Name: Jingyi Yuan
UNI: jy2736
Class: EECS E6892

# Homework 03

## Problem 1

Problem 1.

a). The posterior $p(w, \alpha_1, \dots \alpha_d, \lambda) \simeq q(w, \alpha_1 \dots \alpha_d, \lambda) = q(w) q(\lambda) \prod_{k=1}^{d} q(\alpha_k)$.

$$q(w) \propto \exp\{E_q[\ln p(y|x, w, \lambda) + \ln p(w|\alpha_1, \dots \alpha_d) + \ln p(\alpha_1, \dots \alpha_d) + \ln p(\lambda)]\}$$

$$\propto \exp\{E_q[\ln p(y|x, w, \lambda) + \ln p(w|\alpha_1, \dots \alpha_d)]\}.$$

$$\propto \exp\{\sum_{i=1}^{N} \frac{E_q[\lambda]}{2} (x_i^T w - y_i)^2\} \exp\{-\frac{1}{2} w^T \cdot \mathrm{diag}(E_q(\alpha_1), E_q(\alpha_2) \cdots E_q(\alpha_d)) \cdot w\}$$

$$q(w) = \mathrm{Normal}(w|\mu', \Sigma')$$

$$\Sigma' = \left(\mathrm{diag}(E_q(\alpha_1) \cdots E_q(\alpha_k)) + E_{q(\lambda)}(\lambda) \sum_{i=1}^{N} x_i x_i^T\right)^{-1} \quad \mu' = \Sigma'(E_{q(\lambda)}[\lambda] \sum_{i=1}^{N} y_i x_i).$$

$$q(\lambda) \propto \exp\{E_q[\ln p(y|x, w, \lambda) + \ln p(\lambda)]$$

$$\propto \exp\{\sum_{i=1}^{N} (\frac{1}{2}\ln\lambda - \frac{\lambda E(x_i^T w - y_i)^2}{2})\} \cdot p(\lambda)$$

$$\propto \exp\{\sum_{i=1}^{N} (\frac{1}{2}\ln\lambda - \frac{\lambda}{2} E(x_i^T w - y_i)^2\} \cdot \lambda^{e_0 - 1} e^{-f_0 \lambda}.$$

$$q(\lambda) = \mathrm{Gamma}(\lambda|e', f') \qquad e' = e_0 + \frac{N}{2} \qquad f' = f_0 + \frac{1}{2} \sum_{i=1}^{N} E_{q(w)}[(y_i - x_i^T w)^2].$$

$$q(\alpha_k) \propto \exp\{E_{q(w_k)}[\ln p(w_k|\alpha_k) + \ln p(\alpha_k)\}$$

$$\propto \exp\{-\frac{1}{2}\ln\alpha_k - \frac{\alpha_k E(w_k^2)}{2}\} \cdot \alpha_k^{a_0 - 1} e^{-b_0 \alpha_k}.$$

$$q(\alpha_k) = \mathrm{Gamma}(\alpha_k|a', b') \qquad a_k' = a_0 - \frac{1}{2} \qquad b_k' = b_0 + \frac{1}{2} E(w_k^2).$$

b). pseudo-code.

1. Initialize $\mu_0'$, $\Sigma_0'$, $e_0'$, $f_0'$, $a_{k0}'$ and $b_{k0}'$ in some way.

2. For iteration $t = 1, \dots, T$.

- update $q(w)$ by setting

$$\mu_t' = \Sigma_t' \left( \frac{e_{t-1}}{f_{t-1}} \sum_{i=1}^{N} y_i x_i \right)$$

$$\Sigma_t' = \left[ \mathrm{diag}\left( \frac{a_{1,t-1}'}{b_{1,t-1}'}, \dots \frac{a_{d,t-1}'}{b_{d,t-1}'} \right) + \frac{e_{t-1}'}{f_{t-1}'} \sum_{i=1}^{N} x_i x_i^T \right]^{-1}.$$

- update $q(\lambda)$ by setting

$$e' = e_0 + \frac{N}{2}$$

$$f' = f_0 + \frac{1}{2}\left[ \sum_{i=1}^{N} (y_i - x_i^T \mu_t')^2 + x_i^T \Sigma_t' x_i \right]$$

- update $q(\alpha_k)$ by setting.

$$a_{k,t}' = a_{0,t}' + \frac{1}{2}$$

$$b_{k,t}' = b_{0,t}' + \frac{1}{2}\left[ \mu_t' \mu_t'^T + \Sigma_t' \right]_{kk}.$$

- Evaluate $\mathcal{L}(\mu_t', \Sigma_t', e_t', f_t', a_{k,t}', b_{k,t}')$ to assess convergence.

