Q.1.

a)

$$\frac{f(x)}{g(x)} = \frac{\frac{\exp(-x^2/2)}{\sqrt{2\pi}(1-\Phi(b))}}{b\exp(-b(x-b))} = \frac{1}{\sqrt{2\pi}(1-\Phi(b))\cdot b} e^{-x^2/2+bx-b^2}$$

$$= \frac{e^{-\frac{x^2}{2}+bx-b^2}}{\sqrt{2\pi}(1-\Phi(b))b} = p(b)\cdot e^{-\frac{x^2}{2}+bx-b^2}$$

where $p(b)$ should be constant and independent of $x$.)

$M = \max_x \frac{f(x)}{g(x)}$, we want to max $p(b)\cdot e^{-\frac{x^2}{2}+bx}$

$\Rightarrow \max_x e^{-\frac{x^2}{2}+bx} \Rightarrow \max_x -\frac{x^2}{2}+bx$

take derivative $-x+b=0$, $b=x$ ( $\frac{f(x)}{g(x)}$ has greatest value )

Then, we have $M = p(b)\cdot e^{-\frac{b^2}{2}+b^2-b^2} = p(b)\cdot e^{-b^2/2}$

$U$ should satisfy $U \leq \frac{f(x)}{Mg(x)}$

$$U \leq \cancel{p(b)}\cdot e^{-\frac{x^2}{2}+bx-b^2}\cdot\frac{1}{\cancel{p(b)}\, e^{-b^2/2}}$$

$$U \leq e^{-\frac{x^2}{2}+bx-b^2+\frac{b^2}{2}}$$

$$U \leq e^{-\frac{x^2}{2}+bx-\frac{b^2}{2}}$$

$$U \leq e^{-\frac{1}{2}(x-b)^2}$$

Then we accept $X = b + \frac{Y}{b}$ if

$$U \leq e^{-\frac{1}{2}(b+\frac{Y}{b}-b)^2}$$

$$U \leq e^{-\frac{1}{2}\cdot(\frac{Y}{b})^2}$$

$$\ln U \leq -\frac{1}{2}\left(\frac{Y}{b}\right)^2$$

$$-2\ln U \geq \frac{Y^2}{b^2}$$

b).

probability of acceptance: $\frac{1}{M} = \frac{\sqrt{2\pi}\, b\,(1-\Phi(b))}{\exp(-b^2/2)}$

$$\lim_{b\to\infty} \frac{\sqrt{2\pi}\, b\,(1-\Phi(b))}{\exp(-b^2/2)} = \lim_{b\to\infty} \frac{\frac{d}{db}\left(\sqrt{2\pi}\, b\,(1-\Phi(b))\right)}{\frac{d}{db}\exp(-b^2/2)}$$ by L'Hoptal Rule

$$\Rightarrow \lim_{b\to\infty} \frac{\sqrt{2\pi}\,(1-\Phi(b)) - \sqrt{2\pi}\, b\,\phi(b)}{-b e^{-\frac{1}{2}b^2}}$$

[note: $\Phi$ is CDF of $N(0,1)$
$\phi$ is pdf of $N(0,1)$ ]

by Mills's ratio

$$1-\Phi(b) \approx \frac{1}{b}\phi(b) \quad \& \quad \phi(b) = \frac{1}{\sqrt{2\pi}} e^{-\frac{b^2}{2}}$$

$$\lim_{b\to\infty} \frac{\sqrt{2\pi}\,\frac{1}{b}\phi(b) - \sqrt{2\pi}\, b\,\phi(b)}{-b e^{-\frac{1}{2}b^2}}$$

$$= \lim_{b\to\infty} \frac{\frac{1}{b} e^{-b^2/2} - b e^{-b^2/2}}{-b e^{-\frac{1}{2}b^2}}$$

$$= \lim_{b\to\infty} \frac{1}{b^2} + 1 = 1$$

The probability equals 1.

c) $\min_{\lambda>0} \max_{x\geq b} \frac{f(x)}{g_\lambda(x)} = \max_{x\geq b} \left( \min_{\lambda>0} \frac{f(x)}{g_\lambda(x)} \right)$

$\max_{x\geq b} \min_{\lambda>0} \ln f(x) - \ln g_\lambda(x) = \ln f(x) - \ln(\lambda \exp(-\lambda(x-b)))$

$= \ln f(x) - \ln\lambda + \lambda x - \lambda b$

take derivative to find min.

