Problem 1

$y_i \overset{ind}{\sim} Normal(x_i^T w, \lambda^{-1})$, $w \sim Normal(0, diag(\alpha_1 \cdots \alpha_d)^{-1})$

$\alpha_k \overset{iid}{\sim} Gamma(a_0, b_0)$, $\lambda \sim Gamma(e_0, f_0)$.

$$Gamma(\eta \mid \tau_1, \tau_2) = \frac{\tau_2^{\tau_1}}{\Gamma(\tau_1)} \eta^{\tau_1 - 1} e^{-\tau_2 \eta}$$

$$P(x, y, w, \alpha, \lambda) = P(y \mid x, w, \lambda) \cdot P(w \mid \alpha) \cdot P(\lambda) \cdot P(\alpha)$$

$\Rightarrow$

$$\ln P(x, y, w, \alpha, \lambda) = \ln P(y \mid x, w, \lambda) + \ln p(w \mid \lambda) + \ln P(\lambda) + \ln P(\alpha)$$

our variables are w, α, λ we need to factorize $q(w, \alpha, \lambda)$ into $q(w)\, q(\alpha)\, q(\lambda)$

for the remainder of the problem we are going to assume that $diag(\alpha_1 \cdots \alpha_d)^{-1} = \Phi^{-1}$

$\Rightarrow$ we need $q(w)\, q(\lambda)\, q(\Phi)$

first lets calculate

$$P(y \mid x, w, \lambda) \sim Normal(x_i^T w, \lambda^{-1})$$

$$P(y \mid x, w, \lambda) = \prod_{i=1}^{n} (\lambda)^{n/2} \cdot (2\pi)^{n/2} \cdot \exp\left\{ -\frac{\lambda}{2} \sum_{i=1}^{n} (y_i - x_i^T w)^2 \right\}$$

$$= (\lambda)^{n/2} (2\pi)^{n/2} \cdot \exp\left\{ -\frac{\lambda}{2} \sum_{i=1}^{n} (y_i - x_i^T w)^2 \right\}$$

$$\Rightarrow \ln p(y \mid x, w, \lambda) = \frac{n}{2} \ln(\lambda) + \frac{n}{2} \ln(2\pi) - \frac{\lambda}{2} \sum (y_i - x_i^T w)^2$$

$$= \frac{n}{2} \ln(\lambda) + \frac{n}{2} \ln(2\pi) - \frac{\lambda}{2} \sum (y_i^2 - 2 y_i x_i^T w + x_i^T w\, x_i w^T$$

for w we need

$$\mathbb{E}_{-q(w)} \left[\ln(p(y \mid x, w, \lambda) + \ln P(w \mid \alpha) \right]$$

taking the term inside and plugging the distribution

$$\frac{n}{2} \ln(\lambda) + \frac{n}{2} \ln(2\pi) - \frac{\lambda}{2} \sum (y_i - x_i^T w)^2 + \ln P(N(0, \Phi))$$

calculating $P(\omega_i \mid \varphi^{-1}) = (2\pi)^{-d/2} \cdot \varphi^{d/2} \exp\{-\frac{\varphi}{2} \omega_i^T \omega_i\}$

$\ln P(\omega \mid \varphi^{-1}) = \frac{d}{2}\ln(2\pi) + \frac{d}{2}\ln(\varphi) + \frac{\varphi}{2}\omega^T\omega$

plugging the above and dropping values that don't depend on ω then inside term becomes

$\Rightarrow -\frac{\lambda}{2}\Sigma(y_i - x_i^T\omega)^2 + (-\frac{\varphi}{2}\omega^T\omega)$

$= -\frac{\lambda}{2}\sum_{i=1}^{n}(y_i^2 - 2\omega^T x_i y_i + \omega^T x_i x_i^T \omega) - (\frac{1}{2}\omega^T \varphi \omega)$