($\mathcal{L}$ is obtained in part c)).

c). $\ln p(y|x) \geq \mathcal{L}(q(w, \alpha, x)) = \sum_{i=1}^{N} E[\ln p(y_i|x_i, w, \lambda)] + E[\ln p(w|\alpha)] + \sum_{k=1}^{d} E[\ln p(\alpha_k)]$
$+ E[\ln p(\lambda)] - E[\ln q(w)] - \sum_{k=1}^{d} E[\ln q(\alpha_k)] - E[\ln q(\lambda)].$

$$E[\ln p(y_i|x_i, w, \lambda)] = \frac{1}{2} E[\ln \lambda] - \frac{1}{2}\ln 2\pi - \frac{1}{2} E[\lambda (x_i^T w - y_i)^2]$$

$$= \frac{1}{2}(\psi(e') - \ln f') - \frac{1}{2}\ln 2\pi - \frac{1}{2} E[\lambda]\, E[(x_i^T \mu' - y_i)^2]$$

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

$$E[\ln p(w|\alpha)] = \sum_{i=1}^{d} E[\ln p(w_k|\alpha_k)]$$

$$E[\ln p(w_k|\alpha_k)] = \frac{1}{2} E[\ln \alpha_k] - \frac{1}{2}\ln 2\pi - \frac{1}{2} E[\alpha_k] \cdot E[w_k^2]$$

$$= \frac{1}{2}(\psi(a_k') - \ln b_k') - \frac{1}{2}\ln 2\pi - \frac{1}{2}\frac{a_k}{b_k}(\mu_t' \mu_t'^T + \Sigma_t')_{kk}$$

$$E[\ln p(\alpha_k)] = a_0 \ln b_0 - \ln \Gamma(a_0) + (a_0 - 1) E[\ln \alpha_k] - b_0 E[\alpha_k]$$

$$= a_0 \ln b_0 - \ln \Gamma(a_0) + (a_0 - 1)\cdot(\psi(a_k') - \ln b_k') - b_0 \cdot \frac{a_k'}{b_k'}.$$

$$E[\ln p(\lambda)] = e_0 \ln f_0 - \ln \Gamma(e_0) + (e_0 - 1) E[\ln \lambda] - f_0 E[\lambda]$$

