EECS 6892 BMML HW#3 UNI: sl3763 Name: Sung-Yen Liu

Problem 1.

$y_i \overset{ind}{\sim} N(0, \text{diag}(\alpha_1, \dots, \alpha_d)^{-1})$

$\alpha_K \overset{iid}{\sim} \text{Gamma}(a_0, b_0)$

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

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

Univariate normal:

$\ln N(x|\mu, \lambda^{-1}) = \frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{\lambda}{2}(x-\mu)^2$
$= -\frac{1}{2}\lambda x^2 + \lambda\mu x + c$

Multivariate normal:

$\ln N_d(x|\mu, \Lambda^{-1}) = -\frac{1}{2}\ln|\Lambda^{-1}| - \frac{d}{2}\ln 2\pi - \frac{1}{2}(x-\mu)^T\Lambda(x-\mu)$
$= -\frac{1}{2}x^T\Lambda x + x^T\Lambda\mu + c$

Gamma:

$\ln \text{Gam}(x|a,b) = a\ln b - \ln\Gamma(a) + (a-1)\ln x - bx$
$= (a-1)\ln x - bx + c$

$q(w, \alpha_1, \dots, \alpha_d, \lambda) \approx p(w, \alpha_1, \dots, \alpha_d, \lambda | y, x)$

$q(w, \alpha_1, \dots, \alpha_d, \lambda) = q(w)\, q(\lambda) \prod_{K=1}^{d} q(\alpha_K)$

Since α_K are independent, assume $q(\alpha) = \prod_{K=1}^{d} q(\alpha_K) = \prod_{i=1}^{d} \text{Gam}(\alpha_i | a_0, b_0)$

$= \prod_{i}^{d} \frac{1}{\Gamma(a_0)} b_0^{a_0} \alpha_i^{a_0 - 1} \exp(-b_0 \alpha_i)$

Joint likelihood: $p(y, w, \alpha, \lambda | x) = p(y|w, \lambda, x)\, p(w|\alpha)\, p(\alpha)\, p(\lambda)$

$$\ln q(\lambda) = \mathbb{E}_{q(\alpha,w)} \ln p(y|w,\lambda,x) + \underbrace{\mathbb{E}_{q(\alpha,w)} \ln p(w|\alpha)}_{\text{const w.r.t } \lambda} + \underbrace{\mathbb{E}_{q(\alpha,w)} \ln p(\alpha)}_{\text{const w.r.t. } \lambda} + \mathbb{E}_{q(\alpha,w)} \ln p(\lambda)$$

$$= \mathbb{E}_{q(\alpha,w)} \ln p(y|w,\lambda,x) + \mathbb{E}_{q(\alpha,w)} \ln p(\lambda) + c$$

$$= \mathbb{E}_{q(\alpha,w)} \left[\sum_{i=1}^{n} \left(\frac{1}{2}\ln\lambda \underbrace{- \frac{1}{2}\ln(2\pi)}_{c} - \frac{\lambda}{2}(y - w^T x)^2 \right) \right] + \mathbb{E}_{q(\alpha,w)} \left[\underbrace{e_0 \ln f_0 - \ln\Gamma(e_0)}_{c} + (e_0 - 1)\ln\lambda - f_0\lambda \right] + c$$

$$= \frac{n}{2}\ln\lambda - \frac{\lambda}{2} y^T y + \lambda \mathbb{E}_{q(w)}[w^T] x^T y - \frac{\lambda}{2} \mathbb{E}_{q(w)}[w^T] x^T x\, \mathbb{E}_{q(w)}[w] + (e_0 - 1)\ln\lambda - f_0\lambda + c$$

$$= \left(e_0 + \frac{n}{2} - 1\right)\ln\lambda - \left\{ \frac{1}{2} y^T y - \mathbb{E}_{q(w)}[w^T] x^T y + \frac{1}{2}\mathbb{E}_{q(w)}[w^T] x^T x\, \mathbb{E}_{q(w)}[w] + f_0 \right\} \lambda + c$$

