Problem 1 (graded by Yiran) - 50 points

(a) 4 points

In a class, among 20 students, 8 are female, and 12 are male. 2 of the female students are taller than 170 cm, and 8 of the male students are taller than 170 cm. Suppose we randomly pick a student, let

x: the student is female;

y: the student is taller than 170 cm.

Then,

$P(x, y)$ is the probability that the student is both female and is taller than 170 cm, which is equal to $2/20 = 0.1$.

$P(x)$ is the probability that the student is female, which is equal to $8/20 = 0.4$.

$P(y|x)$ is the probability that known the student is female, she is taller than 170 cm, which is equal to $2/8 = 0.25$.

$P(y)$ is the probability that the student is taller than 170 cm, which is equal to $(2+8)/20 = 0.5$.

$P(x|y)$ is the probability that known the student is taller than 170 cm, the student is female, which is equal to $2/(2 + 8) = 0.2$.

We see that:

$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$

(b) 8 points

Independent

Let

x : I get an A for Ge/ESE118.

y: The next president of the U.S. is an Republician.

This two events are independent, because x happens, does not affect the probability of y, vice versa.

Let's assume $P(x) = 3/5$, and $P(y) = 1/2$.

Suppose I get an A with $P(x)$. It doesn't affect the election at all, and there is still 1 in 2 odds that the next president will be an Republician. Therefore, to make both happen, $P(x, y) = P(x)P(y)$. Similary, suppose the Republician wins the election with $P(y)$. It doesn't affect my odd to get an A, and to make both happen, $P(x, y) = P(y)P(x)$.

Intuitively, the rule holds because the two events are independent - one happens does not affect the other; therefore, to make both happen, we need to multiply $P(x)$ and $P(y)$.

Dependent

Let

x: The next president of the U.S. is an Democratic.

y: The next president of the U.S. is an Republician.

These two events are not independent, because either of them happens, will affect the proability of the other.

Let's assume $P(x) = P(y) = \frac{1}{2}$. Because it's impossible that the next president is both an Democratic and an Republician, $P(x, y) = 0 \neq P(x)P(y)$.

(c) 8 points

(c.i) 3 points

$$E(x) = \int_{-\infty}^{\infty} xP(x)dx$$

Since x is an odd function, and $P(x)$ is an even function, their product is an odd function. Integration of an odd function over symmetric boundaries as $[-\infty, \infty]$ is 0. Therefore,

$$E(x) = 0$$

(c.ii) 5 points

$$\begin{aligned} E(x^2) &= \int_{-\infty}^{\infty} x^2 P(x)dx \\ &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \end{aligned}$$

Since

$$\left[\exp\left(-\frac{x^2}{2\sigma^2}\right)\right]' = -\frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Then

$$\begin{aligned} E(x^2) &= -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left[\exp\left(-\frac{x^2}{2\sigma^2}\right)\right]' dx \\ &= -\frac{\sigma}{\sqrt{2\pi}} \left[x \exp\left(-\frac{x^2}{2\sigma^2}\right)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \right] \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \end{aligned}$$

Since

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-x^2/2} dx &= \sqrt{2\pi} \\ &= \int_{-\infty}^{\infty} \exp\left[-\frac{\left(\frac{x}{\sigma}\right)^2}{2}\right] d\left(\frac{x}{\sigma}\right) \\ &= \frac{1}{\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \end{aligned}$$