Now we replace λ and φ with the expectations

$\Rightarrow -\frac{E\lambda}{2}(y_i^2 - 2\omega^T x_i y_i + \omega^T x_i x_i^T \omega) - \frac{1}{2}(\omega^T E(\varphi)\omega)$

$\Rightarrow -\frac{1}{2}\left(\omega^T(E(\varphi) + \Sigma\lambda\Sigma x_i x_i^T)\omega - 2\omega^T(E\lambda\Sigma x_i y_i)\right)$

$\Rightarrow \Sigma' = \left(E(\varphi) + E\lambda\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$

$\mu' = \Sigma'(E\lambda\sum_{i=1}^{n} y_i x_i)$

$\Rightarrow q(\omega) = N(\mu', \Sigma')$

We now calculate $q(\lambda)$

$E_{-q(\lambda)}\left[\ln P(y \mid X, \omega, \lambda) + \ln P(\lambda)\right]$

$P(\lambda) \sim \text{Gamma}(e_0, f_0)$

$P(\lambda) = \frac{f_0^{e_0}}{\Gamma(e_0)} \lambda^{e_0 - 1} e^{-f_0 \lambda}$

$\ln P(\lambda) = e_0 \ln(f_0) - \ln(\Gamma(e_0)) + \ln(\lambda)^{e_0 - 1} - f_0\lambda$

Going back to original formula we get

$\frac{n}{2}\ln(\lambda) + \frac{n}{2}\ln(2\pi) - \frac{\lambda}{2}\Sigma(y_i^2 - 2y_i \cancel{x_i^T} + x_i^T\omega \cancel{x_i^T} x_i)$
$+ e_0\ln(f_0) - \ln(\Gamma(e_0)) + (e_0 - 1)\ln(\lambda) - f_0\lambda$

Dropping values that don't include λ

$$E_{q(\lambda)}\left[\frac{n}{2}\ln\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}\left(y_i^2 - 2w^T x_i y_i + w^T x_i x_i^T w\right) + (e_0-1)\ln\lambda - f_0\lambda\right]$$

$$= \frac{n}{2}\ln\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}\left(y_i^2 - 2E(w)^T x_i y_i + x_i^T E(ww^T)x_i\right) + (e_0-1)\ln\lambda - f_0\lambda$$

$$\ln\lambda\left(\frac{n}{2}+e_0-1\right) - \lambda\left(\frac{1}{2}\sum_{i=1}^{n}\left(y_i^2 - 2E(w)^T x_i y_i + x_i^T E[ww^T]x_i\right) + f_0\right)$$

$$e' = \frac{n}{2} + e_0, \quad f' = \frac{1}{2}\sum_{i=1}^{n}\left[(y_i - \mu'^T x_i)^2 + x_i^T\Sigma' x_i\right] + f_0$$

$$q(\lambda) = \text{Gamma}(e', f')$$

And same thing with respect to $\alpha = (\alpha_1 \dots \alpha_d)$
we will solve for any α since αs are independent

$$E_{q(\alpha)}\left[\ln P(w|\alpha) + \ln P(\alpha)\right]$$

~~$\frac{n}{2}\ln(\lambda) - \frac{n}{2}\ln(2\pi) - \frac{\lambda}{2}\sum(y_i^2 - 2w^T x_i y_i + w^T x_i x_i^T w) + a_0 \ln b_0 + \ln\Gamma(a_0) + (a_0-1)\ln\alpha - b_0\alpha$~~
~~drop all the term that are independent of α~~

$$\frac{1}{2}\ln\alpha + \frac{1}{2}\ln(2\pi) - \frac{1}{2}w^T\varphi w) + a_0\ln b_0 - \ln\Gamma(a_0) + (a_0-1)\ln\alpha - b_0\alpha$$

$$E_{q(\alpha)}\left[\frac{1}{2}\ln\alpha - \frac{1}{2}(w^T\varphi w) + (a_0-1)\ln\alpha - b_0\alpha\right]$$

$$\frac{1}{2}\ln\alpha - \frac{1}{2}E(w^Tw) \cdot E(\varphi) + (a-1)\ln\alpha + b_0\alpha$$

$$= \ln \alpha_j \left(\frac{1}{2} + a_0 - 1\right) - \alpha_j \left(\frac{1}{2} E(\omega^T \varphi \omega) + b_0\right)$$