$\Rightarrow q^*(\lambda) = e^{c_1 \ln\lambda - c_2 l + c_3}$ is in the form of Gamma distribution $= \text{Gamma}(\lambda | e_n, f_n)$

with
$$\begin{cases} e_n = e_0 + \frac{n}{2} \\ f_n = f_0 + \frac{1}{2} y^T y - \mathbb{E}_{q(w)}[w^T] x^T y + \frac{1}{2}\mathbb{E}_{q(w)}[w^T] x^T x\, \mathbb{E}_{q(w)}[w] \end{cases}$$

$$\ln q(w) = \mathbb{E}_{q(\alpha,\lambda)} \ln p(y|w,\lambda,x) + \mathbb{E}_{q(\alpha,\lambda)} \ln p(w|\alpha) + \underbrace{\mathbb{E}_{q(\alpha,\lambda)} \ln p(\alpha)}_{\text{const w.r.t } w} + \underbrace{\mathbb{E}_{q(\alpha,\lambda)} \ln p(\lambda)}_{\text{const w.r.t } w}$$

$$= \sum_{i=1}^{n} \mathbb{E}_{q(\alpha,\lambda)} \left[\underbrace{\frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi}_{c} - \frac{\lambda}{2}(y_i - w^T x_i)^2 \right]$$

$$+ \mathbb{E}_{q(\alpha)} \left[\underbrace{-\frac{1}{2}\ln|\text{diag}(\alpha_1, \dots, \alpha_d)^{-1}| - \frac{d}{2}\ln 2\pi}_{\text{const w.r.t } w} - \frac{1}{2} w^T \text{diag}(\alpha_1, \dots, \alpha_d)\, w \right] + c$$

$$= -\frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2}\sum_{i=1}^{n}(y_i - w^T x_i)^2 - \frac{\mathbb{E}_{q(\alpha)}[\mathrm{diag}(\alpha_1,\dots,\alpha_d)]}{2} w^T w + c$$

$$= \underbrace{-\frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2} y^T y}_{c} + \mathbb{E}_{q(\lambda)}[\lambda]\, w^T X^T y - \frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2} w^T X^T X w - \frac{\mathbb{E}_{q(\alpha)}[\mathrm{diag}(\alpha_1,\dots,\alpha_d)]}{2} w^T w + c$$

$$= -\frac{1}{2} w^T \left[\mathbb{E}_{q(\lambda)}[\lambda] X^T X + \mathbb{E}_{q(\alpha)}[\mathrm{diag}(\alpha_1,\dots,\alpha_d)]\right] w + w^T \left[\mathbb{E}_{q(\lambda)}[\lambda] X^T y\right] + c$$

$\Rightarrow$ $q(w)$ is in the form of Normal distribution $= N_d(w \mid \mu_n, \Sigma_n)$,

with
$$\begin{cases} \mu_n = \Sigma_n \mathbb{E}_{q(\lambda)}[\lambda] X^T y \\ \Sigma_n = \left(\mathbb{E}_{q(\alpha)}[\mathrm{diag}(\alpha_1,\dots,\alpha_d)] + \mathbb{E}_{q(\lambda)}[\lambda] X^T X\right)^{-1} \end{cases}$$

$$\ln q(\alpha) = \underbrace{\mathbb{E}_{q(w,\lambda)}[\ln p(y \mid w,\lambda,X)]}_{\text{const w.r.t } \alpha} + \mathbb{E}_{q(w,\lambda)}[\ln p(w \mid \alpha)] + \mathbb{E}_{q(w,\lambda)}[\ln p(\alpha)] + \underbrace{\mathbb{E}_{q(w,\lambda)}[\ln p(\lambda)]}_{\text{const w.r.t } \alpha}$$

$$= \mathbb{E}_{q(w)}[\ln p(w \mid \alpha)] + \mathbb{E}_{q(w,\lambda)}[\ln p(\alpha)] + c$$

$$= \mathbb{E}_{q(w)}\left[-\frac{1}{2}\ln|\mathrm{diag}(\alpha_1,\dots,\alpha_d)^{-1}| \underbrace{- \frac{d}{2}\ln 2\pi}_{c} - \frac{1}{2} w^T \mathrm{diag}(\alpha_1,\dots,\alpha_d) w\right]$$

$$+ \sum_{i=1}^{d} \mathbb{E}_{q(w,\lambda)}\left[\underbrace{a_0 \ln b_0 - \ln\Gamma(a_0)}_{c} + (a_0 - 1)\ln\alpha_i - b_0 \alpha_i\right] + c$$

