STA2102 Homework 1

Kevin Zhang 1002225264

October 18, 2020

Question 1

a)

$$\frac{f(x)}{g(x)} = \frac{b}{\sqrt{2\pi}(1 - \Phi(b))} \exp\left[\frac{-\left(b + \frac{Y}{b}\right)^2}{2} + b\left(b + \frac{Y}{b} - b\right)\right]$$

$$= \frac{b}{\sqrt{2\pi}(1 - \Phi(b))} \exp\left[\frac{-b^2 - 2Y - \frac{Y^2}{b^2}}{2} + Y\right]$$

$$= \frac{b}{\sqrt{2\pi}(1 - \Phi(b))} \exp\left[\frac{-b^2}{2} - \frac{Y^2}{2b^2}\right]$$

$$\text{define } M = \frac{b}{\sqrt{2\pi}(1 - \Phi(b))} \exp\left[\frac{-b^2}{2}\right]$$

$$\text{then } \frac{f(x)}{Mg(x)} = \exp\frac{-Y^2}{2b^2} \leq 1 \text{ since } e^{-x} \leq \ \forall \, x \geq 0$$

Now we accept $X = b + \frac{Y}{b}$ if $U \leq \frac{f(X)}{Mg(X)} = \exp\frac{-Y^2}{2b^2} \implies -2\ln U \geq \frac{Y^2}{b^2}$.

b)

$$\frac{\partial}{\partial x}(\ln f(x) - \ln g(x)) = \frac{\partial}{\partial x}\left(\frac{-x^2}{2} - \ln(\sqrt{2\pi}(1 - \Phi(b))) - \ln(b) + b(x - b)\right)$$

$$= -x + b = 0$$

$$\implies x = b$$

$$\implies M = \frac{1}{\sqrt{2\pi}(1 - \Phi(b))} e^{\frac{-b^2}{2}} \times \frac{1}{b}$$

$$\implies P_{accept} = \frac{1}{M} = \frac{b\sqrt{2\pi}(1 - \Phi(b))}{e^{\frac{-b^2}{2}}}$$

c)

$$\frac{\partial}{\partial \lambda}(\ln f(x) - \ln g_\lambda(x)) = \frac{\partial}{\partial \lambda}\left(\frac{-x^2}{2} - \ln(\sqrt{2\pi}(1 - \Phi(b))) + \lambda(x - b) - \ln \lambda\right) = 0$$

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

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

$$\implies \max_{x \geq b}\min_{\lambda > 0}(\ln f(x) - \ln g_\lambda(x)) = \max_{x \geq b}\left(\frac{-x^2}{2} - \ln(\sqrt{2\pi}(1 - \Phi(b))) + \ln(x - b) + 1\right)$$

$$\implies \frac{\partial}{\partial x}\left(\frac{-x^2}{2} - \ln(\sqrt{2\pi}(1 - \Phi(b))) + \ln(x - b) + 1\right) = 0$$

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

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

$$\implies x = \frac{b + \sqrt{b^2 + 4}}{2} \text{ since } x \leq b \implies \lambda = \frac{1}{\frac{b+\sqrt{b^2+4}}{2} - b} = \frac{2}{\sqrt{b^2 + 4} - b}$$

Question 2

a)

Suppose there exists a unique minimizer for $f_\lambda(x) = \sum_{i=1}^{n}(y_i - \theta_i)^2 + \lambda\sum_{i=2}^{n-1}(\theta_{i+1} - 2\theta_i + \theta_{i-1})^2$ and notice that $f_\lambda(x) \geq 0$ almost surely. If $\hat{\theta}_i = y_i$, then $f_\lambda(\hat{\theta}) = 0 + \lambda\sum_{i=2}^{n-1}(a(i+1) + b - 2(ai + b) + a(i-1) + b)^2 = 0 + 0 = 0$. Therefore, $\hat{\theta} = y$ is the minimizer.

b)

$$Y^* = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & & & & & \\ 0 & \cdots & \cdots & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} & 0 & \cdots & 0 \\ \vdots & & & & & & & & \ddots & & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & \cdots & \sqrt{\lambda} & -2\sqrt{\lambda} & \sqrt{\lambda} \end{pmatrix}$$

c)