Thus we set

$$a' = \left(a_0 + \frac{1}{2}\right)$$
$$b' = \left(b_0 + \frac{1}{2} E(\omega^T \varphi \omega)\right)$$

$$q(\alpha) = \text{Gamma}(a', b')$$

Part b)

~~for t = 1 to T~~

~~Initialize $\mu'_t, \Sigma'_t, e'_t, f'_t, a'_t, b'_t$~~

for q (

Initialize $\mu_0, \Sigma_0, e_0, f_0, a_0, b_0$

set $\mu' = \mu_0, \Sigma' = \Sigma_0, e' = e_0, f' = f_0, a' = a_0, b' = b_0$

for $t = 1$ to T

update $q(\lambda), q(\alpha), q(\omega)$

Based of the values of $\mu', \Sigma', e', f', a', b'$

Check for convergence using $\mathcal{L}(\mu', \Sigma', e', f', a', b')$

~~keep updating~~

Part c

we seek the variational objective function

$\mathcal{L}(a, b, e, f, \mu, \Sigma)$

$$\mathcal{L} = \int q(\omega) q(\lambda) q(\alpha) \ln p(y, x, \omega, \lambda, \alpha) - \int q(\omega) \ln q(\omega)$$

$$- \int q(\lambda) \ln q(\lambda) - \int q(\alpha) \ln q(\alpha)$$

$$= E_q[\ln p(y, x, \omega, \lambda, \alpha)] - E_q[\ln q(\omega)] - E_q[\ln q(\lambda)]$$

$$- E_q[\ln q(\alpha)]$$

taking the first term

$$E_q[\ln p(y, x, \omega, \lambda, \alpha)] = E_q[\ln p(y \mid x, \omega, \lambda, \alpha)] + E_q[\ln p(\omega \mid \alpha)]$$

$$+ E_q[\ln p(\lambda)] + E_q[\ln p(\alpha)]$$

for λ

$$E_q[\ln p(\lambda \mid e_0, f_0)] - E_q[\ln q(\lambda \mid e', f')]$$

$$E_q[(e_0 \ln f_0 - \ln(\Gamma(e_0)) + (e_0 - 1) \ln \lambda - f_0 \lambda) - (e' \ln f' - \ln(\Gamma(e'))$$

$$+ (e' - 1) \ln \lambda - f' \lambda)$$

$$= e_0 \ln f_0 - \ln(\Gamma(e_0)) + (e - 1) E(\ln \lambda) - f_0 E(\lambda) - (e' \ln f' - \ln(\Gamma(e'))$$

$$+ e' - E(\ln \lambda) - f' E(\lambda)$$

$$= e_0 \ln f_0 - \ln(\Gamma(e_0)) + (e - 1)(\psi(e') - \ln f') - f_0 \frac{e'}{f_1} - e' \ln f'$$

$$- \ln(\Gamma(e')) + (e' - 1)(\psi(e') - \ln f') - f' \frac{e'}{f'}$$

$$= e_0 \ln f_0 - \ln \Gamma(e_0)) - (e' \ln f' + \ln(\Gamma(e')) + (e_0 - e')(\psi(e') - f')$$

$$- (f_0 - f_1) \frac{e'}{f'}$$

for α

$$\cancel{E_q[\ln p(\alpha \mid a_0, b_0)] - E[\ln \alpha}$$

$$E_q\left[\sum_{i=1}^{d} \ln p(\alpha \mid a_0, b_0)\right] - E\left[\sum_{i=1}^{d} \ln q(\alpha \mid a', b')\right]$$