$$= -\frac{1}{2}\sum_{i=1}^{d}\ln\alpha_i^{-1} - \frac{1}{2}\mathbb{E}_{q(w)}[w^T \mathrm{diag}(\alpha_1,\dots,\alpha_d) w] + \sum_{i=1}^{d}[(a_0-1)\ln\alpha_i - b_0\alpha_i] + c$$

$$= \sum_{i=1}^{d}(a_0 - \tfrac{1}{2})\ln\alpha_i - b_0\alpha_i - \frac{1}{2}\mathbb{E}_{q(w)}[w^T \mathrm{diag}(\alpha_1,\dots,\alpha_d) w] + c$$

$$\Rightarrow q(\alpha_i) = \left(a_0 + \frac{1}{2} - 1\right)\ln\alpha_i - \alpha_i\left(b_0 + \frac{1}{2}\mathbb{E}_{q(w)}[w_i^2]\right) + c$$

is in the form of Gamma distribution, with (for $i = 1,\dots,d$)

$$\begin{cases} a_n = a_0 + \frac{1}{2} \\ b_n = b_0 + \frac{1}{2}\mathbb{E}_{q(w)}[w_i^2] \end{cases}$$

Therefore, we have the distribution of $q(w)$, $q(\lambda)$ and $q(\alpha_i)$, the whole iteratively-updating algorithm is as following:

VI algorithm:

1. Initialize a_0', b_0', μ_0', Σ_0', e_0', f_0'

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

$$f_0 + \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}_{q(w)}[(y_i - w^T x_i)^2]$$
$$= f_0 + \frac{1}{2}\sum_{i=1}^{n}[(y_i - \mu_{t-1}' x_i)^2 + x_i^T \Sigma_{t-1}' x_i]$$
$$\|$$

- Update $q_t(\lambda)$ by setting $\begin{cases} e_t' = e_0 + \frac{n}{2} \\ f_t' = f_0 + \frac{1}{2} y^T y - \mathbb{E}_{q(w)}[W^T] X^T y + \frac{1}{2}\mathbb{E}_{q(w)}[W^T] X^T X \mathbb{E}_{q(w)}[W] \end{cases}$

- Update $q_t(w)$ by setting $\begin{cases} \mu_t' = \Sigma_t' \mathbb{E}_{q_t(\lambda)}[\lambda] X^T y \\ \Sigma_t' = \left(\mathbb{E}_{q_t(\alpha)}[\mathrm{diag}(\alpha_1, \ldots, \alpha_d)] + \mathbb{E}_{q_t(\lambda)}[\lambda] X^T X\right)^{-1} \end{cases}$

- Update $q_t(\alpha_i)$ by setting (for each $i = 1 \sim d$) $\begin{cases} a_t' = a_0 + \frac{1}{2} \\ b_t' = b_0 + \frac{1}{2}\mathbb{E}_{q_t(w)}[w_i^2] \end{cases}$

With $\mathbb{E}_{q_t(\alpha)}[\mathrm{diag}(\alpha_1, \ldots, \alpha_d)]$ is a diagonal matrix A with elements

$$A_{ii} = \mathbb{E}_{q_t(\alpha_i)}[\alpha_i] = \frac{a_t'}{b_t'}$$

$$\mathbb{E}_{q_t(\lambda)}[\lambda] = \frac{e_t'}{f_t'}, \quad \mathbb{E}_{q_t(w)}[w_i^2] = (\mu_{ti}')^2 + (\Sigma_t')_{ii}$$

Joint likelihood: $p(y, w, \alpha, \lambda | X) = p(y | w, \lambda, X)\, p(w|\alpha)\, p(\alpha)\, p(\lambda)$

Variational object function:

$$\mathcal{L}(a_t', b_t', \mu_t', \Sigma_t', e_t', f_t') = \sum_{i=1}^{n} \mathbb{E}[\ln p(y_i | x_i, w, \lambda)] + \mathbb{E}[\ln p(w|\alpha)] + \sum_{k=1}^{d} \mathbb{E}[\ln P(\alpha_k)] + \mathbb{E}[\ln p(\lambda)]$$
$$- \mathbb{E}[\ln q(w)] - \sum_{k=1}^{d} \mathbb{E}[\ln q(\alpha_k)] - \mathbb{E}_q[\ln q(\lambda)]$$