$\frac{d}{d\lambda} \ln f(x) - \ln\lambda + \lambda x - \lambda b = 0$

$-\frac{1}{\lambda} + x - b = 0$

$-\frac{1}{\lambda} = b + x$

$\lambda = (x-b)^{-1}$

Plug $\lambda = (x-b)^{-1}$ into function

$\ln f(x) - \ln[(x-b)^{-1} \exp(x-b)^{-1}(x-b))]$

$= \ln \frac{\exp(-x^2/2)}{\sqrt{2\pi}(1-\Phi(b))} - \ln(x-b)^{-1} + 1$

$= -\frac{x^2}{2} - \ln[\sqrt{2\pi}(1-\Phi(b))] - \ln(x-b)^{-1} + 1$

take derivative to find maximum.

$$\frac{d}{dx}\left[-\frac{x^2}{2} - \ln[\sqrt{2\pi}(1-\Phi(b))] - \ln(x-b)^{-1} + 1\right] = 0$$

$$\frac{d}{dx}\left[-\frac{x^2}{2} - \ln[\sqrt{2\pi}(1-\Phi(b))] + \ln(x-b) + 1\right] = 0$$

since $\ln(x-b)^{-1} = -\ln(x-b)$

$$-x + \frac{1}{x-b} = 0$$

$$x^2 - bx - 1 = 0$$

$$x = \frac{b+\sqrt{b^2+4}}{2} \quad (x \geq b,\ b \geq 0)$$

$$\lambda = \lambda(b) = (x-b)^{-1} = \left(\frac{b+\sqrt{b^2+4}}{2} - b\right)^{-1}$$

$$= \left(\frac{-b+\sqrt{b^2+4}}{2}\right)^{-1}$$

$$= \frac{2}{-b+\sqrt{b^2+4}}$$

Thus, $\lambda = \dfrac{2}{-b+\sqrt{b^2+4}}$ maximizes the probability of acceptance

Q2.

1) We have $y_i = a \times i + b$, assume $\theta_i = y_i$,

$$\theta_{i+1} - 2\theta_i + \theta_{i-1} = y_{i+1} - 2y_i + y_{i-1}$$

$$= a(i+1) + b - 2(ai+b) + a(i-1) + b$$

$$= ai + a + b - 2ai - 2b + ai - a + b$$

$$= 0$$

Thus, when $\hat{\theta}_i = y_i$

$$\sum_{i=1}^{n} (y_i - \theta_i)^2 + \lambda \sum_{i=2}^{n-1} (\theta_{i+1} - 2\theta_i + \theta_{i-1})^2 = \sum_{i=1}^{n} 0 + \lambda \sum_{i=2}^{n-1} 0 = 0$$

which is minimized.

2b) $\hat{\theta} = y_i$ for all $i$ minimize $\|y^* - x\theta\|^2$

$$\sum_{i=1}^{n} (y_i - \theta_i)^2 + \lambda \sum_{i=2}^{n-1} (\theta_{i+1} - 2\theta_i + \theta_{i-1})^2 = 0$$

$$\left\| \begin{matrix} y_1 - \theta_1 \\ \vdots \\ y_n - \theta_n \\ \sqrt{\lambda}(\theta_3 - 2\theta_2 + \theta_1) \\ \vdots \\ \sqrt{\lambda}(\theta_n - 2\theta_{n-1} + \theta_{n-2}) \end{matrix} \right\|^2 \Rightarrow y^* = \begin{pmatrix} y_1 \\ \vdots \\ y_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{(2n-2)\times 1}$$

$$X = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & 1 \\ \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} & 0 \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & \cdots & \cdots & \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} \end{bmatrix}_{(2n-2)\times n}$$

$$\therefore y^* - X\theta = \begin{pmatrix} y_1 \\ \vdots \\ y_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & 1 \\ \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} & 0 \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & \cdots & \cdots & \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} \end{bmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}_{n\times 1}$$

c).