$$d a_0 \ln b_0 - d \ln\Gamma(a_0) + (a_0-1)\sum_{k=1}^{d} E(\ln \alpha_k) - b_0 \sum_{i=1}^{d} E(\alpha_k)$$

$$-\sum_{k=1}^{d}\left[a' \ln b'_k - \ln\Gamma(a') + (a'-1)E(\ln\alpha_k) - b'_k E[\alpha_k]\right]$$

$$d a_0 \ln b_0 - d\ln\Gamma(a_0) + (a-1)\sum_{k=1}^{d}\left(\psi(a') - \ln b'_k\right) - b_0 \sum_{i=1}^{d} \frac{a'}{b'_k}$$

$$-\sum_{k=1}^{d}\left[a' \ln b'_k - \ln\Gamma(a') + (a'-1)(\psi(a') - \ln b'_k) - b'_k \frac{a'}{b'_k}\right]$$

$$= d(a_0 \ln b_0 - \ln\Gamma(a_0)) - (a' \sum_{k=1}^{d} \ln b'_k - d\ln\Gamma(a')) + (a_0 - a')\sum_{i=1}^{d}[(\psi(a')$$

$$- \ln b'_k)] - b_0 \sum_{k=1}^{d}\left[\frac{a'}{b'_k}\right] + d a'$$

for $E_q\left[\sum_{i=1}^{n} \ln p(y_i \mid x_i, \omega, \lambda)\right]$

$$\frac{n}{2} E(\ln\lambda) - \frac{n}{2}\ln 2\pi - \frac{E(\lambda)}{2}\sum_{i=1}^{n}\left(y_i^2 - 2E(\omega^T) x_i y_i + x_i^T E[\omega\omega^T] x_i\right]$$

$$= \cancel{\frac{n}{2} E(\ln \lambda)}$$

$$\frac{n}{2}(\psi(e') - \ln f') - \frac{n}{2}\ln 2\pi - \frac{f'}{2e'}\sum_{i=1}^{n}\left(y_i^2 - 2\mu'^T x_i y_i + x_i^T[\Sigma' + \mu'\mu'^T] x_i\right.$$

$$= \frac{n}{2}(\psi(e') - \ln f') - \frac{f'}{2e'}\sum_{i=1}^{n}\left[y_i - 2\mu'^T x_i y_i + x_i^T [\Sigma' + \mu'\mu'^T] x_i\right]$$

$E_q[\ln p(w|\alpha)] - E_q[\ln q(w|\mu', \Sigma)]$

$\frac{1}{2}\sum_{i=1}^{k} E(\ln\alpha_k) - \frac{d}{2}\ln 2\pi - \frac{1}{2}(E(w^T \operatorname{diag}(\alpha) w\} + \frac{1}{2}\ln\|\Sigma'\|$

$+ \frac{k}{2}\ln 2\pi + \frac{1}{2}E[(w-\mu')^T \Sigma^{-1}(w-\mu')]$

we calculate the expectation

$E(\ln\alpha_k) = \psi(a'_k) - \ln b'_k$

$E(w^T \operatorname{diag}(\alpha) w\} = \sum_{i=1}^{d} E(w_k^2) E(\alpha_k) = \sum_{k=1}^{d} [\Sigma'_k + [\mu'\mu'^T]_k \frac{a'}{b'_k}$

$E\{(w-\mu')^T \Sigma^{-1}(w-\mu')\} = E[\Sigma(w^T w - 2w\mu' + \mu'^T\mu'$

$= E(w^T \Sigma^{-1} w) - 2E(w)\Sigma'^{-1}\mu' + \mu'^T \Sigma'^{-1}\mu'$

$= E(w^T \Sigma'^{-1} w] - \mu'^T \Sigma'^{-1}\mu'$, this evaluates to

$$\sum_{k=1}^{d} [\Sigma'^{-1}_k \Sigma'_k] = d$$

pluggin in the Expectation and dropping cst,

$\frac{1}{2}\sum_{k=1}^{d} (\psi(a'_k) - \ln b'_k) - \frac{1}{2}\sum_{k=1}^{d} [[\Sigma'_k + \mu'\mu'^T]_k \frac{a'}{b'_k} + \frac{1}{2}\ln|\Sigma'|$