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

① if $q(\lambda) \sim \mathrm{Gamma}(e, f)$, $\mathbb{E}_{q(\lambda)}[\ln \lambda] = \psi(e) - \ln f$

② $\mathbb{E}[\lambda (y_i - x_i^T w)^2] = \mathbb{E}[\lambda]\,\mathbb{E}[(y_i - x_i^T w)^2] = \frac{e}{f} \cdot \mathbb{E}[(y_i - x_i^T w)^2]$

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

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

$$\mathbb{E}[\alpha_k] \cdot \mathbb{E}[w_k^2]$$

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

(4) Last three terms: $\mathbb{E}[\ln q(w)]$, $\sum_{k=1}^{d} \mathbb{E}[\ln q(\alpha_k)]$, $\mathbb{E}[\ln q(\lambda)] \Rightarrow$ Entropy of each distribution

$$\mathbb{E}[\ln q(\lambda)] = \ln \Gamma(a) - (a-1)\psi(a) - \ln b + a$$
$$\mathbb{E}[\ln q(w)] = -\left(\frac{1}{2}\ln|\Sigma| + \frac{d}{2}(1+\ln(2\pi))\right)$$

(by appendix from PRML)

Combine them all together, we have

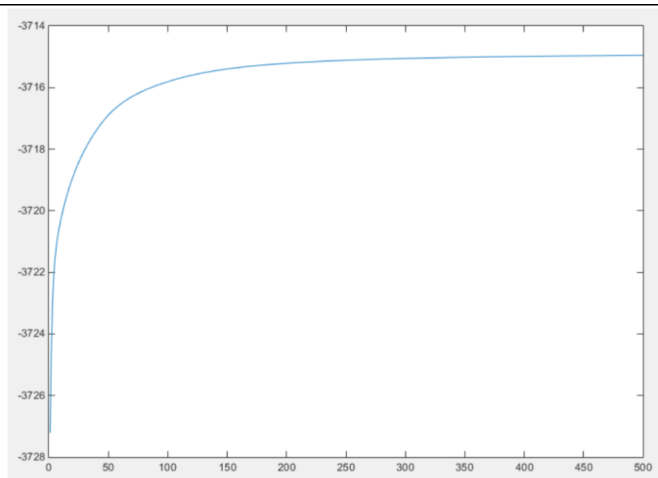
$$\mathcal{L}(a_t', b_t', \mu_t', \Sigma_t', e_t', f_t') = \left[\frac{n}{2}(\psi(e_t') - \ln f_t') - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\cdot\frac{e_t'}{f_t'}\sum_{i=1}^{n}\left((y_i - \mu_t'^T x_i)^2 + x_i^T \Sigma_t' x_i\right)\right]$$
$$+ \sum_{k=1}^{d}\left[\frac{1}{2}(\psi(a_{tk}') - \ln(b_{tk}')) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\left(\frac{a_{tk}'}{b_{tk}'}\right)\cdot\left[(\mu_{tk}')^2 + (\Sigma_t')_{kk}\right]\right]$$
$$+ \sum_{k=1}^{d}\left[a_0 \ln b_0 - \ln\Gamma(a_0) + (a_0-1)(\psi(a_{tk}') - \ln(b_{tk}')) - b_0\frac{a_{tk}'}{b_{tk}'}\right]$$
$$+ \left(e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0-1)(\psi(e_t') - \ln(f_t')) - f_0\frac{e_t'}{f_t'}\right)$$
$$+ \left(\frac{1}{2}\ln|\Sigma_t'| + \frac{d}{2}(1+\ln(2\pi))\right) - \sum_{k=1}^{d}\left[a_{tk}'\ln b_{tk}' - \ln\Gamma(a_{tk}') + (a_{tk}'-1)(\psi(a_{tk}') - \ln(b_{tk}')) - a_{tk}'\right]$$
$$- \left[e_t' \ln f_t' - \ln\Gamma(e_t') + (e_t'-1)(\psi(e_t') - \ln(f_t')) - e_t'\right]$$