Let $\hat{\theta}^{(n)}$ and $\hat{\theta}^{(n+1)}$ denote the iterated values at nth and $(n+1)$th steps, respectively. Then $\hat{\theta}_{\bar{w}}^{(n)} = \hat{\theta}_{\bar{w}}^{(n+1)}$. Therefore,

$$||Y - X_{\bar{w}}\hat{\theta}_{\bar{w}}^{(n+1)} - X\hat{\theta}_w^{(n+1)}||^2 = ||Y - X_{\bar{w}}\hat{\theta}_{\bar{w}}^n - X\hat{\theta}_w^{(n+1)}||^2 \leq ||Y - X_{\bar{w}}\hat{\theta}_{\bar{w}}^n - X\hat{\theta}_w^n||^2$$

The last inequality holds because of the definition of $\hat{\theta}_w^{(n+1)}$ as the minimizer of the function

$$f_n(\theta_w) = ||Y - X_{\bar{w}}\hat{\theta}_{\bar{w}}^n - X\theta_w||^2$$

.

d)

```r
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
}
x <- scan('yield.txt')
p <- seq(5, 50, 5)
for(p_temp in p){
  r <- HP(x,lambda=2000, p=p_temp, niter=1000)
  plot(r$ss, main=sprintf('Objective Function at p = %d', p_temp))
}
```

