Homework 4 (HW4 Part B)

Oussama Ennafii & Sammy Khalife

25/02/2015

Step 1

1. The equality is trivial: we just have to set $\vartheta = \theta$.

To prove the inequality

$$Q_\gamma(\vartheta, \theta) \geq (f + g)(\vartheta)$$

we just need to prove that:

$$\forall \vartheta, \theta \in \Theta \quad f(\vartheta) \leq f(\theta) + < \nabla f(\theta), \vartheta - \theta > + \frac{1}{2\gamma} ||\vartheta - \theta||^2$$

Set $\vartheta, \theta \in \Theta$. Let's take $t \in [0, 1]$, we define:

$$F(t) \triangleq f(\vartheta_t) - f(\theta) - < \nabla f(\theta), \vartheta_t - \theta > - \frac{1}{2\gamma} ||\vartheta_t - \theta||^2 \tag{1}$$

where $\vartheta_t = t\vartheta + (1 - t)\theta$. Since, $\frac{d}{dt}\vartheta_t = \vartheta - \theta$, F is differentiable and:

$$\begin{aligned}
\frac{d}{dt}F(t) &= < \nabla f(\vartheta), \vartheta - \theta > - < \nabla f(\theta), \vartheta - \theta > - \frac{1}{\gamma} ||\vartheta - \theta||^2 \\
&= < \nabla f(\vartheta) - \nabla f(\theta), \vartheta - \theta > - \frac{1}{\gamma} ||\vartheta - \theta||^2 \\
&\leq ||\nabla f(\vartheta) - \nabla f(\theta)||.||\vartheta - \theta|| - ||\vartheta - \theta||^2 \\
&\leq (L - \frac{1}{\gamma}) ||\vartheta - \theta||^2 \\
&\leq 0
\end{aligned}$$

The first inequality is the result of Cauchy-Schwartz inequality while the second one is due to the Lipschitz nature of the gradient. The third inequality comes from the fact that $0 < \gamma \leq \frac{1}{L}$.
This means that F in decreasing in $[0, 1]$, so: $F(0) \geq F(1)$:

$$F(0) = 0$$

and :

$$F(1) = f(\vartheta) - f(\theta) - < \nabla f(\theta), \vartheta - \theta > - \frac{1}{2\gamma} ||\vartheta - \theta||^2$$

2. Let $n \in \mathbb{N}$:

$$\begin{aligned}(f+g)(\theta_n) &= Q_{\gamma_{n+1}}(\theta_n,\theta_n)\\ &\geq Q_{\gamma_{n+1}}(\theta_{n+1},\theta_n)\\ &\geq (f+g)(\theta_{n+1})\end{aligned}$$

The first inequality is the result of the definition of θ_{n+1}. The second one comes from the first question.

3. Set $\vartheta,\theta\in\Theta$. we develop:

$$\frac{1}{2\gamma}||\vartheta-\theta-\gamma\nabla f(\theta)||^2 = \frac{1}{2\gamma}||\vartheta-\theta||^2 - <\vartheta-\theta,\nabla f(\theta)> + \frac{\gamma}{2}||\nabla f(\theta)||^2$$

This means that :

$$Q_\gamma(\vartheta,\theta) = g(\vartheta)+f(\theta)+\frac{1}{2\gamma}||\vartheta-\theta-\gamma\nabla f(\theta)||^2-\frac{\gamma}{2}||\nabla f(\theta)||^2$$

4. Taking off the constants w.r.t to ϑ, we get:

$$\begin{aligned}\theta_{n+1} &= \arg\min_{\vartheta\in\Theta} Q_{\gamma_{n+1}}(\vartheta,\theta_n)\\ &= \arg\min_{\vartheta\in\Theta} g(\vartheta)+\frac{1}{2\gamma_{n+1}}||\vartheta-\theta_n-\gamma_{n+1}\nabla f(\theta_n)||^2\\ &= \arg\min_{\vartheta\in\Theta}\sum_{i=1,\dots,d}\lambda|\vartheta_i|+\frac{1}{2\gamma_{n+1}}(\vartheta_i-(\theta_n)_i-\gamma_{n+1}(\nabla f(\theta_n))_i)^2\end{aligned}$$