=> Objective function L

is

$(e_0 \ln f_0 - \ln\Gamma(e_0)) - (e'\ln f' + \ln\Gamma(e')) + (e_0 - e')(\psi(e') - \ln f')(f_0 - f')\frac{e'}{f'}$

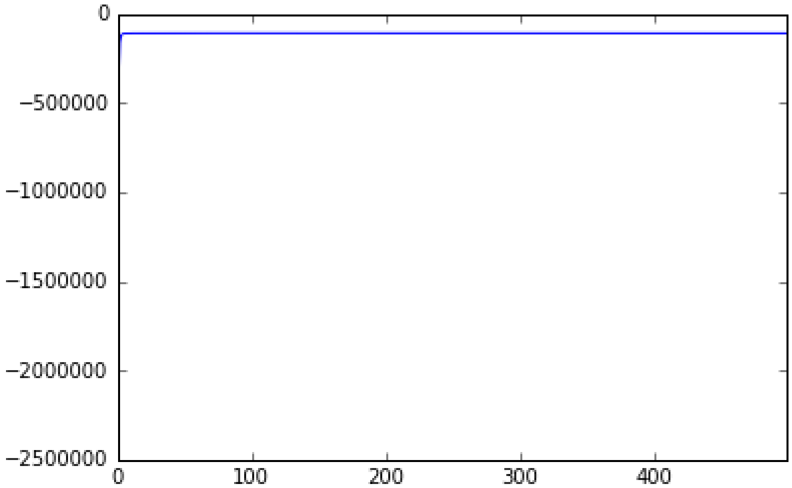
$+ d(a_0 \ln b_0 - \ln\Gamma(a_0)) - (a'\sum\ln b_k - d\ln\Gamma(a')) + (a_0 - a')\sum_{k=1}^{d}[(\psi(a') - \ln b'_k)]$

$- b_0 \sum_{k=1}^{d}\left[\frac{a'}{b'_k}\right] + d a' + \frac{1}{2}\sum_{k=1}^{d}(\psi(a'_k) - \ln b'_k) - \frac{1}{2}\sum_{k=1}^{d}[\Sigma'_k + \mu'\mu'^T]_k \frac{a'}{b'_k}]$

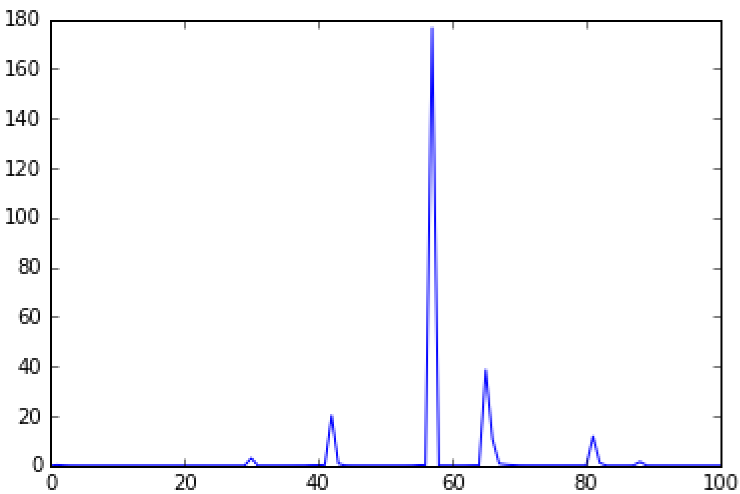
$+ \frac{1}{2}\ln|\Sigma'| + \frac{n}{2}(\psi(e') - \ln f') - \frac{f'}{2e'}\sum_{i=1}^{n}[y_i^2 - 2\mu'^T x_i y_i + x_i^T[\Sigma' + \mu'\mu'^T]x_i]$