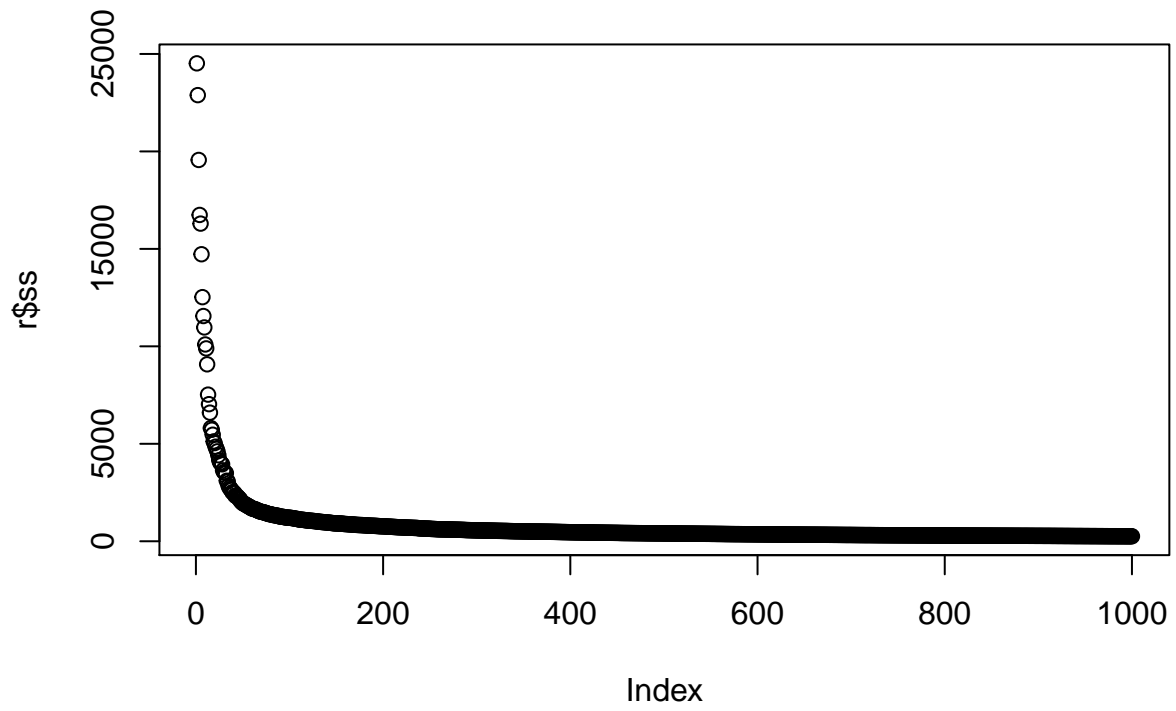# Objective Function at p = 5
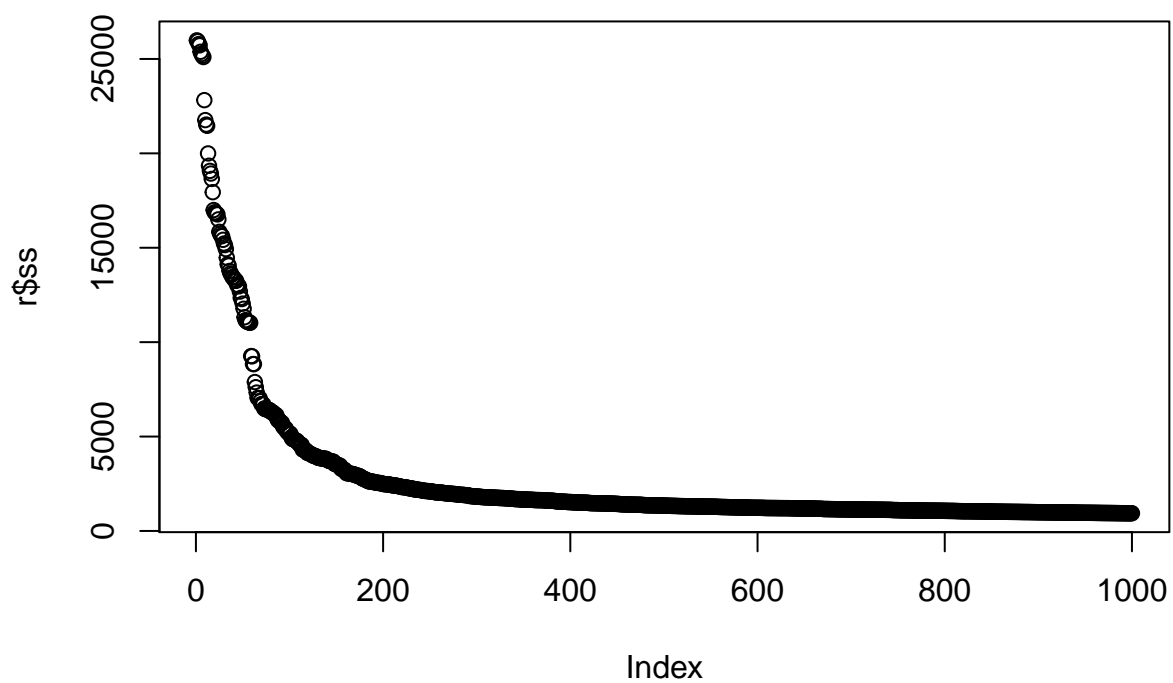