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

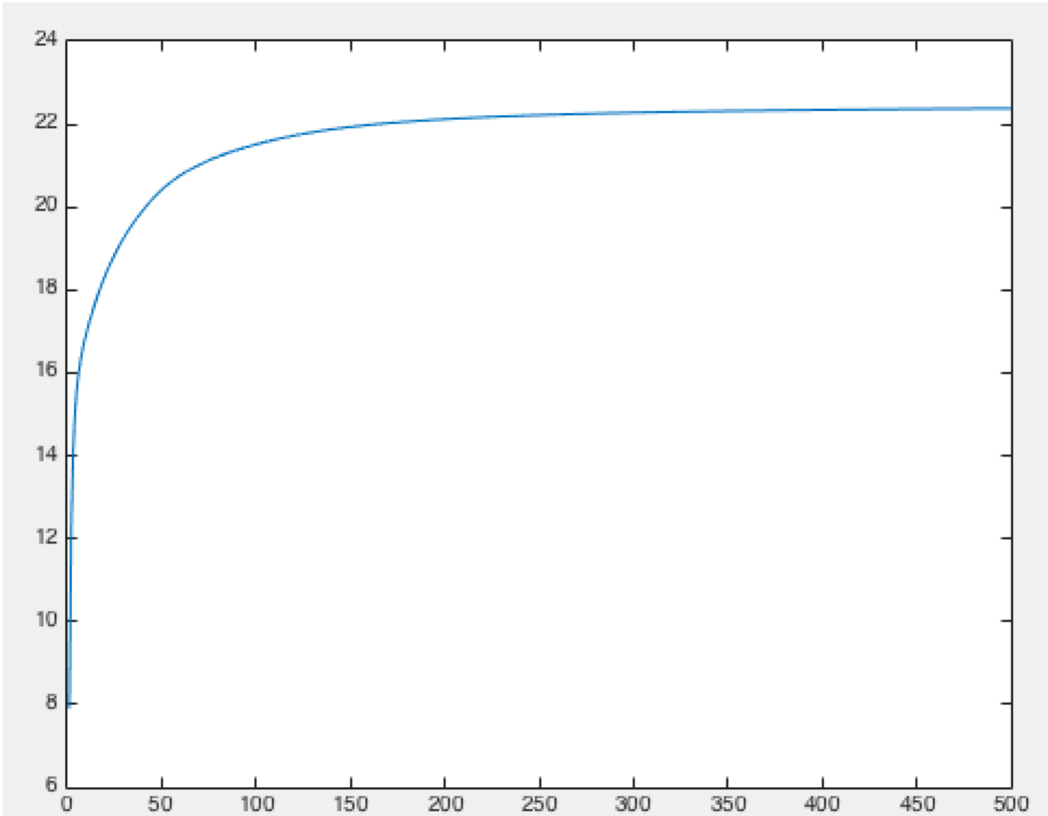
$$
\begin{aligned}
\therefore \ln p(y|x) \geq \mathcal{L}(q(w,\alpha,x)) = & \sum_{i=1}^{N} \left\{ \frac{1}{2}[\psi(e') - \ln f'] - \frac{1}{2}\ln 2\pi - \frac{1}{2}\frac{e'}{f'}\left[(x_i^T \mu' - y_i)^2 + x_i^T \Sigma' x_i\right] \right\} \\
& + \sum_{k=1}^{d} \left\{ \frac{1}{2}[\psi(a_k') - \ln b_k'] - \frac{1}{2}\ln 2\pi - \frac{1}{2}\frac{a_k'}{b_k'} (\mu_t' \mu_t'^T + \Sigma_t')_{tt} \right\} \\
& + \sum_{k=1}^{d} \left\{ a_0 \ln b_0 - \ln\Gamma(a_0) + (a_0 - 1)(\psi(a_k') - \ln b_k') - b_0 \frac{a_k'}{b_k'} \right\} \\
& + e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0 - 1)(\psi(e') - \ln f') - f_0 \frac{e'}{f'} \\
& + \left[ \frac{d}{2}(1 + \ln 2\pi) + \frac{1}{2}\ln|\Sigma'| \right] \\
& + \left\{ e' - \ln f' + \ln\Gamma(e') + (1 - e')\psi(e') \right\} \\
& + \sum_{k=1}^{d} \left\{ a_k' - \ln b_k' + \ln\Gamma(a_k') + (1 - a_k')\psi(a_k') \right\}.
\end{aligned}
$$

$$
\begin{aligned}
\therefore \ln p(y|x) \geq & \ \mathcal{L}(q(w,\alpha,x)) \\
= & \sum_{i=1}^{N} \left\{ \frac{1}{2}[\psi(e') - \ln f'] - \frac{1}{2}\frac{e'}{f'}\left[(x_i^T \mu' - y_i)^2 + x_i^T \Sigma' x_i\right] \right\} + \text{constant} \\
& + \sum_{k=1}^{d} \left\{ \frac{1}{2}[\psi(a_k') - \ln b_k'] - \frac{1}{2}\frac{a_k'}{b_k'} (\mu_t' \mu_t'^T + \Sigma_t')_{kk} \right\} + \text{constant} \\
& + \sum_{k=1}^{d} \left\{ (a_0 - 1)(\psi(a_k') - \ln b_k') - b_0 \frac{a_k'}{b_k'} \right\} + \text{constant}. \\
& + (e_0 - 1)(\psi(e') - \ln f') - f_0 \frac{e'}{f'} + \text{constant} \\
& + \frac{1}{2}\ln|\Sigma'| + \text{constant} \\
& + \sum_{k=1}^{d} \left\{ a_k' - \ln b_k' + \ln\Gamma(a_k') + (1 - a_k')\psi(a_k') \right\} \\
& + e' - \ln f' + \ln\Gamma(e') + (1 - e')\psi(e')
\end{aligned}
$$