So we can just minimize on each direction independently:

$$(\theta_{n+1})_i = \arg\min_{\vartheta_i}\lambda|\vartheta_i|+\frac{1}{2\gamma_{n+1}}(\vartheta_i-(\theta_n)_i-\gamma_{n+1}(\nabla f(\theta_n))_i)^2$$

We set : $u=\theta_n-\gamma_{n+1}\nabla f(\theta_n)$. We rewrite the last equation as:

$$(\theta_{n+1})_i = \arg\min_{\vartheta_i}\lambda|\vartheta_i|+\frac{1}{2\gamma_{n+1}}(\vartheta_i-u_i)^2$$

So we consider the function:

$$L_u^\gamma(x)\triangleq\lambda|x|+\frac{1}{2\gamma}(x-u)^2$$

It is obvious that :

$$L_u^\gamma(x)=L_{-u}^\gamma(x)$$

So L is symmetric w.r.t. u. It means that we can just focus on the case where: $u\in\mathbb{R}_+$:

- L_u^γ is decreasing in $\mathbb{R}_-$ from $+\infty$ to $\frac{1}{2\gamma}u^2$, because:

$$\forall x\in\mathbb{R}_-\quad L_u^\gamma(x)=-\lambda x+\frac{1}{2\gamma}(x-u)^2$$

- Since:
$$\forall x \in \mathbb{R}_+ \quad \frac{d}{dx} L_u^\gamma(x) = \lambda + \frac{1}{\gamma}(x - u)$$
Two cases present themselves:
  (i). Either $u - \gamma\lambda \geq 0$ and then L_u^γ decreases in $[0, u - \gamma\lambda]$ from $\frac{1}{2\gamma}u^2$ to $\lambda u - \frac{\lambda^2\gamma}{2}$ and then starts increasing in $[u - \gamma\lambda, +\infty)$ from $\lambda u - \frac{\lambda^2\gamma}{2}$ to $+\infty$.
  (ii). Or $u - \gamma\lambda < 0$, which means that L_u^γ increases in $\mathbb{R}_+$ from $\frac{1}{2\gamma}u^2$ to $+\infty$

We make use of the symmetry of L w.r.t. u and we summuries the previous discussion:

$$\arg\min L_u^\gamma(x) = \begin{cases} u - \gamma\lambda & , \text{if } u \geq \gamma\lambda \\ 0 & , \text{if } u \in (-\lambda\gamma, \gamma\lambda) \\ u + \gamma\lambda & , \text{if} u \leq -\gamma\lambda \end{cases}$$

Now we can conclude that :

$$\theta_{n+1} = P_{\gamma_{n+1}}(\theta_n - \gamma_{n+1}\nabla f(\theta_n))$$

Step 2

1. For $i = 1, \ldots, N$, we introduce: $\tilde{u}_i \triangleq x_i'\beta + \sigma z_i' U$

$$\begin{aligned} \mathbb{P}(Y_i = y_i|\mathbf{U}) &= s(\tilde{u}_i)^{y_i}(1 - s(\tilde{u}_i))^{1-y_i} \\ &= \frac{e^{\tilde{u}_i y_i}}{(1 + e^{\tilde{u}_i})^{y_i}} \times \frac{1}{(1 + e^{\tilde{u}_i})^{1-y_i}} \\ &= \frac{e^{\tilde{u}_i y_i}}{(1 + e^{\tilde{u}_i})} \end{aligned}$$