For set 1:

a)



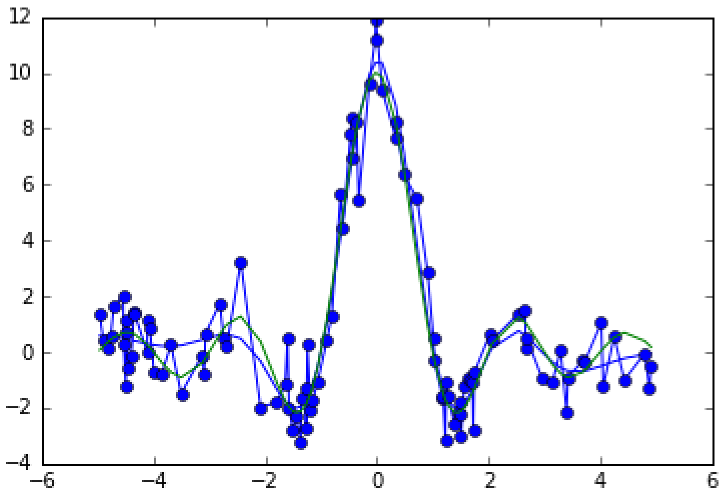
b)



c)

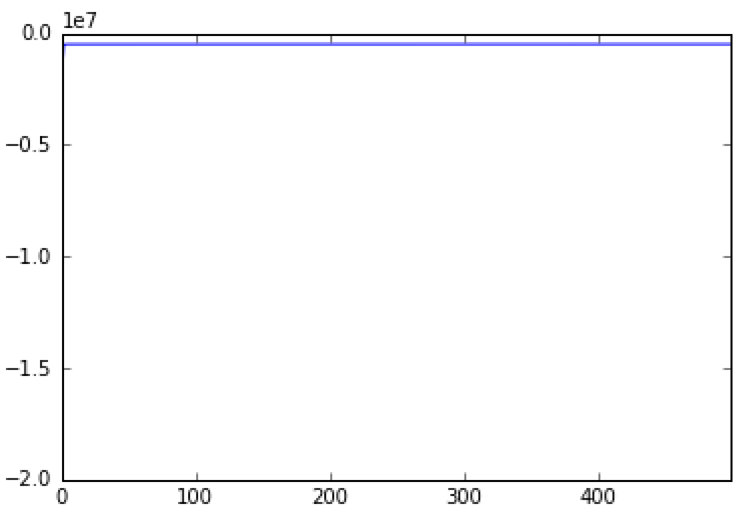
1.03035968976

d)

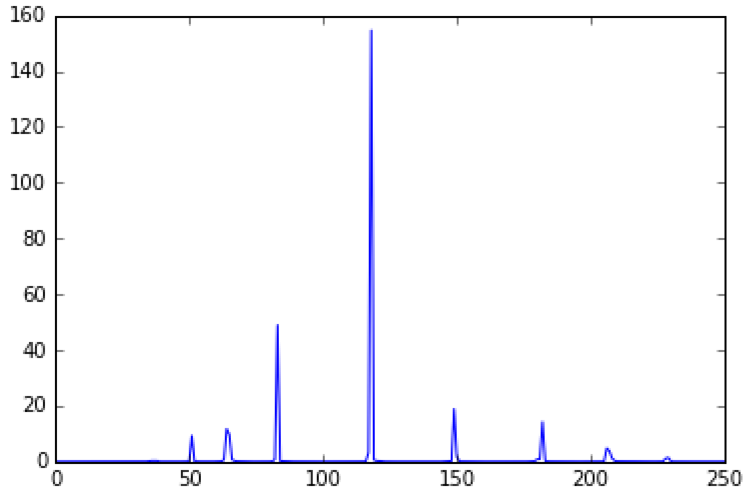


For set2

a)