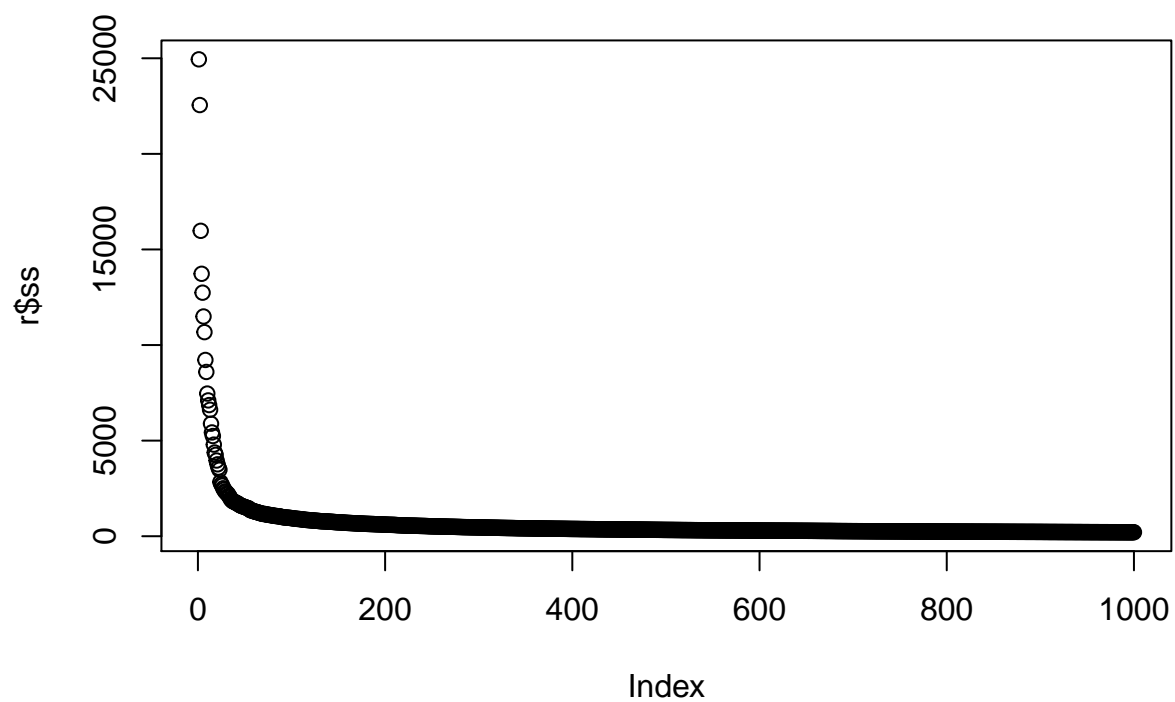# Objective Function at p = 10
