid \alpha, \beta) = \prod_{i=1}^{n} \frac{\exp\left(-\frac{x_i - \alpha}{\beta}\right)}{\beta\left[1 + \exp\left(-\frac{x_i - \alpha}{\beta}\right)\right]^2}$$

(b)

$$L(\alpha, \beta) = -\log\left(\prod_{i=1}^{n} \frac{\exp\left(-\frac{x_i - \alpha}{\beta}\right)}{\beta\left[1 + \exp\left(-\frac{x_i - \alpha}{\beta}\right)\right]^2}\right)$$

$$= -\left\{\sum_{i=1}^{n} \log \exp\left(-\frac{x_i - \alpha}{\beta}\right) - \log \beta\left[1 + \exp\left(-\frac{x_i - \alpha}{\beta}\right)\right]^2\right\}$$

$$= \sum_{i=1}^{n}\left\{\frac{x_i - \alpha}{\beta} + \log\beta + 2 \cdot \log\left(1 + \exp\left(-\frac{x_i - \alpha}{\beta}\right)\right)\right\}$$

$$= \sum_{i=1}^{n} \frac{x_i - \alpha}{\beta} + n \cdot \log\beta + 2\sum_{i=1}^{n} \log\left(1 + \exp\left(-\frac{x_i - \alpha}{\beta}\right)\right)$$

$$\log'(\alpha) = \frac{1}{\alpha}$$

$$\frac{\partial L}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[\sum_{i=1}^{n}\left(\frac{x_i}{\beta} - \frac{\alpha}{\beta}\right) + n \log \beta + 2 \cdot \sum_{i=1}^{n} \log\left(1 + \exp\left(\underbrace{-\frac{x_i - \alpha}{\beta}}_{\frac{\partial}{\partial \alpha} = \frac{1}{\beta}}\right)\right)\right]$$

$$= \sum_{i=1}^{n}\left(-\frac{1}{\beta}\right) + 0 + 2\sum_{i=1}^{n}\left(\frac{1}{1+\exp\left(-\frac{x_i-\alpha}{\beta}\right)} \cdot \exp\left(-\frac{x_i-\alpha}{\beta}\right) \cdot \frac{1}{\beta}\right)$$

$$= \frac{-n}{\beta} + \frac{2n}{\beta}\sum_{i=1}^{n} \frac{\exp\left(-\frac{x_i-\alpha}{\beta}\right)}{1+\exp\left(-\frac{x_i-\alpha}{\beta}\right)} \quad \#$$

$$\frac{\partial L}{\partial \beta} = \frac{\partial}{\partial \beta}\left[\sum_{i=1}^{n}\frac{x_i - \alpha}{\beta} + n \cdot \log \beta + 2 \sum_{i=1}^{n} \log\left(1 + \exp\left(\underbrace{-\frac{x_i-\alpha}{\beta}}_{\frac{\partial}{\partial \beta} = \cancel{(-1)}\cancel{(-1)}(x_i-\alpha)\beta^{-2}}\right)\right)\right]$$

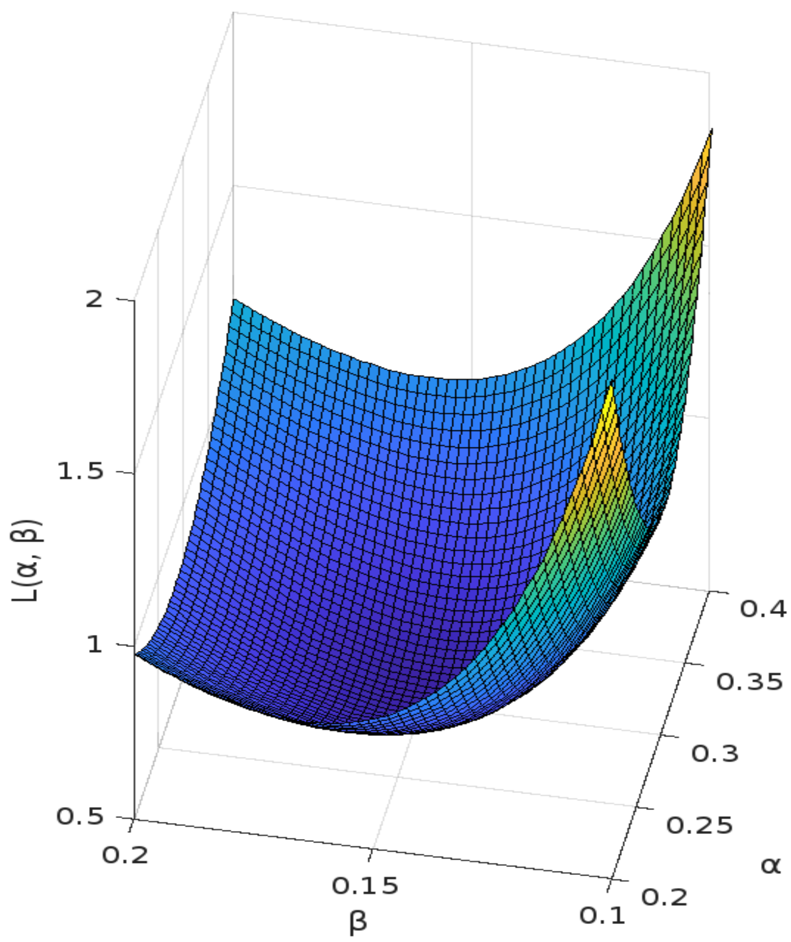
$$= \left[\sum_{i=1}^{n}(-1)\cdot\frac{x_i-\alpha}{\beta^2}\right] + \frac{n}{\beta} + 2\sum_{i=1}^{n}\left(\frac{1}{1+\exp\left(-\frac{x_i-\alpha}{\beta}\right)} \exp\left(-\frac{x_i-\alpha}{\beta}\right) \cdot \frac{x_i-\alpha}{\beta^2}\right)$$