Then

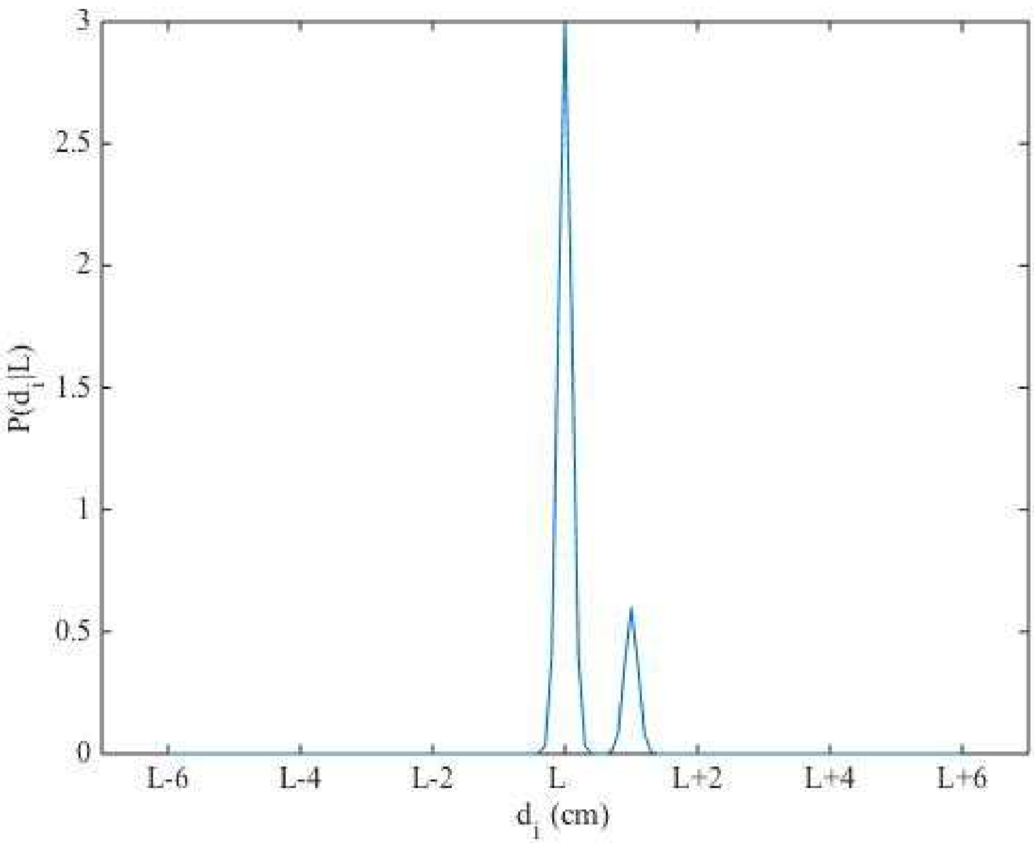
$$E(x^2) = \frac{\sigma}{\sqrt{2\pi}} \sqrt{2\pi}\sigma = \sigma^2$$

(d) 30 points

(d.i) 10 points

$$
\begin{aligned}
P(d_i|L) &= \frac{3}{4}\mathcal{N}(L,\sigma) + \frac{1}{4}\mathcal{N}(L+1,\sigma) \\
&= \frac{3}{4}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(d_i-L)^2}{2\sigma^2}\right) + \frac{1}{4}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(d_i-(L+1))^2}{2\sigma^2}\right) \\
&= \frac{15}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-50(d_i-L)^2\right) + \frac{3}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-50(d_i-(L+1))^2\right)
\end{aligned}
$$

where $\sigma = 0.1$ cm, and L, d_i are in cm.



(d.ii) 10 points

From Bayes' theorem,

$$
\begin{aligned}
P(L|d_i) &= \frac{P(d_i|L)P(L)}{P(d_i)} \\
&\propto P(d_i|L) = \frac{3}{4}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(d_i-L)^2}{2\sigma^2}\right) + \frac{1}{4}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(d_i-(L+1))^2}{2\sigma^2}\right)
\end{aligned}
$$

Where we assume the prior distribution $P(L)$ is uniform, and $P(d_i)$ is a constant (but different for different d_i).

The constant ahead of $P(d_i|L)$ (denoted as "c") is determined by

$$\int_{-\infty}^{\infty} P(L|d_i)dL \quad = \int_{-\infty}^{\infty} cP(d_i|L)dL \quad = 1$$

We can re-write $P(d_i|L)$ as

$$P(d_i|L) \quad = \quad \frac{3}{4}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(L-d_i)^2}{2\sigma^2}\right) + \frac{1}{4}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(L-(d_i-1))^2}{2\sigma^2}\right)$$