EECS 6892 BMML HW#3 UNI: sl3763 Name: Sung-Yen Liu

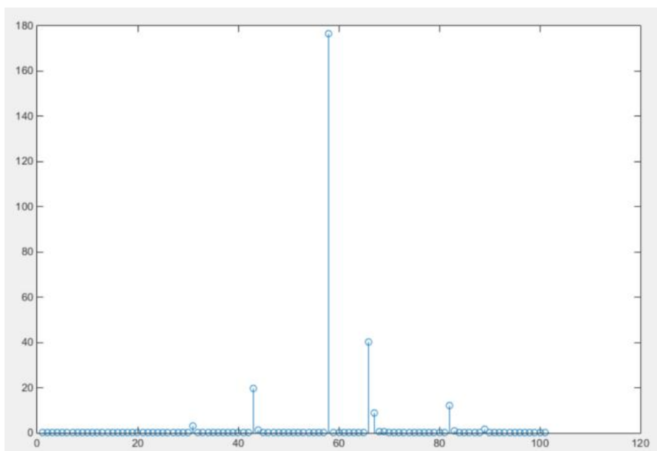
Problem 2.

data1.mat:

a) Plot of variational objective function



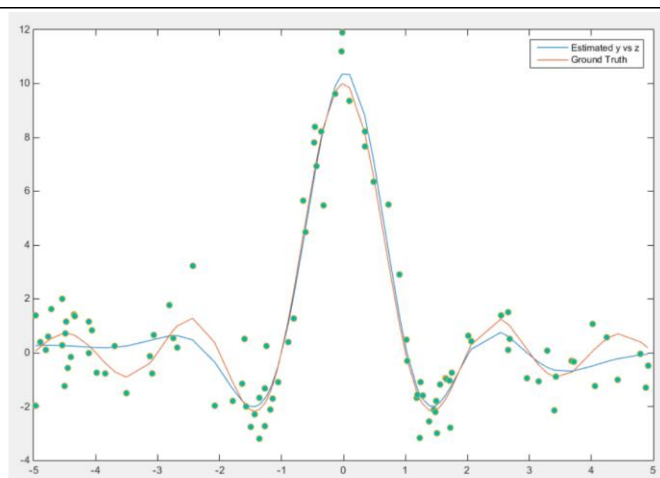
b) Plot of $1/E_q[\alpha_k]$



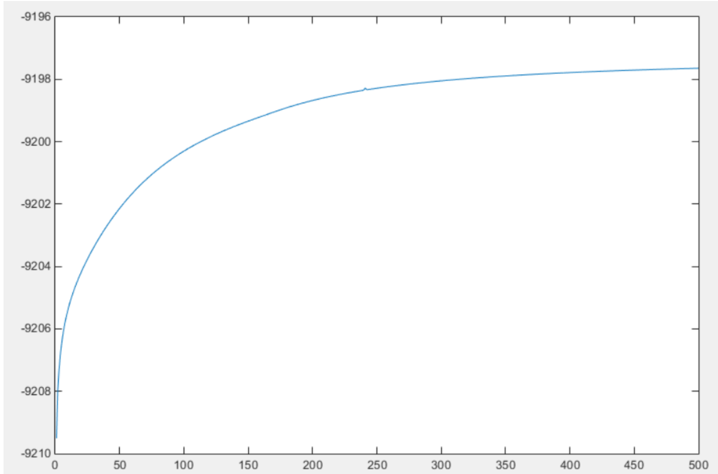
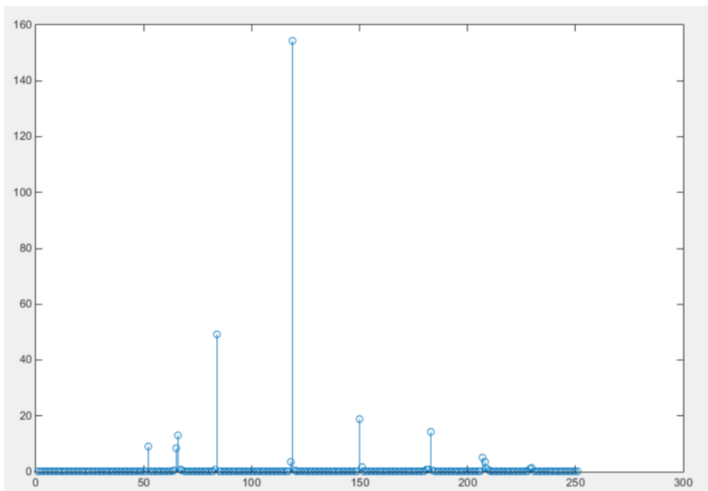
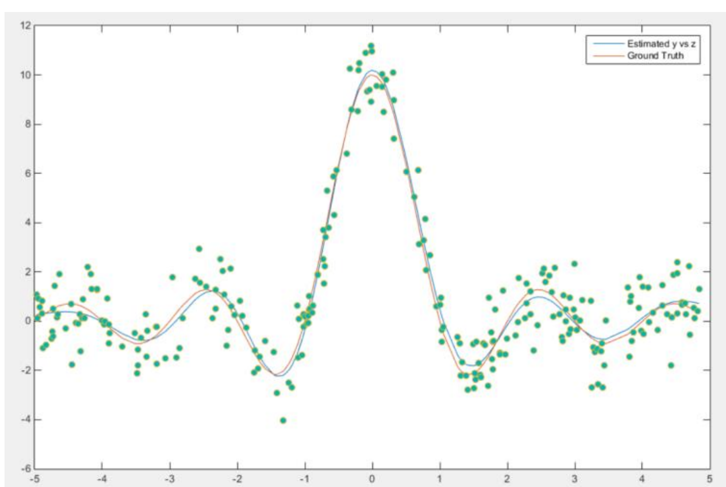
c) $1/E_q[\lambda]$