$$= \frac{-1}{\beta^2}\sum_{i=1}^{n}(x_i-\alpha) + \frac{n}{\beta} + \frac{2n}{\beta^2}\sum_{i=1}^{n}\frac{\exp\left(-\frac{x_i-\alpha}{\beta}\right)(x_i-\alpha)}{1+\exp\left(-\frac{x_i-\alpha}{\beta}\right)} \quad \#$$

$$\nabla C = \begin{bmatrix} \dfrac{-n}{\beta} + \dfrac{2n}{\beta}\displaystyle\sum_{i=1}^{n} \frac{\exp\left(-\frac{x_i-\alpha}{\beta}\right)}{1+\exp\left(-\frac{x_i-\alpha}{\beta}\right)} \\[2em] \dfrac{-1}{\beta^2}\displaystyle\sum_{i=1}^{n}(x_i-\alpha) + \dfrac{n}{\beta} + \dfrac{2n}{\beta^2}\displaystyle\sum_{i=1}^{n}\frac{\exp\left(-\frac{x_i-\alpha}{\beta}\right)(x_i-\alpha)}{1+\exp\left(-\frac{x_i-\alpha}{\beta}\right)} \end{bmatrix}$$

(c) I used MATLAB with sample code provided on Canvas to do the homework.

```
close all;
clear all;
x = [.30753, .56678, -.25177, .37243, .26375];
n = 5;
% create a grid of points for α, β
a = linspace(.2,.4,50);
b = linspace(.1,.2,50);
[alpha_grid, beta_grid] = meshgrid(a, b);
L = zeros(1, n);
disp(size(alpha_grid, 1));
disp(size(alpha_grid, 2));
% along with row number
for i = 1:size(alpha_grid, 1)
   % along with column number
   for j = 1:size(alpha_grid, 2)
       alpha = alpha_grid(i, j);
       beta = beta_grid(i, j);
       % calculate f(x|α,β) for each x value
       f_x = exp(-(x - alpha) ./ beta) ./ (beta * (1 + exp(-(x - alpha) ./ beta)).^2);
       % calculate L(α, β)
       L(i, j) = -log(prod(f_x));
   end
end
subplot(121);
surf(alpha_grid, beta_grid, L);
xlabel('α');
ylabel('β');
zlabel('L(α, β)');
```



(d)