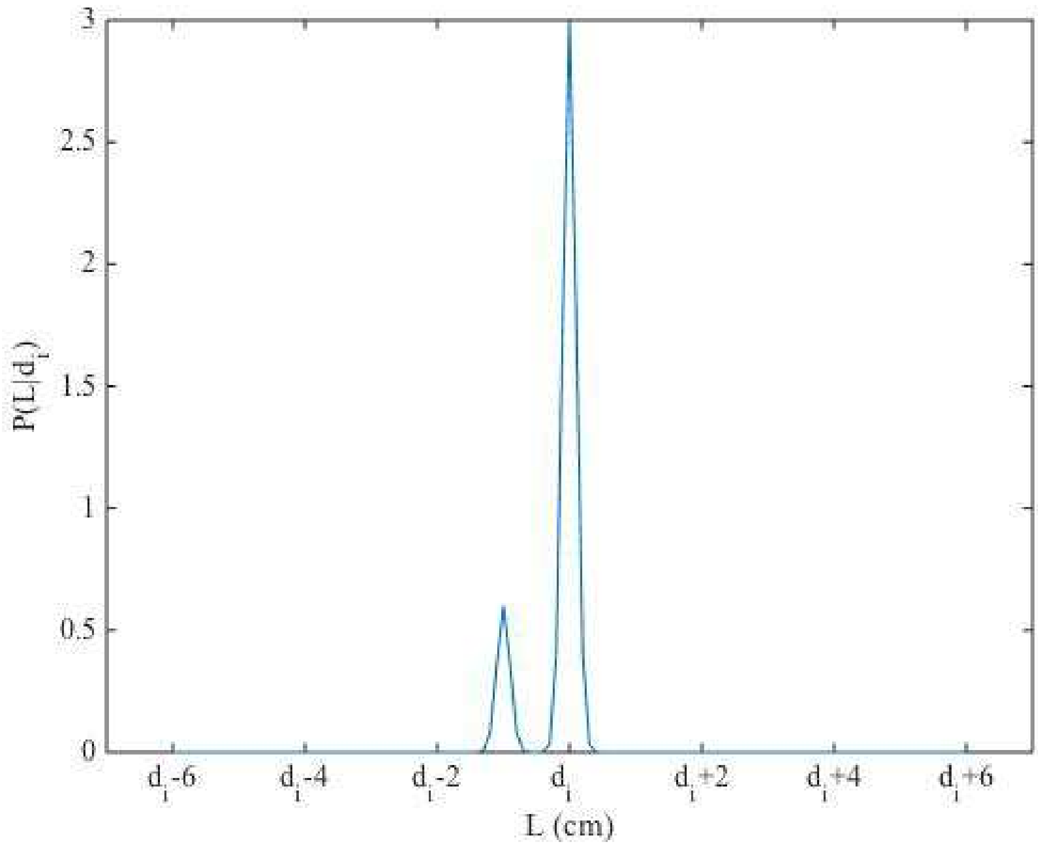
which is a summation of two normal distributions (centered at d_i, and $d_i - 1$), weighted by $\frac{3}{4}$ and $\frac{1}{4}$.

Then

$$\int_{-\infty}^{\infty} \quad P(d_i|L)dL \quad = 1$$

Therefore, the constant $c = 1$

$$\begin{aligned}
P(L|d_i) &= P(d_i|L) \\
&= \frac{3}{4}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(L-d_i)^2}{2\sigma^2}\right) + \frac{1}{4}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(L-(d_i-1))^2}{2\sigma^2}\right) \\
&= \frac{15}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-50(L-d_i)^2\right) + \frac{3}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-50(L-(d_i-1))^2\right)
\end{aligned}$$