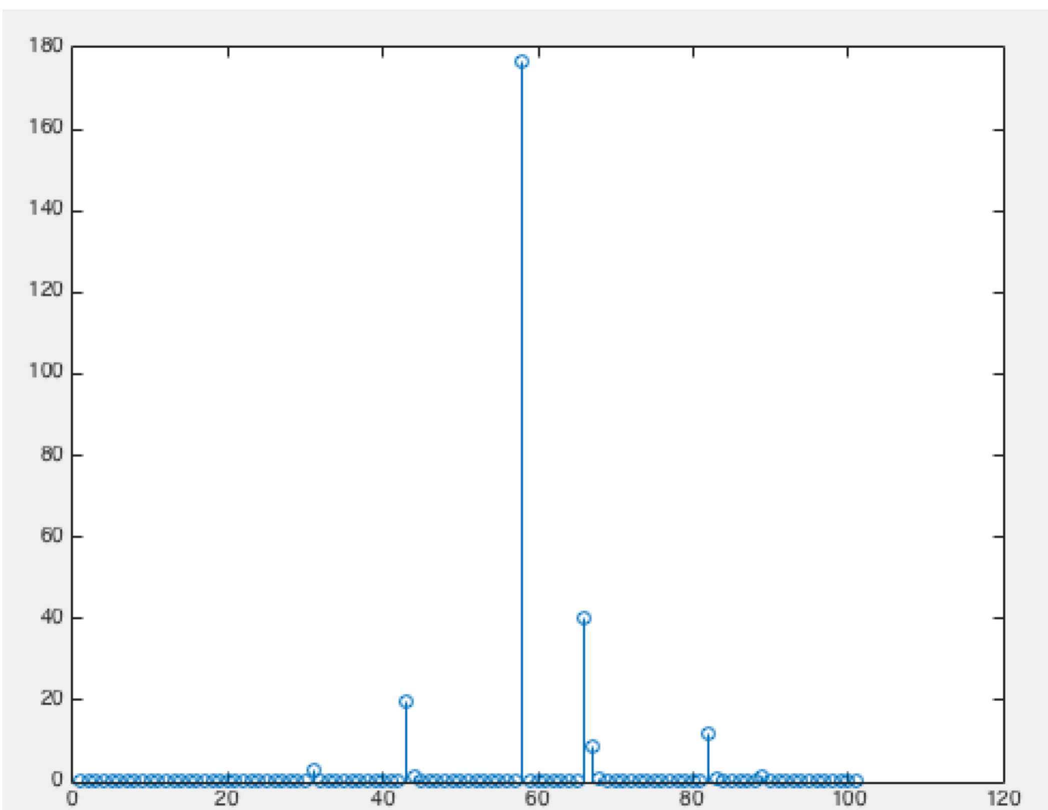
## Problem 2

**data1:**

**a)**


24
22
20
18
16
14
12
10
8
6
0
50
100
150
200
250
300
350
400
450
500

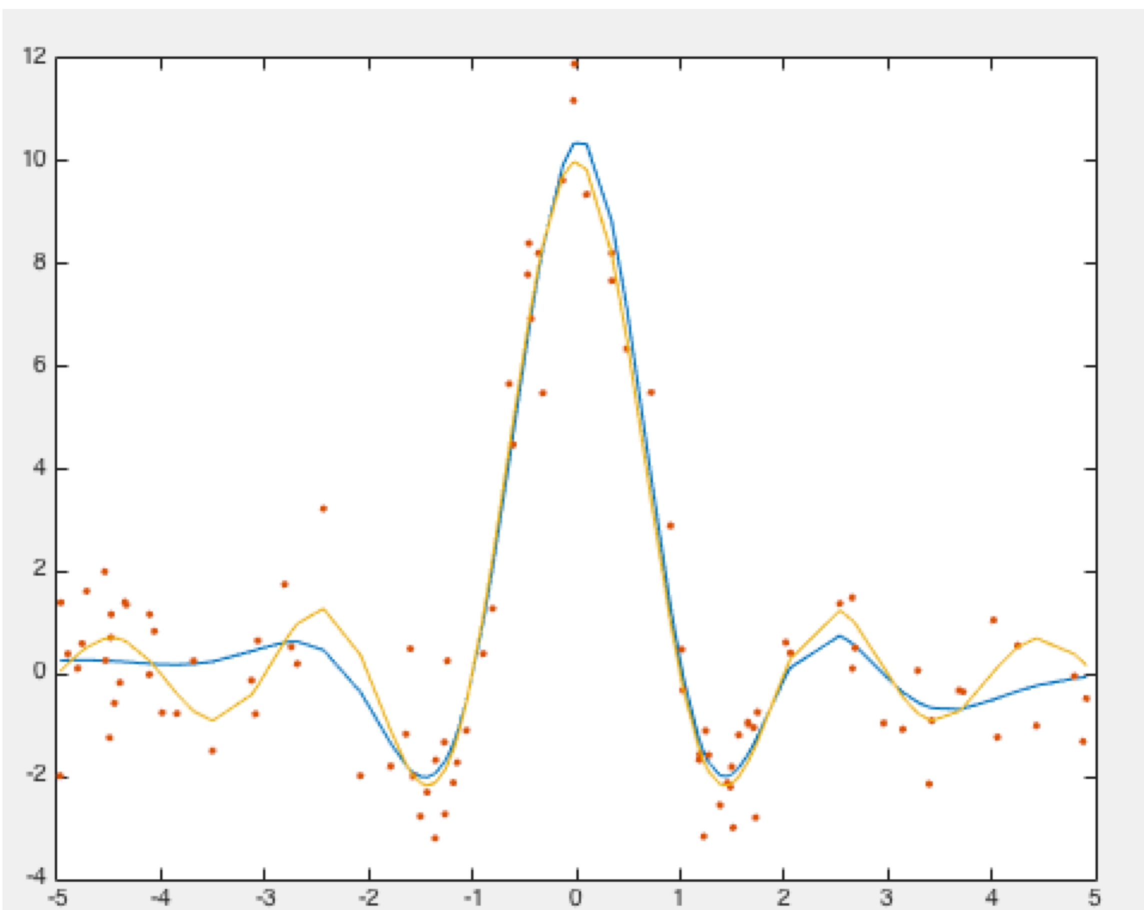
**b)**


180
160
140
120
100
80
60
40
20
0
0