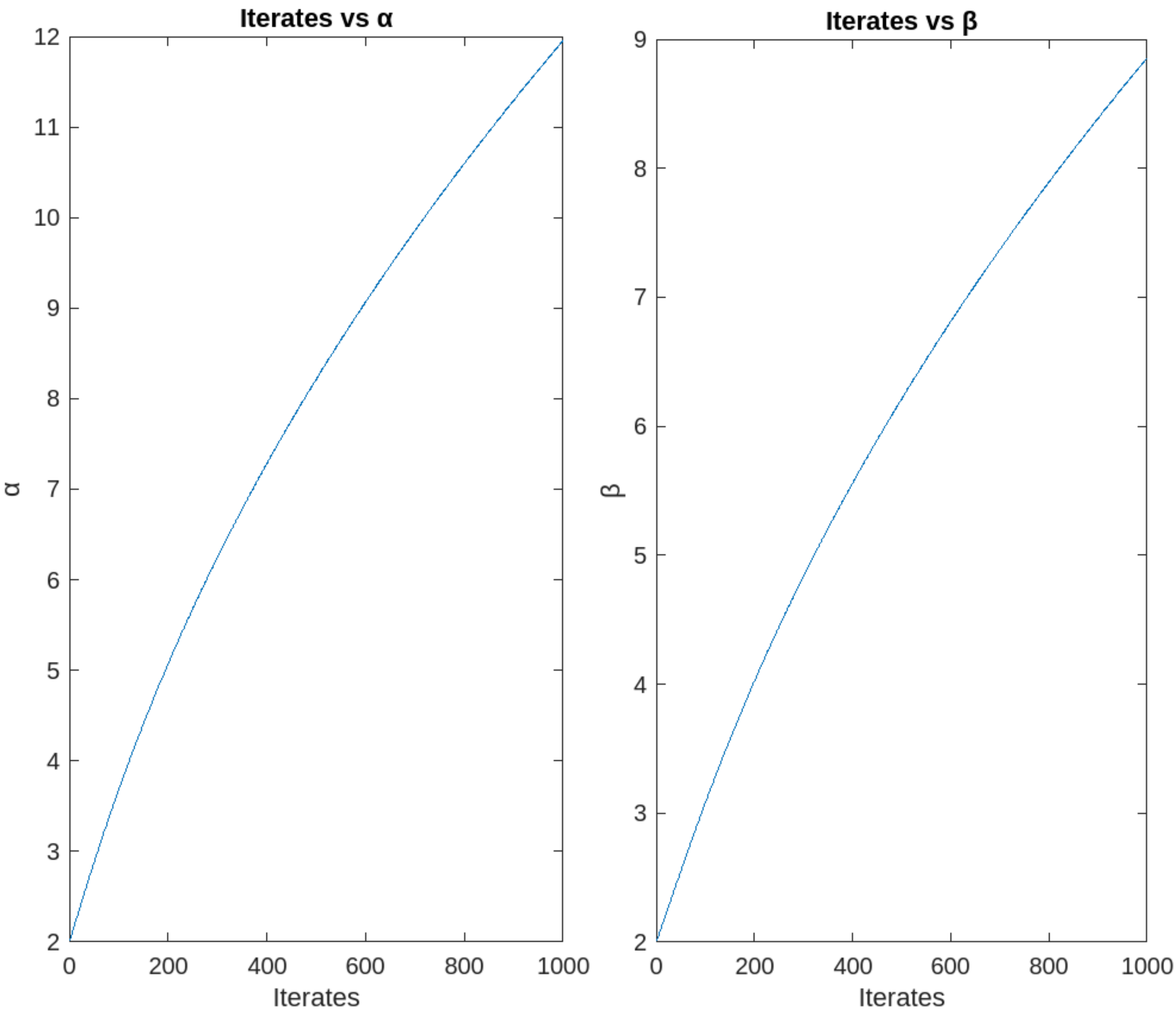
```
n = 5;
x = [0.30753, 0.56678, -0.25177, 0.37243, 0.26375];
eta = 0.001;
iterates = 1000;
p = [2.0; 2.0];

% Gradient
gradient = @(alpha, beta) [
   -n/beta + 2*n/beta^2 * sum((x - alpha) .* exp(-(x - alpha) / beta) ./ (1 + exp(-(x -
alpha) / beta)));
   -1/beta^2 * sum(x - alpha) + n/beta + 2*n/beta^2 * sum(exp(-(x - alpha) / beta) .* (x
- alpha) ./ (1 + exp(-(x - alpha) / beta)))
];

% 1000 loops
alpha_values = zeros(iterates, 1);
beta_values = zeros(iterates, 1);
for i = 1:iterates
   grad = gradient(p(1), p(2));
   p = p - eta * grad;
   alpha_values(i) = p(1);
   beta_values(i) = p(2);
end

% final estimates of alpha and beta
fprintf('estimate of α: %.3f\n', p(1));
fprintf('estimate of β: %.3f\n', p(2));

% Plot
figure;
subplot(1, 2, 1);
plot(1:iterates, alpha_values);
xlabel('Iterates');
ylabel('α');
title('Iterates vs α');
subplot(1, 2, 2);
plot(1:iterates, beta_values);
xlabel('Iterates');
ylabel('β');
title('Iterates vs β');
```