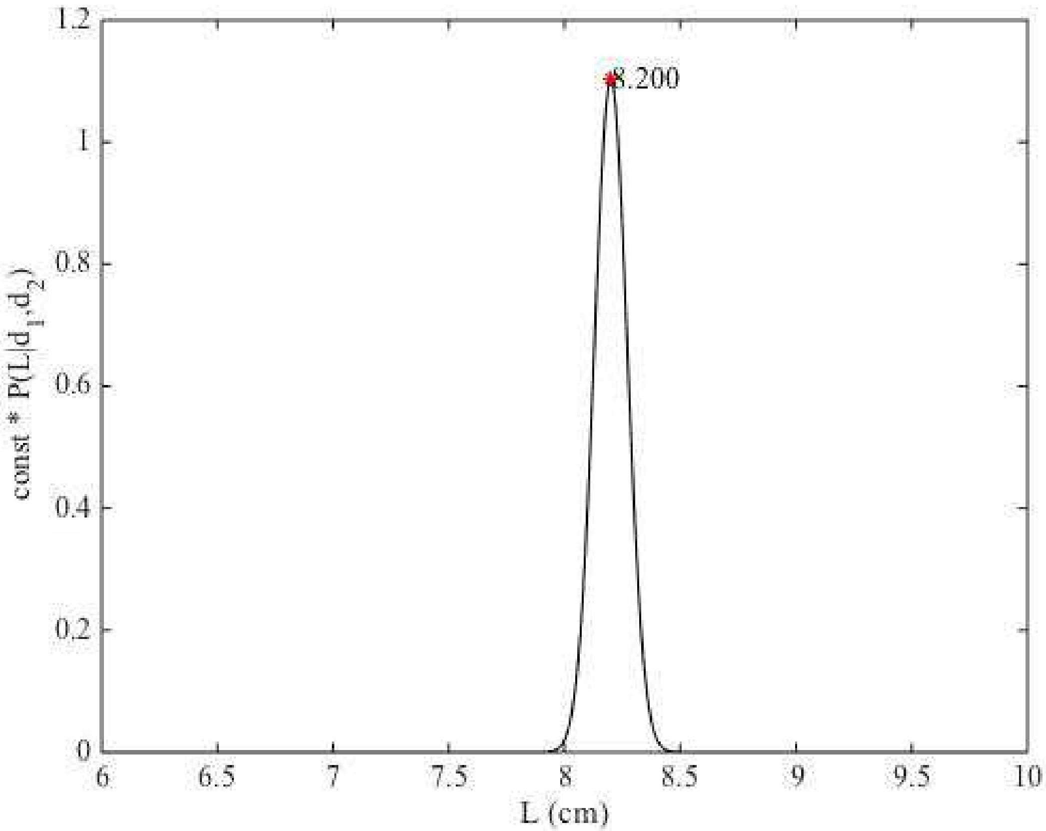
(d.iii) 10 points

We can use the distribution of L from the first measurement as prior distribution to compute a posterior distribution after the second measurement

$$P(L|d_1 = 8.3, d_2 = 9.1) \quad \propto \quad P(d_2 = 9.1|L)P(L|d_1 = 8.3)$$

Since we only care about the maximum of the LHS, instead of its value; for the RHS, absortbing the $\frac{1}{4}\frac{1}{\sigma\sqrt{2\pi}}$ terms into the constant

$$\begin{aligned} P(L|d_1 = 8.3, d_2 = 9.1) \quad \propto \quad & \left[3\exp\left(-50(9.1 - L)^2\right) + \exp\left(-50(8.1 - L)^2\right)\right] \\ & \cdot \left[3\exp\left(-50(L - 8.3)^2\right) + \exp\left(-50(L - 7.3))^2\right)\right] \end{aligned}$$



From the plot, we see that the best estimate of L is 8.2.

Because the two measurement differs ≈ 1, it's likely that in the second measurement, the one quarter chance of additional 1 cm happens. After the 1 cm correction, the second measurement should be 8.1. The mean of 8.3 and 8.1 is 8.2, which is our estimation through the analysis above.

problem 2 (graded by Kangchen) - 50 points

(a)12 points

$P(\boldsymbol{m}) \propto e^{-F(\boldsymbol{m})}$where $F(\boldsymbol{m}) = \sum \frac{(d_k - g_k(\boldsymbol{m}))^2}{2\sigma_k^2}$

(b)12 points

Since $d'_k = d_k/\sigma_k$, the relation between $\boldsymbol{d'}$ and $\boldsymbol{d}$ can be written in matrix form $\boldsymbol{d'} = \boldsymbol{W}\boldsymbol{d}$ where $W_{ij} = \delta_{ij}\frac{1}{\sigma_j}$ ($\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$).

(c)12 points

$$F = \frac{1}{2}(\boldsymbol{d'} - \boldsymbol{g'}(\boldsymbol{m}))^T(\boldsymbol{d'} - \boldsymbol{g}'(\boldsymbol{m}))$$

the gradient: $\boldsymbol{\nabla} F = \hat{\boldsymbol{G}'}^T(\boldsymbol{d'} - \boldsymbol{g'}(\boldsymbol{m}))$

the approximated hessian: $\boldsymbol{H} = (\hat{\boldsymbol{G}'}^T\hat{\boldsymbol{G}'})$

So the least squares solution: $\Delta\boldsymbol{m} = (\hat{\boldsymbol{G}'}^T\hat{\boldsymbol{G}'})^{-1}\hat{\boldsymbol{G}'}^T(\boldsymbol{d'} - \boldsymbol{g'}(\boldsymbol{m}))$

Substitute $\hat{\boldsymbol{G}'} = \boldsymbol{W}\hat{\boldsymbol{G}}$, $\boldsymbol{d'} = \boldsymbol{W}\boldsymbol{d}$, $\boldsymbol{g'} = \boldsymbol{W}\boldsymbol{g}$, $\boldsymbol{C} = \boldsymbol{W}^T\boldsymbol{W}$