20
40
60
80
100
120

**c)**

```
>> Eqlambda

Eqlambda =

    0.9261

>> 1/Eqlambda

ans =

    1.0798
```

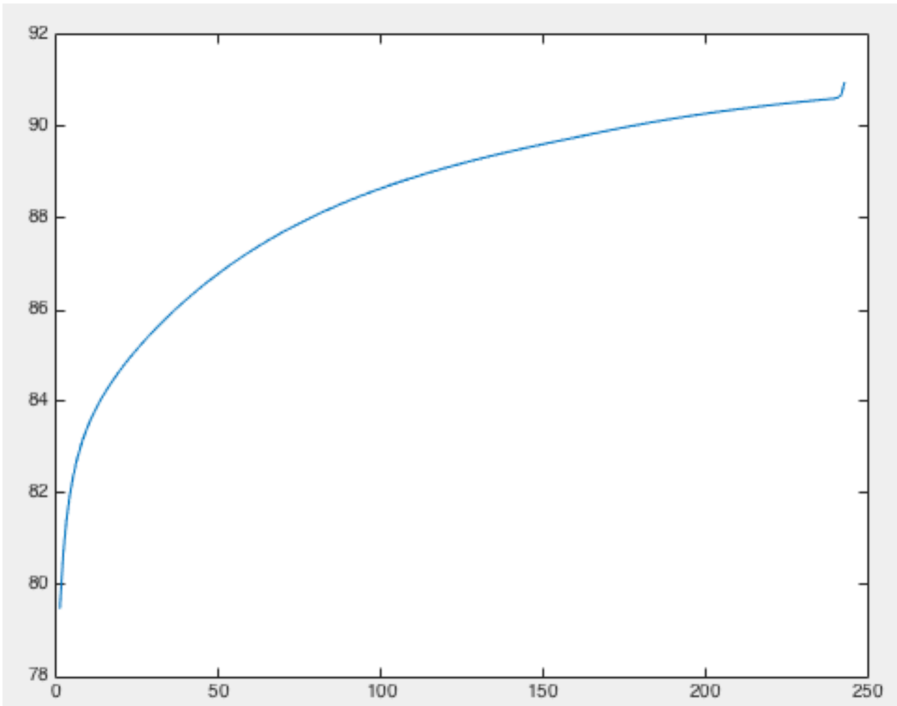
**d)**



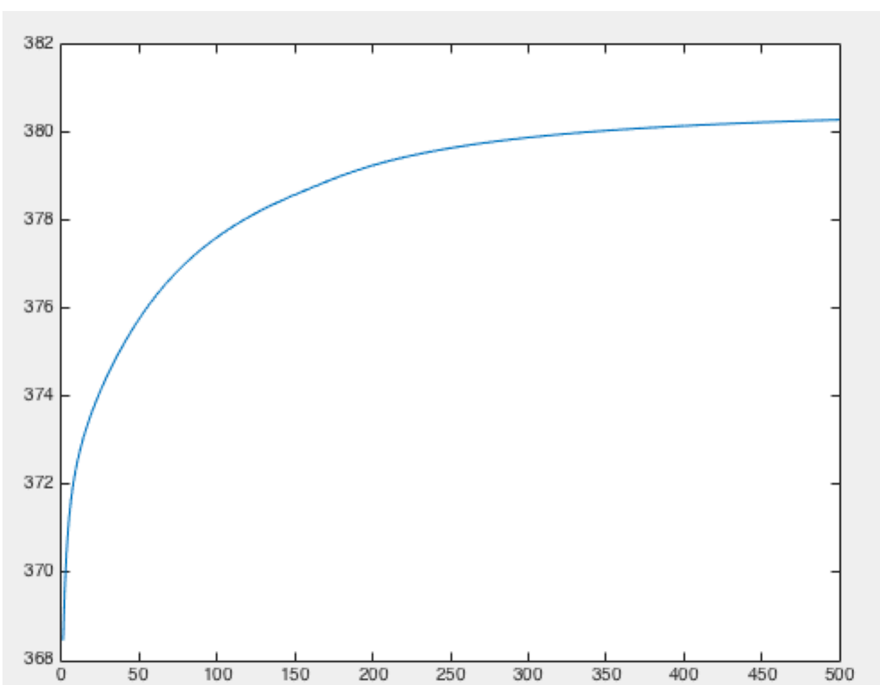
**data2:**

**a)**

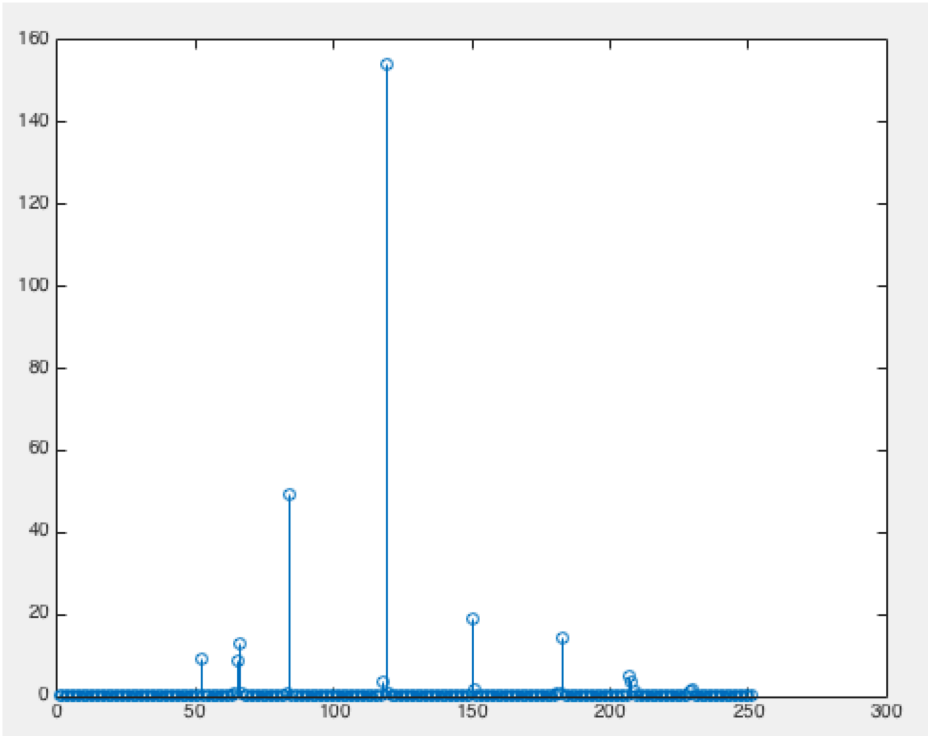
When t becomes large, the L goes to INF in MATLAB.




We change E[lnq(w)] to be $\frac{1}{2}|10^{*}sigma|$ instead of $\frac{1}{2}|sigma|$, which is 0 in MATLAB. Then the variational function is like:



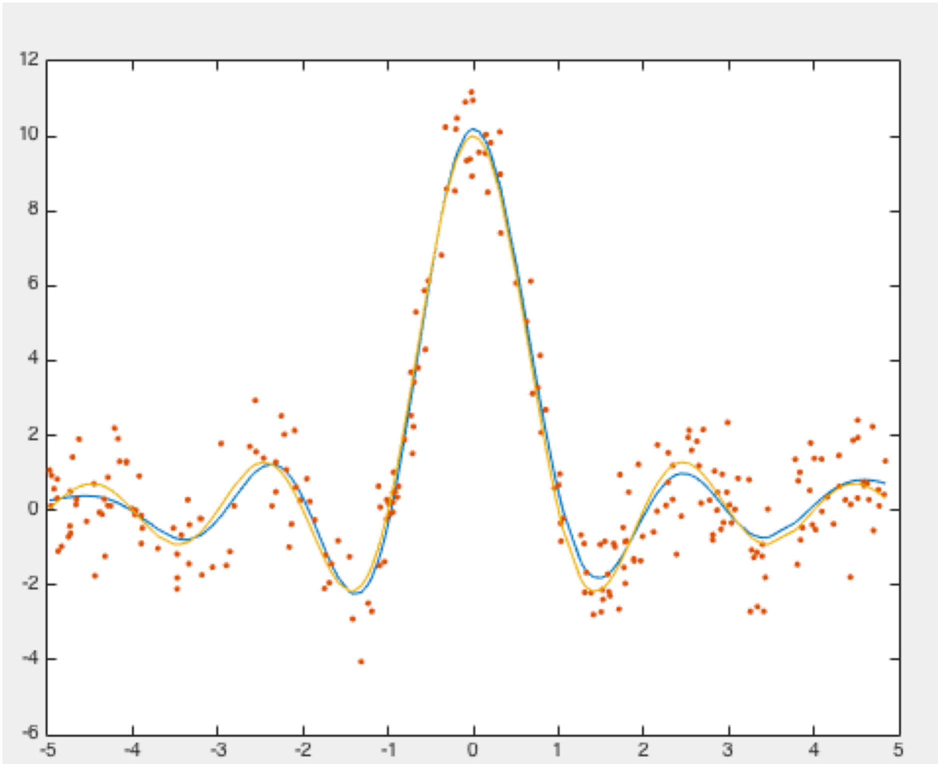
**b)**




**c)**

```
Eqlambda =

    1.1118

>> 1/Eqlambda

ans =

    0.8994
```

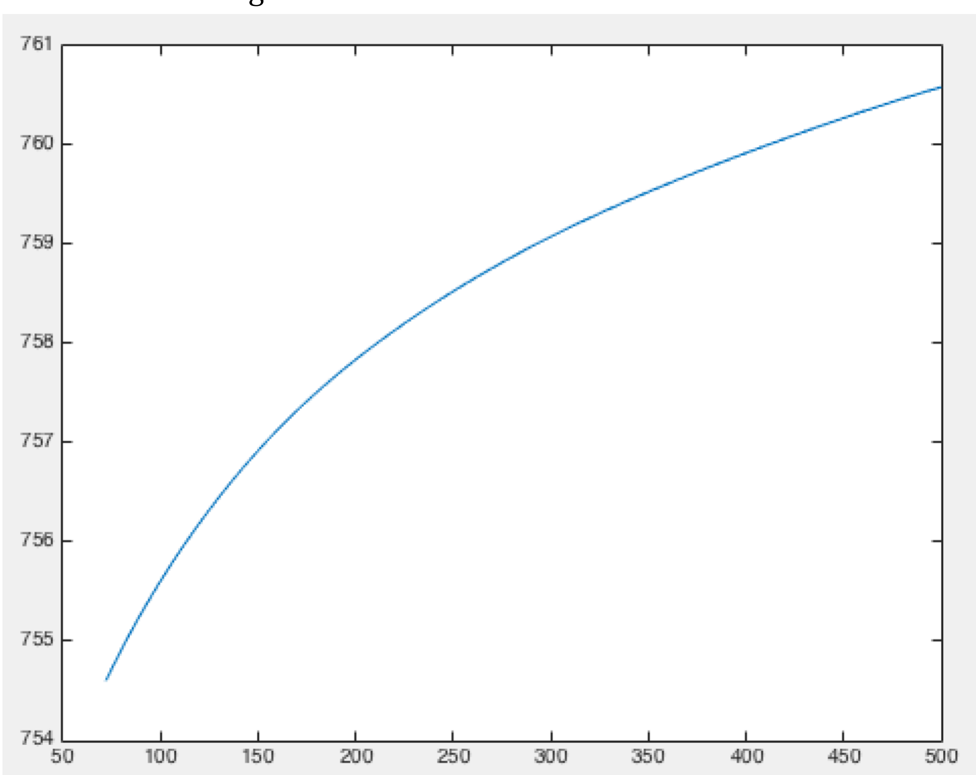
**d)**



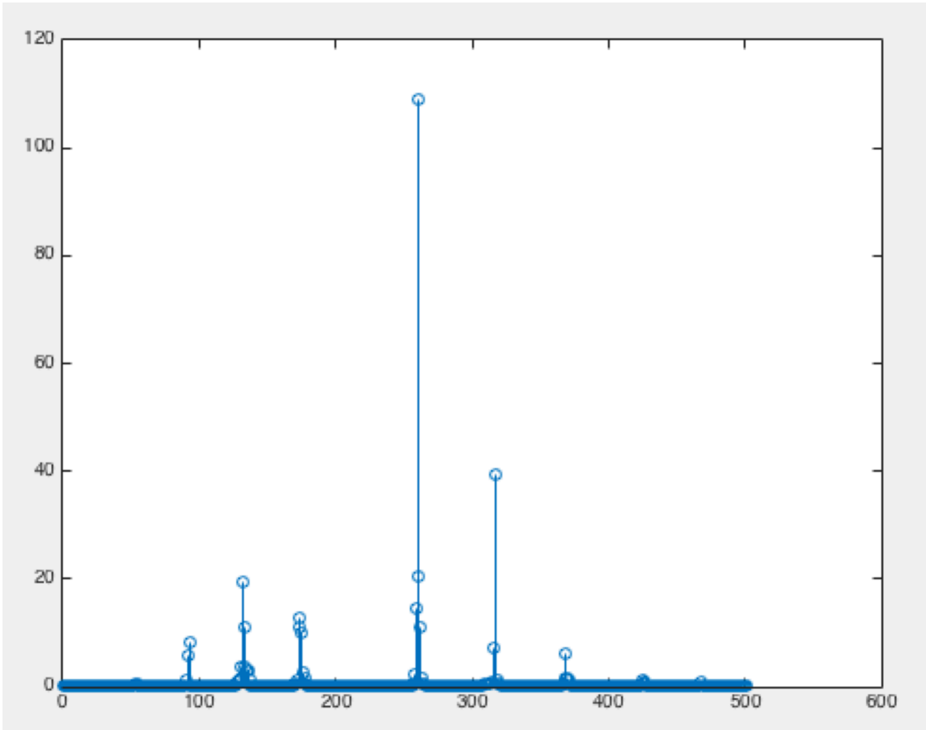