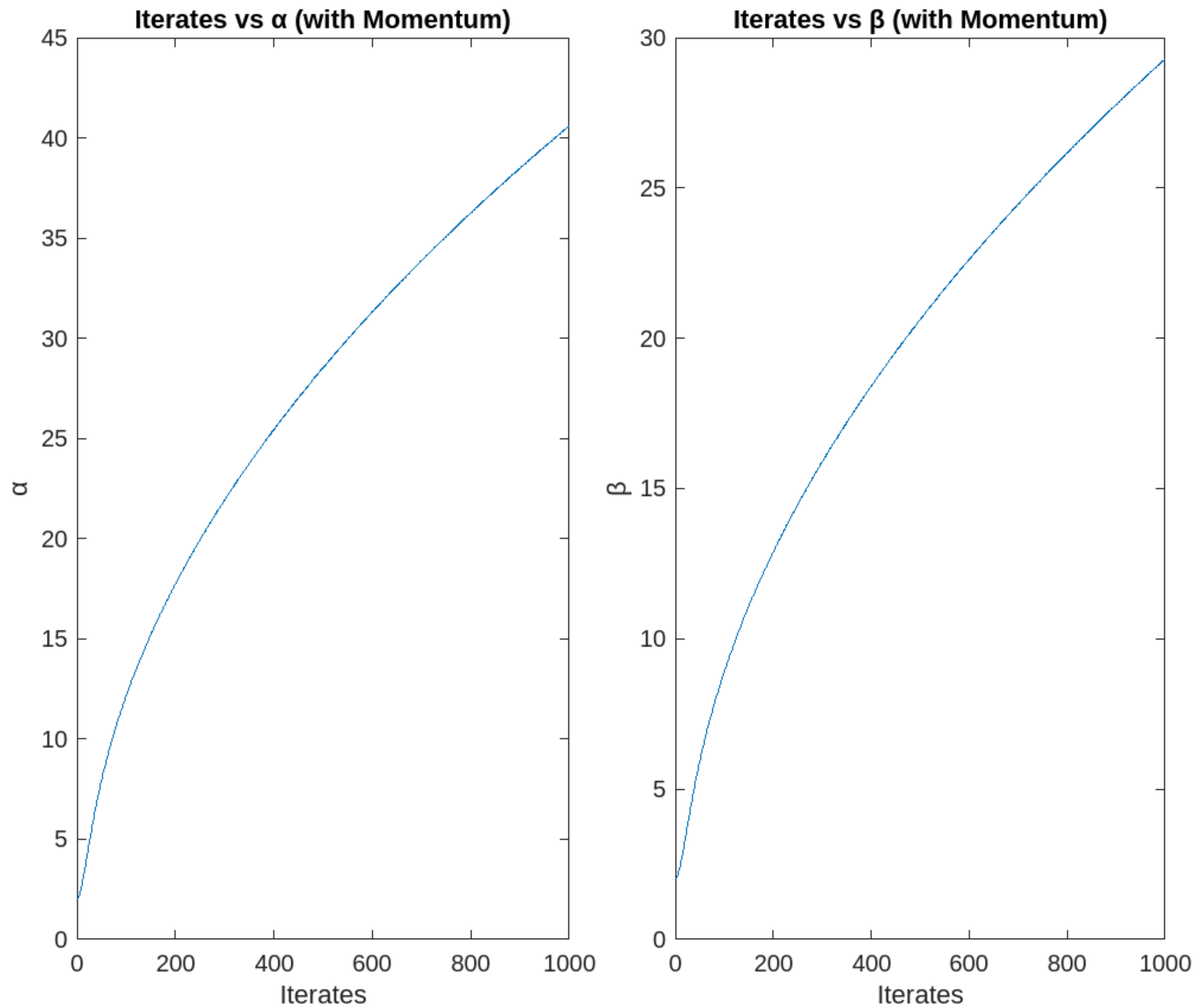

```
The estimate of α: 11.958
The estimate of β: 8.857
```

(e)

```
n = 5;
x = [.30753, .56678, -.25177, .37243, .26375];
eta = .001;
gamma = .9;
iterates = 1000;
p = [2.0; 2.0];
% Gradient
gradient = @(alpha, beta) [
   -n/beta + 2*n/beta^2 * sum((x - alpha) .* exp(-(x - alpha) / beta) ./ (1 +
exp(-(x - alpha) / beta)));
   -1/beta^2 * sum(x - alpha) + n/beta + 2*n/beta^2 * sum(exp(-(x - alpha) / beta)
.* (x - alpha) ./ (1 + exp(-(x - alpha) / beta)))
];
theta = [0;0];
% 1000 loops
alpha_values = zeros(iterates, 1);
beta_values = zeros(iterates, 1);
for i = 1:iterates
   grad = gradient(p(1), p(2));
   theta = gamma * theta + eta * grad;
   p = p - theta;
   alpha_values(i) = p(1);
   beta_values(i) = p(2);
end

% final estimates of alpha and beta
fprintf('estimate of α: %.3f\n', p(1));
fprintf('estimate of β: %.3f\n', p(2));

% Plot iterative values of alpha and beta
figure;
subplot(1, 2, 1);
plot(1:iterates, alpha_values);
xlabel('Iterates');
ylabel('α');
title('Iterates vs α (with Momentum)');
subplot(1, 2, 2);
plot(1:iterates, beta_values);
xlabel('Iterates');
ylabel('β');
title('Iterates vs β (with Momentum)');
```




```
The estimate of α with momentum: 40.606
The estimate of β with momentum: 29.296
```

From my plots, there seems no explicit difference. However, the momentum should be faster.