At ith iteration, the minimize function we have is.

$$\|y^* - X_{\bar{w}} \hat{\theta}_{\bar{w}(i)} - X_w \theta_{w(i)}\|$$

And for i+1 iteration, we want to minimize

$$\| y^* - X_{\bar{w}} \hat{\theta}_{\bar{w}(i)} - X_w \theta_w \|$$

then we have

$$\|y^* - X_{\bar{w}} \hat{\theta}_{\bar{w}(i)} - X_w \hat{\theta}_{w(i+1)}\| \leq \|y^* - X_{\bar{w}} \hat{\theta}_{\bar{w}(i)} - X_w \hat{\theta}_{w(i)}\|$$

Thus, the objective function is non-increasing from one iteration to the next.

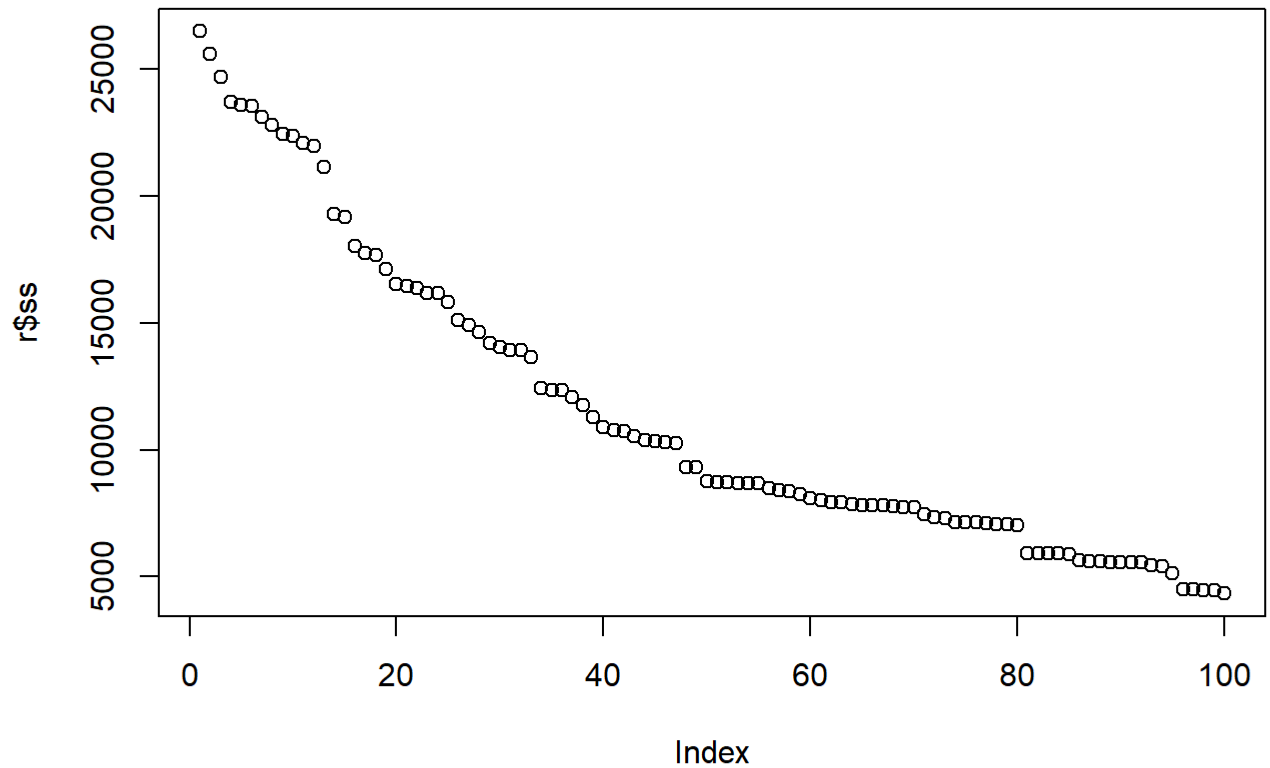
# STA410 A2

```
HP <- function(x,lambda,p=20,niter=200) {
        n <- length(x)
        a <- c(1,-2,1)
        aa <- c(a,rep(0,n-2))
        aaa <- c(rep(aa,n-3),a)
        mat <- matrix(aaa,ncol=n,byrow=T)
        mat <- rbind(diag(rep(1,n)),sqrt(lambda)*mat)
        xhat <- x
        x <- c(x,rep(0,n-2))
        sumofsquares <- NULL
        for (i in 1:niter) {
           w <- sort(sample(c(1:n),size=p))
           xx <- mat[,w]
           y <- x - mat[,-w]%*%xhat[-w]
           r <- lsfit(xx,y,intercept=F)
           xhat[w] <- r$coef
           sumofsquares <- c(sumofsquares,sum(r$residuals^2))
           }
        r <- list(xhat=xhat,ss=sumofsquares)
        r
        }
```

```
data<-scan("C:/Users/WLJY8/Desktop/Courses/YEAR 4/STA410/A2/yield.txt")
```