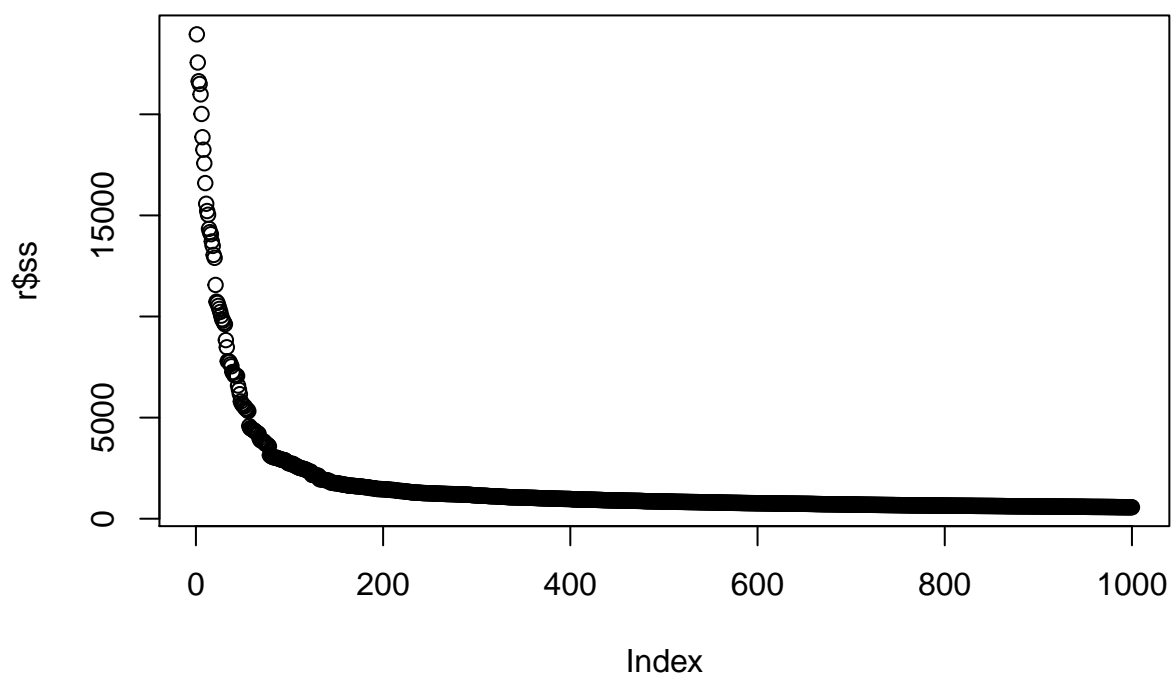
Objective Function at p = 15
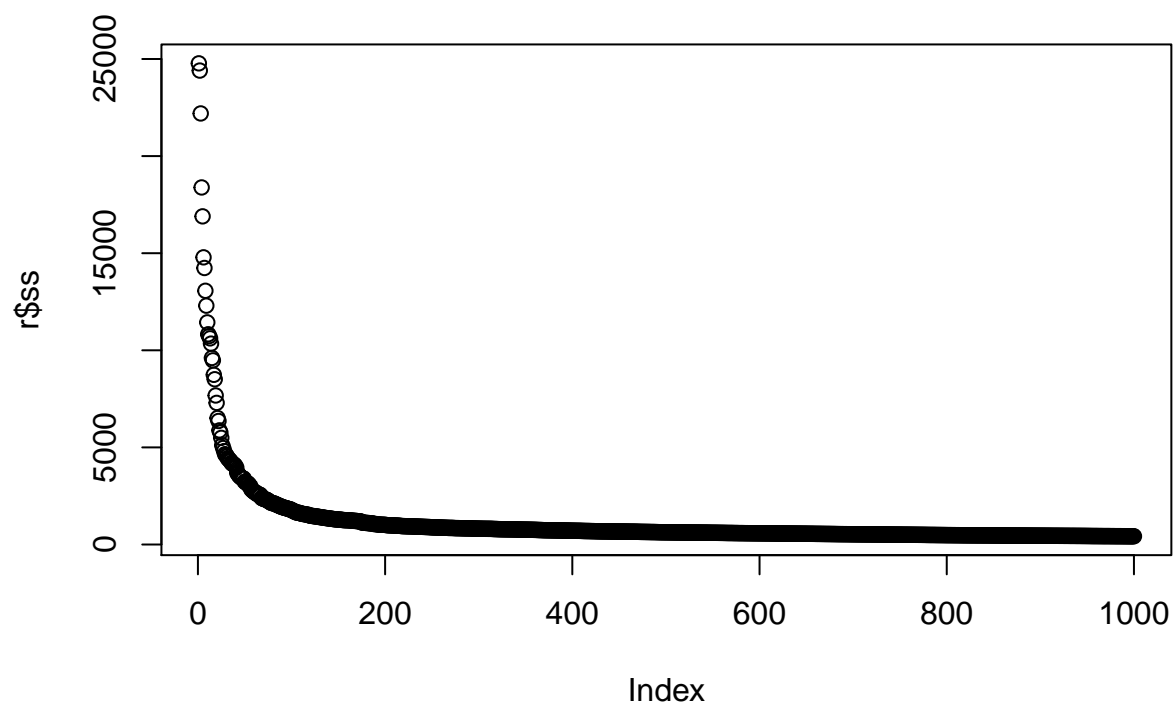