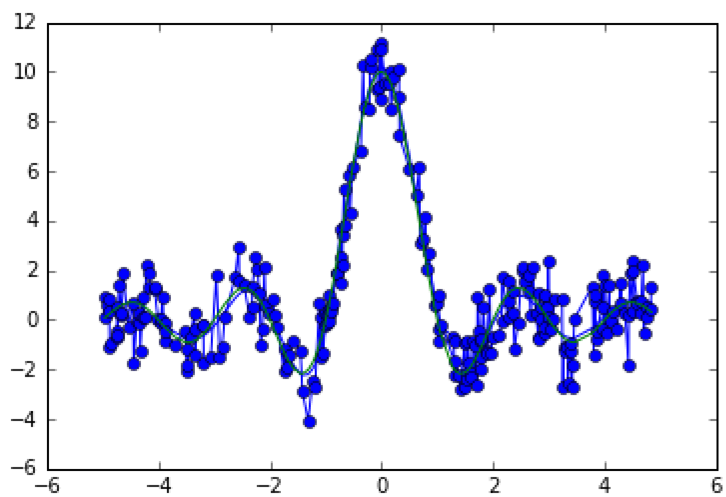
b)



c)

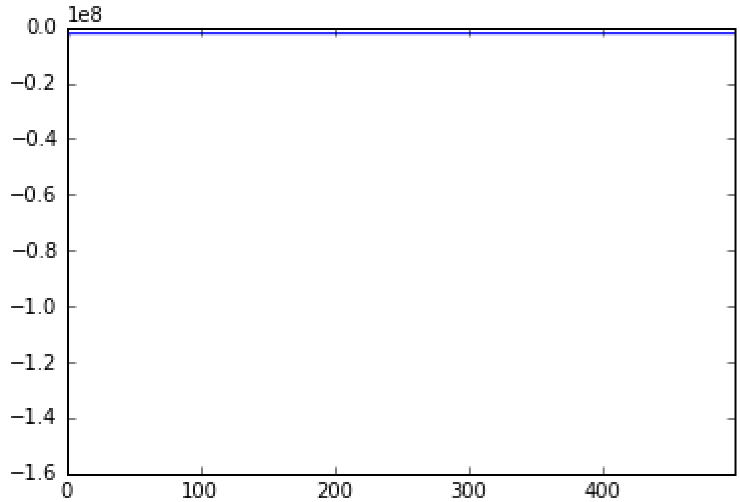
0.904139816884

d)

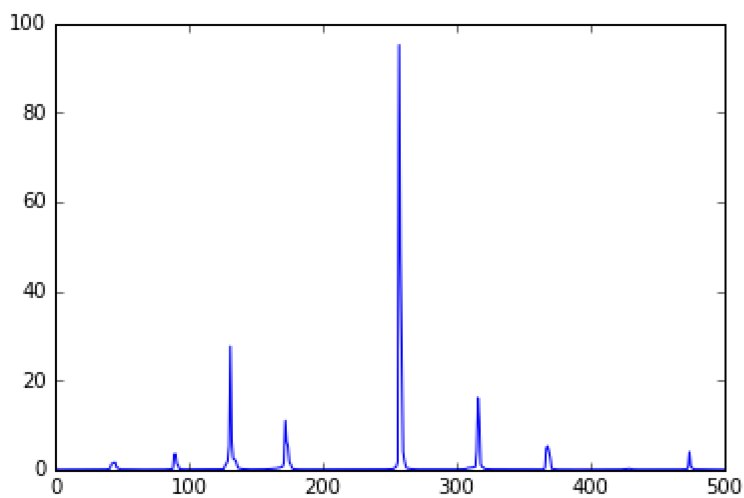


Set 3)

a)



b)



c)

0.974731805221

d)