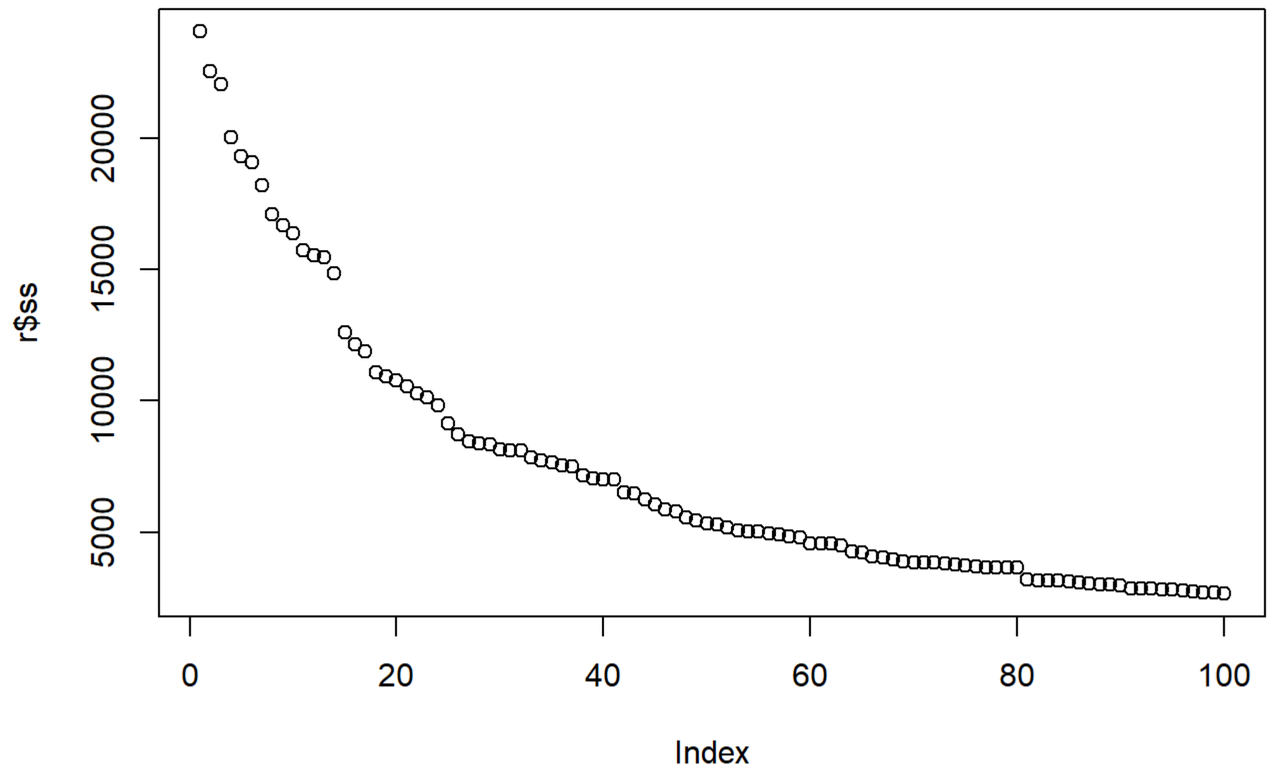
## p=5

```
r <- HP(data,lambda=2000,p=5,niter=100)
plot(r$ss)
```