Objective Function at p = 20
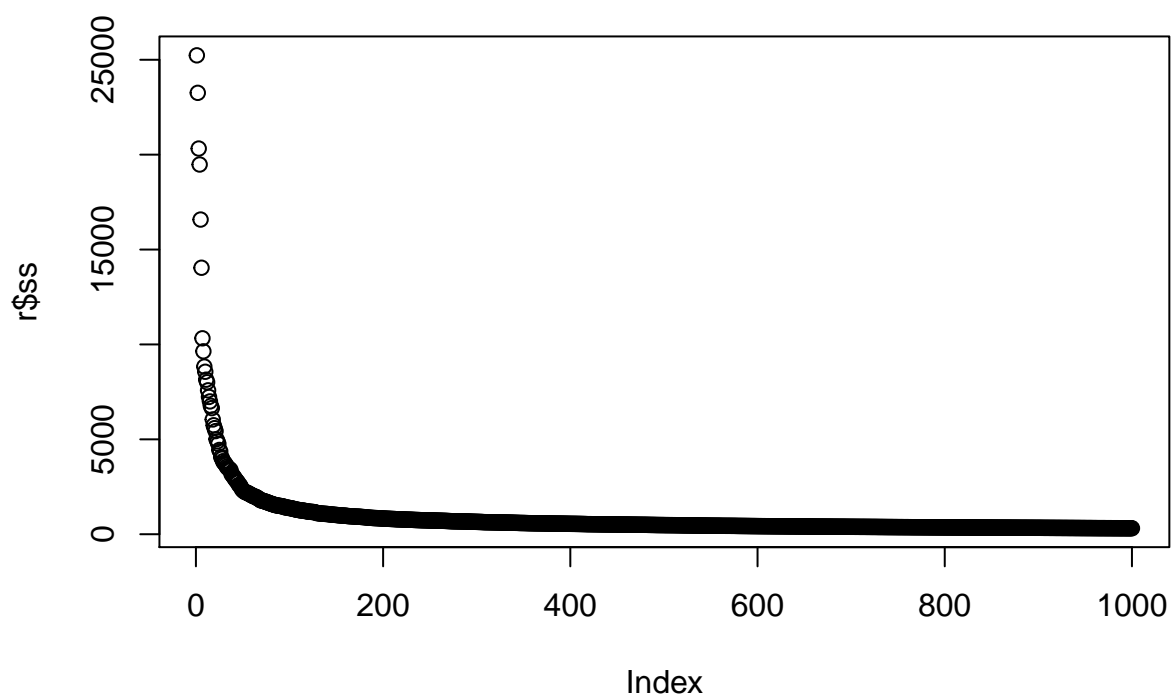