1.0798

d) Estimated y vs. z, Ground Truth and Scatter Plot

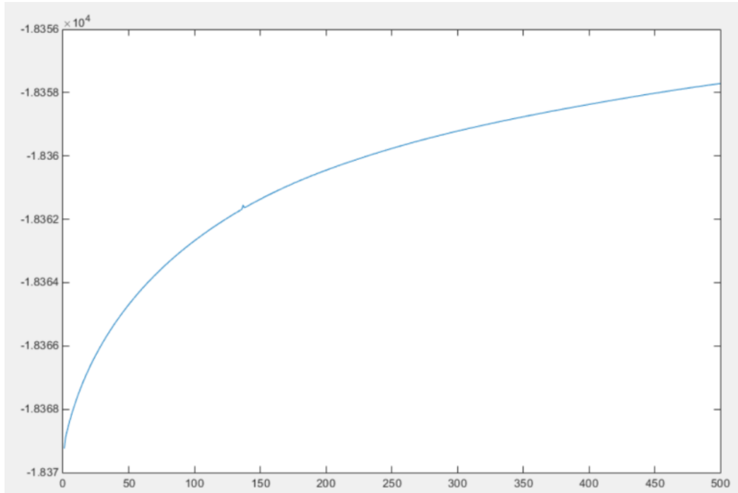


data2.mat

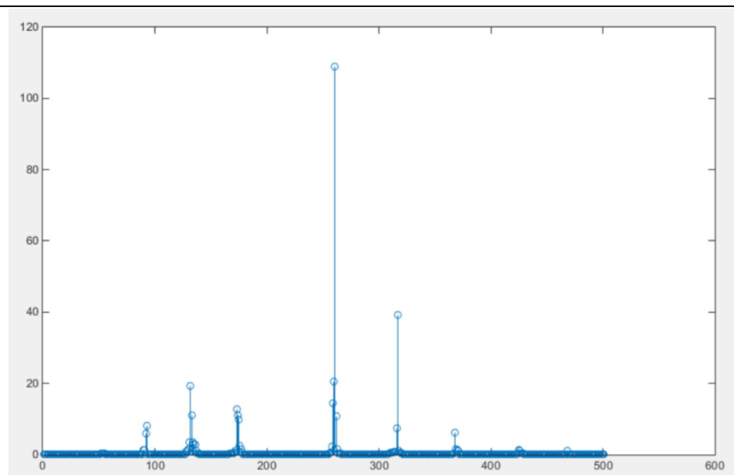
a) Plot of variational objective function

b) Plot of $1/E_q[\alpha_k]$

c) $1/E_q[\lambda]$
0.8994
d) Estimated y vs. z, Ground Truth and Scatter Plot


data3.mat

a) Plot of variational objective function



b) Plot of $1/E_q[\alpha_k]$



c) $1/E_q[\lambda]$

0.9781

d) Estimated y vs. z, Ground Truth and Scatter Plot