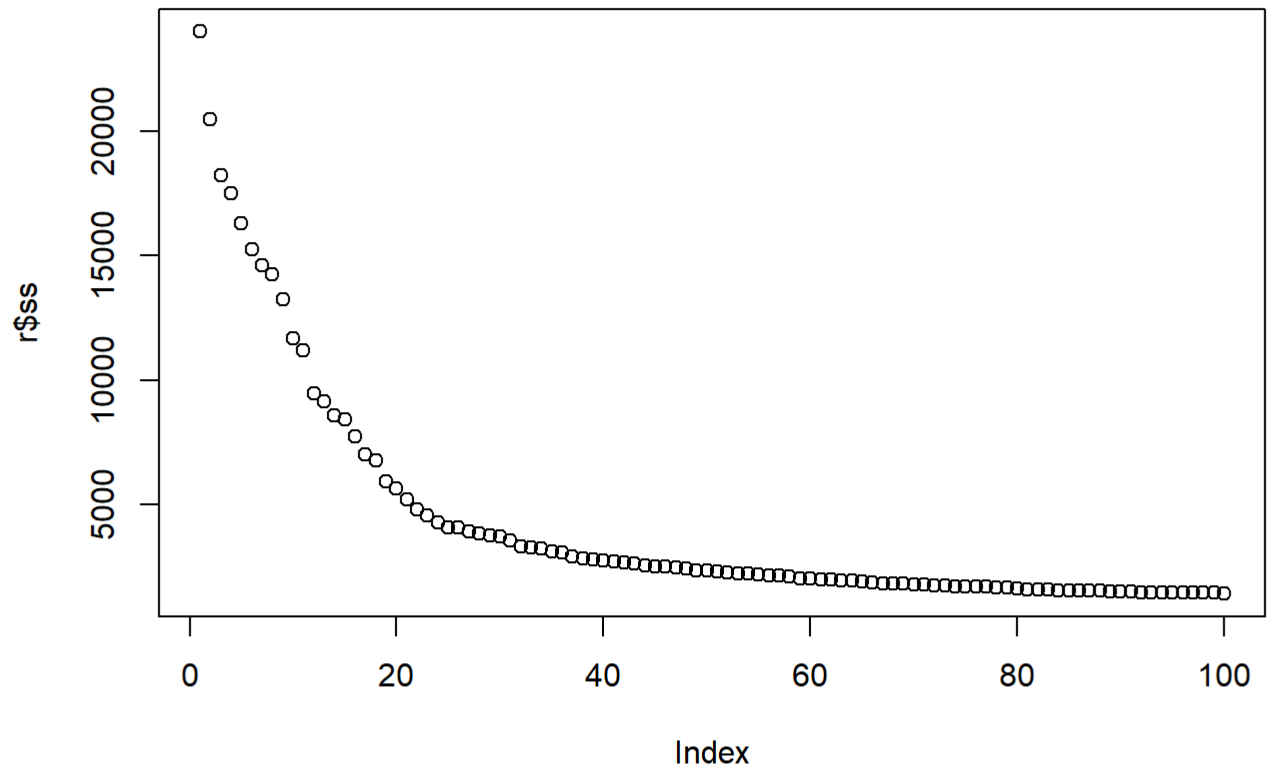

p=10

```
r <- HP(data,lambda=2000,p=10,niter=100)
plot(r$ss)
```



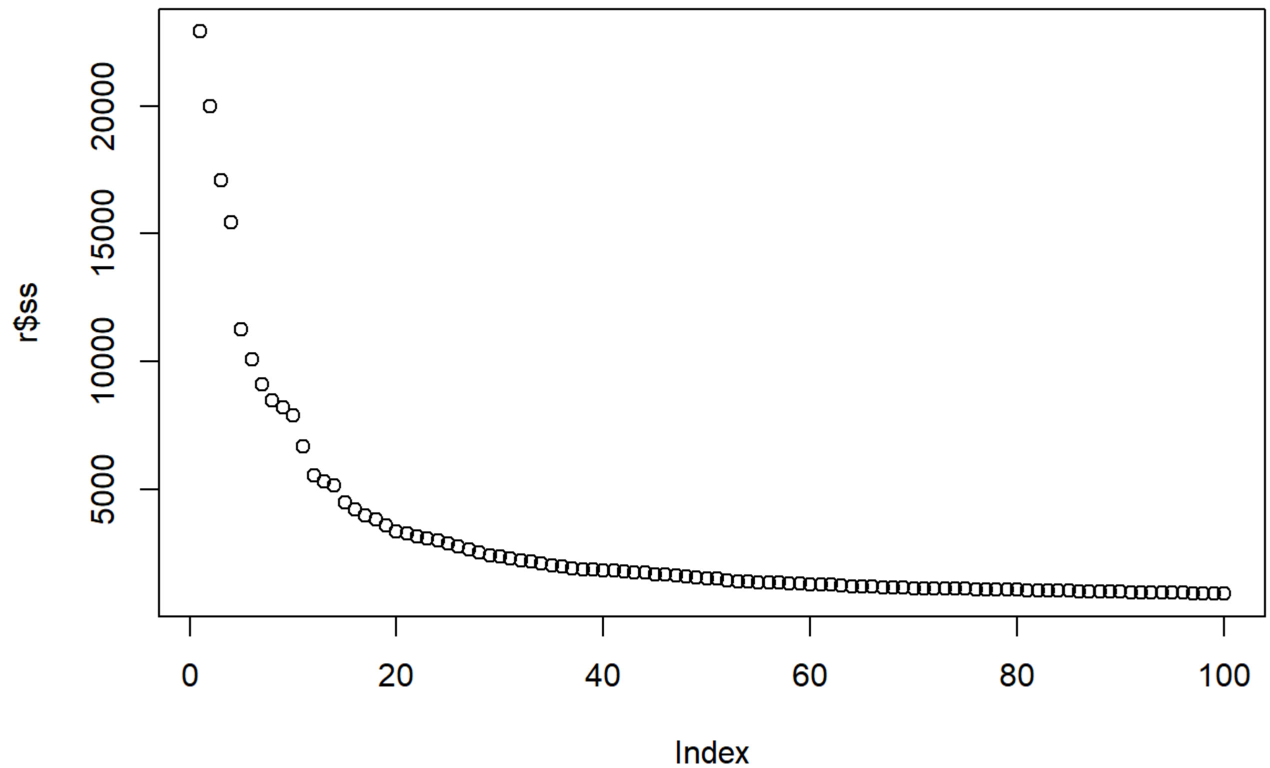

p=20

```
r <- HP(data,lambda=2000,p=20,niter=100)