$\Delta\boldsymbol{m} = (\hat{\boldsymbol{G}}^T\boldsymbol{W}^T\boldsymbol{W}\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}^T\boldsymbol{W}^T\boldsymbol{W}(\boldsymbol{d} - \boldsymbol{g}(\boldsymbol{m})) = (\hat{\boldsymbol{G}}^T\boldsymbol{C}\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}^T\boldsymbol{C}(\boldsymbol{d} - \boldsymbol{g}(\boldsymbol{m}))$

(d)14 points

```
function [ Grad, Hess] = compute_gradient_approx_hess( x,y,M,residue,weigh)

W =diag(weigh);

C=W'*W;

xs = M(1);
ys = M(2);
zs = M(3);
p = M(4);

R = ((x - xs).^2 + (y - ys).^2 + zs^2);

dx = x-xs;
dy = y-ys;

Ghat(:,1) = (3.*p.*zs.*(dx))./((R).^(5/2));
Ghat(:,2) = (3.*p.*zs.*(dy))./((R).^(5/2));
Ghat(:,3) = p./(R).^(3/2) - (3*p.*zs.^2)./(R).^(5/2);
Ghat(:,4) = zs./(R).^(3/2);

Grad = (residue')*C*Ghat;



Hess = (Ghat')*C*Ghat;
%this is the apprximated Hessian;


Hess = 0.5*(Hess + Hess');

end
```

```

function [M]=nonlinear_solver(x,y,d,Minit,w)

M=Minit;

for ii = 1:1:1000

r=compute_residue(x,y,M,d);


[Grad,Hess]=compute_gradient_approx_hess(x,y,M,r,w);

```

```
deltaM= (Hess)\Grad';

M=M+deltaM;

 if (norm(r)<1e-7)
     break;
end
end
disp(r)
end
```

```
function [ r ] = compute_residue( x,y,M,d )

xs = M(1);
ys = M(2);
zs = M(3);
p = M(4);
r =d- p*zs./((x - xs).^2 + (y - ys).^2 + zs^2).^(3/2);
end
```

```
%%%problem 1d
x = [0 11 15 6 -7 3]';
y = [0 0 6 13 10 -7]';
d = [0.103 0.162  0.065  0.036 0.025 0.169]';
M0 = [8 -5 10 30]';  %initial guess
Ms = nonlinear_solver(x,y,d,M0,[1,1,0.2,1,2.5,2.5]);
```

$$\boldsymbol{m} = [8.3068, -5.3425, 11.8179, 31.8569]^T$$

$$\boldsymbol{error} = [-4.99 \times 10^{-5}, 1.20 \times 10^{-5} - 2.95 \times 10^{-3}, 3.59 \times 10^{-4}, -2.87 \times 10^{-5}, 2.74 \times 10^{-6}]$$

We can find that since we put a smaller weight on station 3, its error is the largest. Since we put a larger weight on station 5 6 , their errors are smaller.

(Extra Credit) Problem 3 (graded by Yiran) - 25 points

(a) 5 points

Since $\boldsymbol{g_i}$ has M elements, the maximum dimension of $\boldsymbol{G}$ spanned by $\{\boldsymbol{g_i}, i = 1...N\}$ is M. Therefore,

$$dim\,(R(\boldsymbol{G})) \quad \leq \quad M < N = dim(\mathbb{R}^N)$$

(b) 10 points

$$\boldsymbol{H_{ij}} \quad = \quad \boldsymbol{g_i}^T \boldsymbol{g_j}$$

The diagonal elements of $\boldsymbol{H}$ are the squared lengths of the column vectors of $\boldsymbol{G}$, the off-diagnoal elements measure how much the column vectors of $\boldsymbol{G}$ project onto each other. Suppose we project a vector $\boldsymbol{d}$ into the column space of $\boldsymbol{G}$, the coordinates are $(m_1, ..., m_M)$. If the length of certain column vector $\boldsymbol{g_i}$ is small, an error in $\boldsymbol{d}$ will cause a big error in m_i. The error in $\boldsymbol{d}$ also tends to affect those m_i similarly if their corresponding column vectors $\boldsymbol{g_i}$ are near parallel.

(c) 5 points

$$\boldsymbol{G}^T \boldsymbol{d} \quad = \quad (\boldsymbol{g_1}^T \boldsymbol{d}, ..., \boldsymbol{g_M}^T \boldsymbol{d})^T$$

which is the projection of $\boldsymbol{d}$ into the model space.

(d) 5 points

This equation implies that $\boldsymbol{Gm}$ equals the projection of $\boldsymbol{d}$ in the model space. Thus, the least-squares solution $\boldsymbol{m}$ is the coordinate of $\boldsymbol{d}$ in the model space.