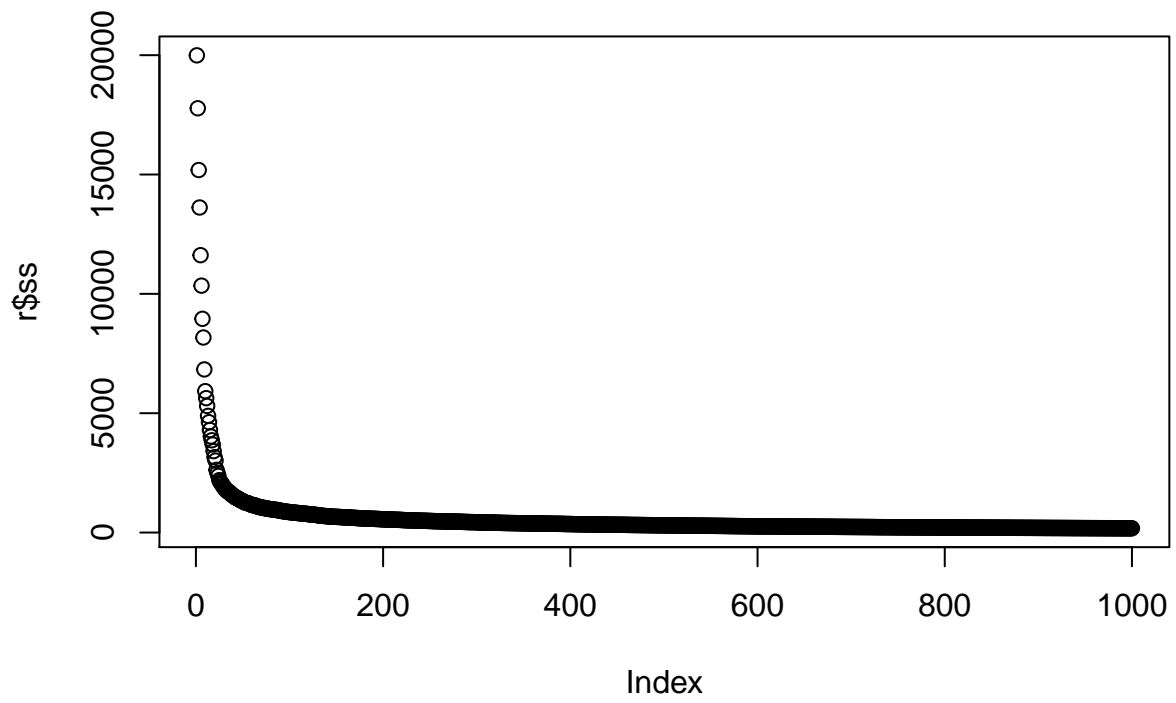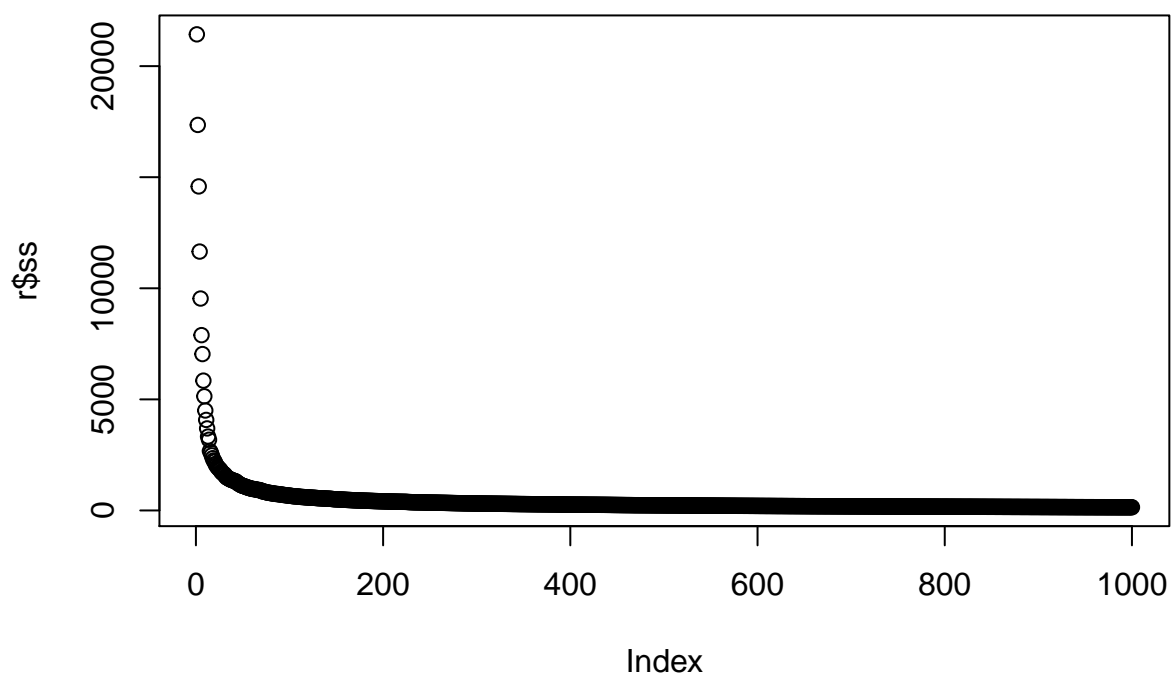# Objective Function at p = 25