This means that:

$$\begin{aligned} \mathbb{P}((Y_1, \ldots, Y_N) = (y_1, \ldots, y_N)|\mathbf{U}) &= \prod_{i=1,\ldots,N} \frac{e^{\tilde{u}_i y_i}}{(1 + e^{\tilde{u}_i})} \\ &= \frac{\exp(\sum_{i=1,\ldots,N} \tilde{u}_i y_i)}{\prod_{i=1,\ldots,N}(1 + e^{\tilde{u}_i})} \end{aligned}$$

$$\begin{aligned} \mathbb{P}((Y_1, \ldots, Y_N) = (y_1, \ldots, y_N)) &= \int_{\mathbb{R}} \mathbb{P}((Y_1, \ldots, Y_N) = (y_1, \ldots, y_N)|\mathbf{U} = u)\phi(u)du \\ &= \int_{\mathbb{R}} \frac{\exp(\sum_{i=1,\ldots,N} \tilde{u}_i y_i)}{\prod_{i=1,\ldots,N}(1 + e^{\tilde{u}_i})}\phi(u)du \end{aligned}$$

Now we can write that:

$$l(\theta|Y_1,\dots,Y_N) = \log(\int_{\mathbb{R}} \frac{\exp(\sum_{i=1,\dots,N}(x_i'\beta+\sigma z_i'u)Y_i)}{\prod_{i=1,\dots,N}(1+\exp(x_i'\beta+\sigma z_i'u))}\phi(u)du)$$

2. We can derive inside the integral using the bounded convergence since the function is continous:

$$\begin{aligned}
\nabla l(\theta) &= \frac{1}{\int_{\mathbb{R}} \frac{\exp(\sum_{i=1,\dots,N}(x_i'\beta+\sigma z_i'u)Y_i)}{\prod_{i=1,\dots,N}(1+\exp(x_i'\beta+\sigma z_i'u))}\phi(u)du} \times \nabla \int_{\mathbb{R}} \frac{\exp(\sum_{i=1,\dots,N}(x_i'\beta+\sigma z_i'u)Y_i)}{\prod_{i=1,\dots,N}(1+\exp(x_i'\beta+\sigma z_i'u))}\phi(u)du) \\
&= \frac{1}{\exp(l(\theta))} \times \nabla \int_{\mathbb{R}} \exp(\sum_{i=1,\dots,N}(x_i'\beta+\sigma z_i'u)Y_i - \log(1+\exp(x_i'\beta+\sigma z_i'u)))\phi(u)du \\
&= \exp(-l(\theta)) \int_{\mathbb{R}} \nabla \exp(l_c(\theta|\mathbf{u}))\phi(u)du \\
&= \int_{\mathbb{R}} \nabla l_c(\theta|\mathbf{u}) \exp(l_c(\theta|\mathbf{u}) - l(\theta))\phi(u)du \\
&= \int_{\mathbb{R}} \sum_{i=1}^{N}(Y_i \begin{bmatrix} x_i \\ z_i'u \end{bmatrix} - \frac{\exp(x_i'\beta+\sigma z_i'u)}{1+\exp(x_i'\beta+\sigma z_i'u)} \begin{bmatrix} x_i \\ z_i'u \end{bmatrix}) \exp(l_c(\theta|\mathbf{u}) - l(\theta))\phi(u)du
\end{aligned}$$

Step 3

1.

$$
\begin{aligned}
\tilde{\pi}_\theta(u,w) &= \pi_\theta(u)(\prod_{i=1}^{N}\bar{\pi}(w_i;x_i'\beta+\sigma z_i'u)) \\
&= \frac{1}{exp(l(\theta))}\phi(u)exp(l_c(\theta|u))(\prod_{i=1}^{N}\bar{\pi}(w_i;x_i'\beta+\sigma z_i'u)) \\
&= \frac{1}{exp(l(\theta))}\phi(u)\prod_{i=1}^{N}Zcosh(\frac{x_i'\beta+\sigma z_i'u}{2})\rho(w_i)\mathbf{1}_{\mathbb{R}^+}(w_i)exp(-w(x_i'\beta+\sigma z_i'u)^2/2) \\
&\qquad \frac{exp(Y_i(x_i'\beta+\sigma z_i'u))}{1+exp(x_i'\beta+\sigma z_i'u)}
\end{aligned}
$$