**data3:**

**a)**

It does not converge when t = 500.



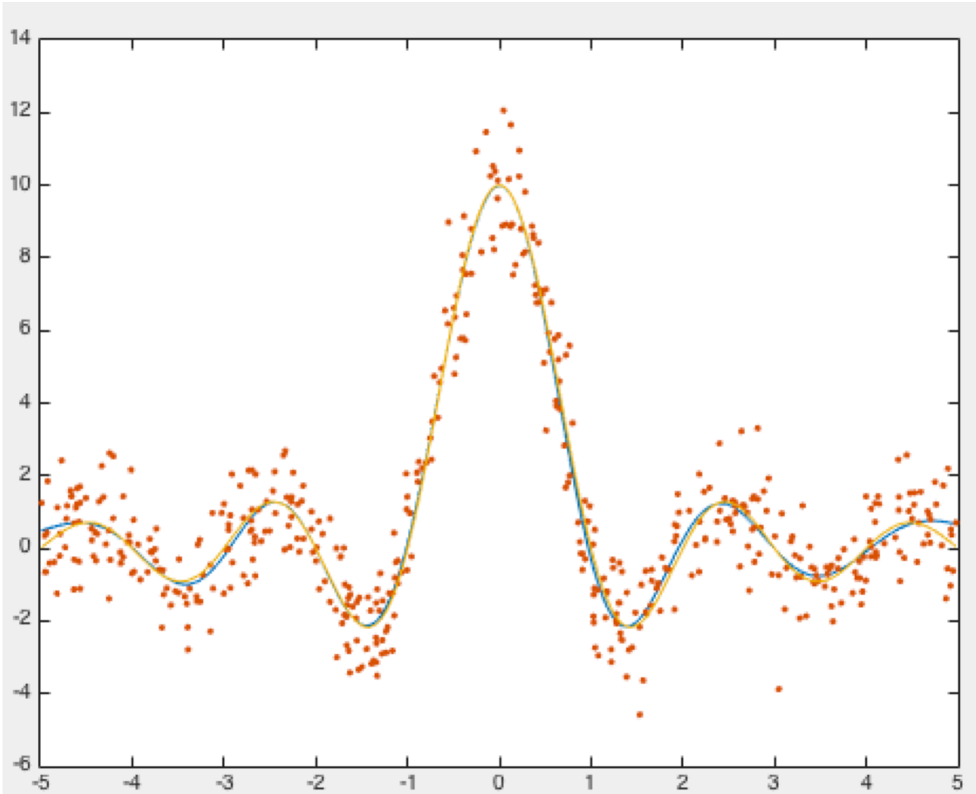
**b)**




**c)**

```
Eqlambda =

    1.0224

>> 1/Eqlambda

ans =

    0.9781
```

**d)**


14
12
10
8
6
4
2
0
-2
-4
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5