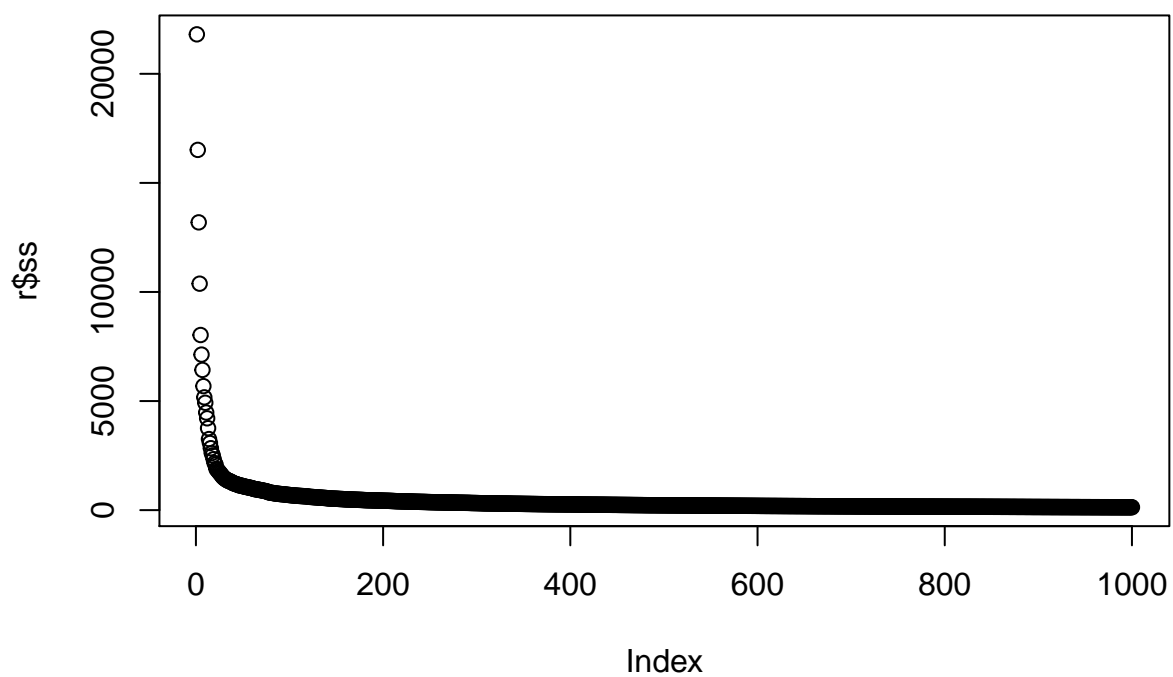# Objective Function at p = 30
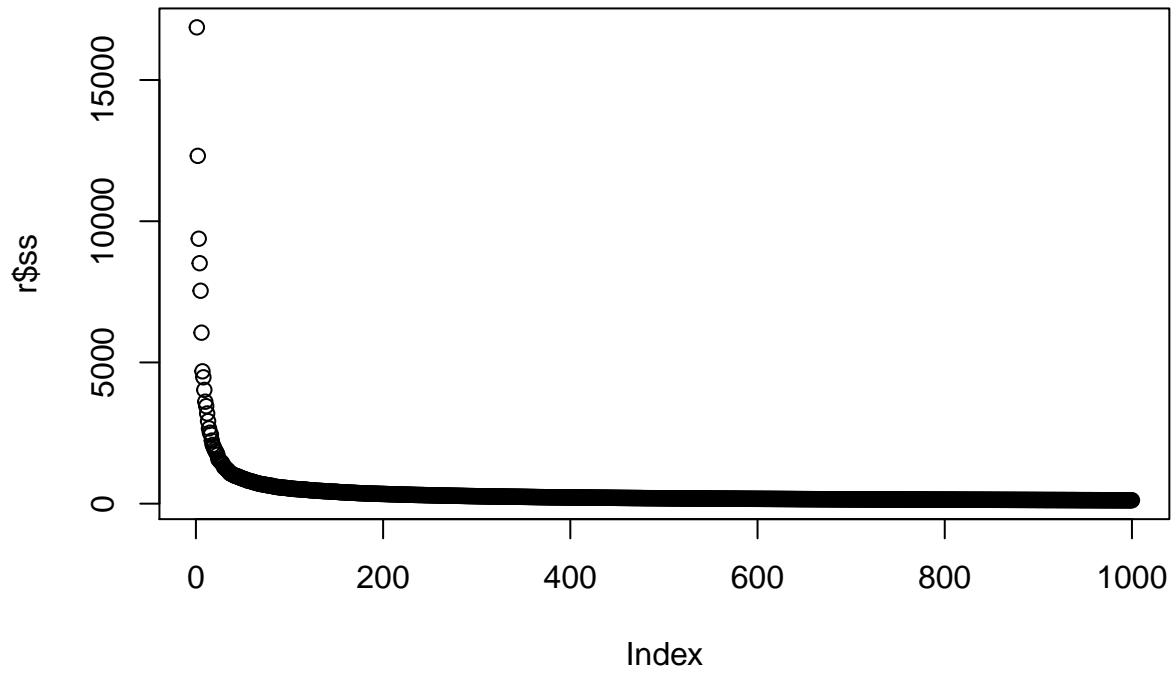
Objective Function at p = 35

Objective Function at p = 40

Objective Function at p = 45

Objective Function at p = 50

As p increases, the rate of convergence gets faster.