plot(r$ss)
```



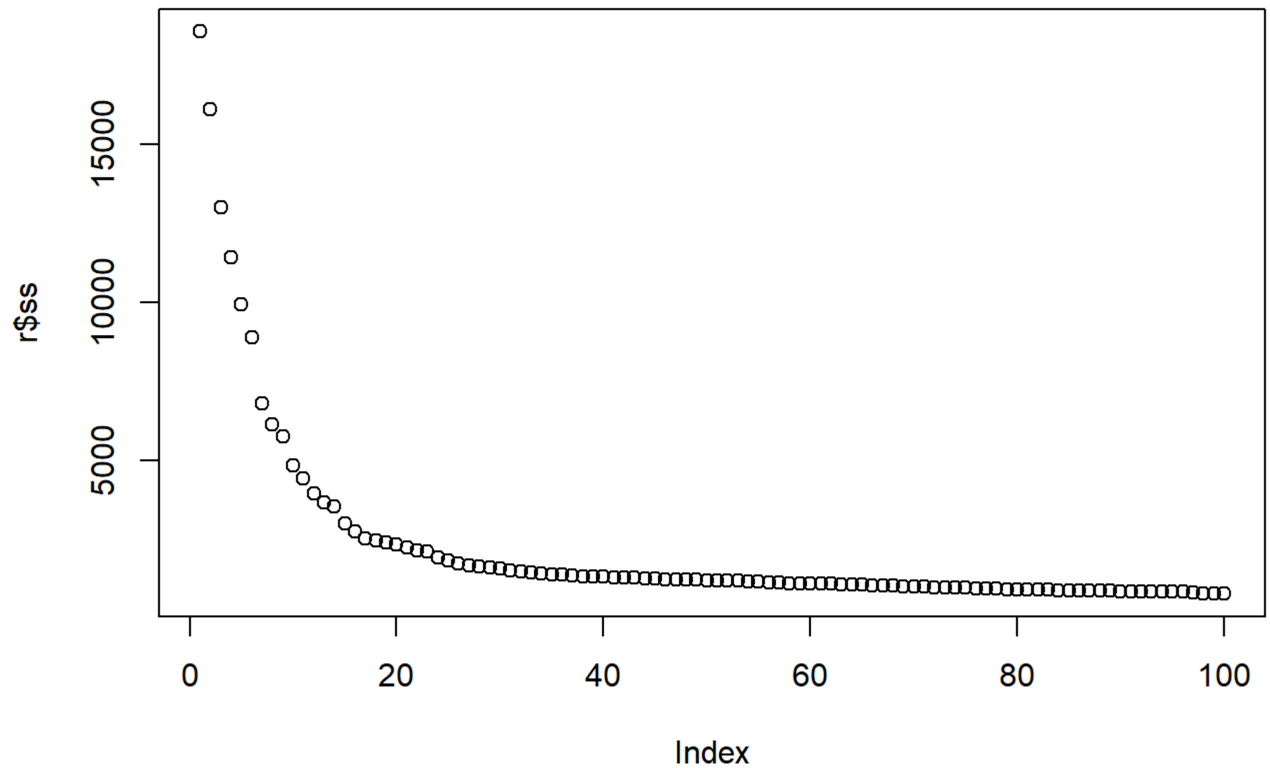

p=30

```
r <- HP(data,lambda=2000,p=30,niter=100)
plot(r$ss)
```



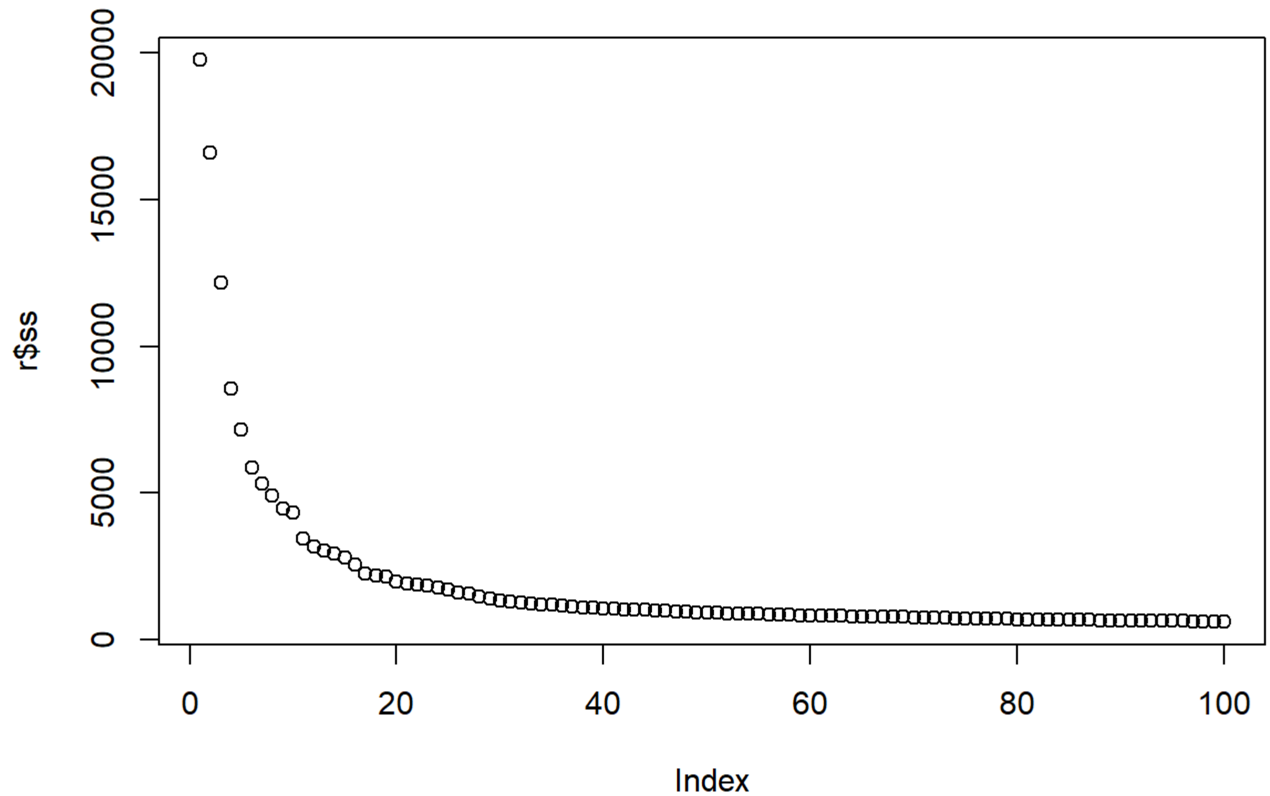

p=40

```
r <- HP(data,lambda=2000,p=40,niter=100)
plot(r$ss)
```




p=50

```
r <- HP(data,lambda=2000,p=50,niter=100)
plot(r$ss)
```




As p increases, the objective function value decreases more quickly as a function of the number of iterations.