Writing $cosh(x) = exp(-x/2)(1+exp(x))\frac{1}{2}$, we obtain :

$$
\begin{aligned}
\tilde{\pi}_\theta(u,w) &= \frac{1}{exp(l(\theta))}\phi(u)\prod_{i=1}^{N}Z\rho(w_i)\mathbf{1}_{\mathbb{R}^+}(w_i)exp(Y_i(x_i'\beta+\sigma z_i'u)-(x_i'\beta+\sigma z_i'u)/2-w_i(x_i'\beta+\sigma z_i'u)^2/2) \\
&= \frac{1}{exp(l(\theta))}\phi(u)Z^N\prod_{i=1}^{N}exp(Y_ix_i'\beta-x_i'\beta/2) \\
&\qquad \prod_{i=1}^{N}\rho(w_i)\mathbf{1}_{\mathbb{R}^+}(w_i)exp(\sigma(Y_i-1/2)z_i'u-w_i(x_i'\beta+\sigma z_i'u)^2/2) \\
&= C(\theta)\phi(u)\prod_{i=1}^{N}\rho(w_i)\mathbf{1}_{\mathbb{R}^+}(w_i)exp(\sigma(Y_i-1/2)z_i'u-w_i(x_i'\beta+\sigma z_i'u)^2/2)
\end{aligned}
$$

With $\boxed{C(\theta) = \frac{1}{exp(l(\theta))}\phi(u)Z^N\prod_{i=1}^{N}exp(Y_ix_i'\beta-x_i'\beta/2)}$,

$$
\boxed{\tilde{\pi}_\theta(u,w) = C(\theta)\phi(u)\prod_{i=1}^{N}\rho(w_i)\mathbf{1}_{\mathbb{R}^+}(w_i)exp(\sigma(Y_i-1/2)z_i'u-w_i(x_i'\beta+\sigma z_i'u)^2/2)}
$$

2. Let us factorize the expression of $\tilde{\pi}_\theta(u,w)$ under the form :

$$
\begin{aligned}
\tilde{\pi}_\theta(u,w) &= h(\theta,w,\beta,\sigma,...)exp(-\frac{1}{2}u^tu-\sum_{i=1}^{N}\frac{w_i}{2}\sigma^2(z_i'u)^2+\sum_{i=1}^{N}\sigma(Y_i-\frac{1}{2})z_i'u-w_ix_i'\beta\sigma z_i'u) \\
&= h(\theta,w,\beta,\sigma,...)exp(-\frac{1}{2}u^t(I+\sigma^2\sum_{i=1}^{n}\frac{w_i}{2}z_iz_i')u+\sigma<u,\sum_{i=1}^{N}((Y_i-\frac{1}{2})-w_ix_i'\beta)z_i>)
\end{aligned}
$$

Where h does not depend on u. We can directly identify the density of a Gaussian :

$$\tilde{\pi}_\theta(u|w) \quad \propto \quad exp(-\frac{1}{2}u^t(I+\sigma^2\sum_{i=1}^{n}\frac{w_i}{2}z_iz_i')u + \sigma < u, \sum_{i=1}^{N}((Y_i-\frac{1}{2})-w_ix_i'\beta)z_i >)$$

i.e (we can identify since the Gaussian is normalized)

$$\boxed{\tilde{\pi}_\theta(u|w) \quad = \mathcal{N}(\mu_\theta, \Gamma_\theta(w))}$$

With $\boxed{\Gamma_\theta(w) = (I+\sigma^2\sum_{i=1}^{n}\frac{w_i}{2}z_iz_i')^{-1}}$, and $\boxed{\mu_\theta(w) = \sigma\Gamma_\theta(w)\sum_{i=1}^{N}((Y_i-\frac{1}{2})-w_ix_i'\beta)z_i}$.

3. We use the same reasoning as the previous question.

$$\begin{aligned}
\tilde{\pi}_\theta(w|u) \quad & \underset{w}{\propto} \quad \prod_{i=1}^{N}\rho(w_i)\mathbf{1}_{R^+}(w_i)exp(-\frac{w_i}{2}(x_i'\beta+\sigma z_i'u)^2) \\
& \underset{w}{\propto} \quad \prod_{i=1}^{N}Zcosh(\frac{x_i'\beta+\sigma z_i'u}{2})\mathbf{1}_{R^+}(w_i)exp(-\frac{w_i}{2}(x_i'\beta+\sigma z_i'u)^2)
\end{aligned}$$

Since $cosh(x) = cosh(|x|)$,

$$\begin{aligned}
\tilde{\pi}_\theta(w|u) \quad & \underset{w}{\propto} \quad \prod_{i=1}^{N}Zcosh(\frac{|x_i'\beta+\sigma z_i'u|}{2})\mathbf{1}_{R^+}(w_i)exp(-\frac{w_i}{2}(x_i'\beta+\sigma z_i'u)^2) \\
& \underset{w}{\propto} \quad \prod_{i=1}^{N}\bar{\pi}(w_i;|x_i'\beta+\sigma z_i'u|)
\end{aligned}$$

Since the last line a function normalized in w, i.e

$$\begin{aligned}
\int_\Omega\prod_{i=1}^{N}\bar{\pi}(w_i;|x_i'\beta+\sigma z_i'u|)dw \quad & = \quad \prod_{i=1}^{N}\int_{\Omega_i}\bar{\pi}(w_i;|x_i'\beta+\sigma z_i'u|)dw_i \\
& = \quad 1
\end{aligned}$$

We have

$$\boxed{\tilde{\pi}_\theta(w|u) = \prod_{i=1}^{N}\bar{\pi}(w_i;|x_i'\beta+\sigma z_i'u|)}$$

4.

5. To sample $(w_1, ..., w_N, u)$ from $\tilde{\pi}_\theta(u, w)$, we will use a Gibbs Sampler

Algorithm 1 Gibbs Sampler to sample from $\tilde{\pi}_\theta$

Given, $w_1^{(0)}, ..., w_N^{(0)}$

Loop on k
Draw $u^{(k+1)}$ from $\pi_\theta(.|w^{(k)}) = \mathcal{N}(\mu_\theta, \Gamma_\theta(\omega))$
-for i=1:N
$w_i^{(k+1)}$ drawn thanks to HW1Sampler$(.,|x_i'\beta + \sigma z_i' u^{(k+1)}|/2)$
-end for i
end k
return $(u^{(k)}, w^{(k)})_{1\leq k\leq K}$

Cf code for implementation.

6. One can use a sample generated by the previous algorithm to approximate the integral of $\nabla l(\theta)$. Let us denote $H_\theta(u) = \sum_{i=1}^N (Y_i - s(x_i'\beta + \sigma z_i' u)) \binom{x_i}{z_i' u}$. Then,

$$\nabla l(\theta) = \int\int H_\theta(u)\tilde{\pi}_\theta(u,w)dudw \sim \frac{1}{K}\sum_{k=1}^K H_\theta(u^k)$$

Algorithm 2 Algorithm to approximate $\nabla l(\theta)$

Sample a chain $(u^{(k)}, w^{(k)})_{1\leq k\leq K}$ from distribution $\tilde{\pi}_\theta$ with algorithm 1.

return $\frac{1}{K}\sum_{k=1}^K H_\theta(u^k)$

Cf code for implementation.

Step 4

1. c.f. the code directory $\rightarrow' \, dataset\_generator.m'$.