# Supoptimality Examples

*David Elsheimer*

*2/26/2020*

The `boxls` function is used to compute the least squares estimator $\hat{\beta}$ subject to the constraints that $\beta_j \in [0, 1]$ for $j = 1, \ldots, 20$. Using `boxls_gap`, the suboptimality of the output of `boxls` is computed. `factr` and `maxit` are changed around to see how `boxls` output changes.

**Answer:**

```
y<-as.matrix(read.csv("y.csv",header=FALSE))
X<- as.matrix(read.csv("X.csv",header=FALSE))



betas<- as.matrix(RStudio2020::boxls(X,y,runif(20),rep(0,20),rep(1,20),factr=1e-7,maxit=1e4)$par)
betas
```

```
##            [,1]
##  [1,] 0.5322818
##  [2,] 0.5155891
##  [3,] 0.2895702
##  [4,] 0.3370818
##  [5,] 0.1425379
##  [6,] 0.0000000
##  [7,] 0.0000000
##  [8,] 0.0000000
##  [9,] 0.1452307
## [10,] 0.5202368
## [11,] 0.1855009
## [12,] 0.6620775
## [13,] 1.0000000
## [14,] 1.0000000
## [15,] 0.6725628
## [16,] 0.1206004
## [17,] 0.0000000
## [18,] 0.0000000
## [19,] 0.0000000
## [20,] 0.0000000
```

```
RStudio2020::boxls_gap(X,y,betas,rep(0,20),rep(1,20))
```

```
##              [,1]
## [1,] 8.608673e-08
```

```
betas<- as.matrix(RStudio2020::boxls(X,y,runif(20),rep(0,20),rep(1,20),factr=1e-7,maxit=1e1)$par)
RStudio2020::boxls_gap(X,y,betas,rep(0,20),rep(1,20))
```

```
##             [,1]
## [1,] 1.79329e-07
```

```
betas<- as.matrix(RStudio2020::boxls(X,y,runif(20),rep(0,20),rep(1,20),factr=1e7,maxit=1e4)$par)
RStudio2020::boxls_gap(X,y,betas,rep(0,20),rep(1,20))
```

```
##             [,1]
```

```
## [1,] 0.007092113
```

```
betas<- as.matrix(RStudio2020::boxls(X,y,runif(20),rep(0,20),rep(1,20),factr=1e7,maxit=1e1)$par)
RStudio2020::boxls_gap(X,y,betas,rep(0,20),rep(1,20))
```

```
##              [,1]
## [1,] 0.001959748
```

```
betas<- as.matrix(RStudio2020::boxls(X,y,runif(20),rep(0,20),rep(1,20),factr=1e17,maxit=1e4)$par)
RStudio2020::boxls_gap(X,y,betas,rep(0,20),rep(1,20))
```

```
##          [,1]
## [1,] 3794.284
```

```
betas<- as.matrix(RStudio2020::boxls(X,y,runif(20),rep(0,20),rep(1,20),factr=1e17,maxit=1e1)$par)
RStudio2020::boxls_gap(X,y,betas,rep(0,20),rep(1,20))
```

```
##          [,1]
## [1,] 3145.659
```

The fewer iterations and the greater the factor, the less optimal our output becomes, as $g(\boldsymbol{\beta})$ gets larger and larger when these are increased.

**Step 4:** Recall the first order optimality condition

$$\langle \nabla \ell(\boldsymbol{\beta}^{\star}), \boldsymbol{\beta} - \boldsymbol{\beta}^{\star} \rangle \geq 0 \qquad \text{for all } \boldsymbol{\beta} \in \mathcal{D}.$$

The left hand side is a directional derivative of $\ell$ evaluated at a global minimizer of $\ell$ pointing into the feasible set $\mathcal{D}$.

1000 replicates of random $\boldsymbol{\beta}$s drawn from $\mathcal{D} = [0,1]^{20}$ are generated and the directional derivative at $\hat{\boldsymbol{\beta}}$ and the output of `boxls` are computed using these 1000 $\boldsymbol{\beta}$s.

```
library(Matrix)
b<-runif(20)
lb<-rep(0,20)
ub<-rep(1,20)

ddiv1<-Matrix(NA,nrow=0,ncol=1)
ddiv2<-Matrix(NA,nrow=0,ncol=1)
ddiv3<-Matrix(NA,nrow=0,ncol=1)
betamin1<-RStudio2020::boxls(X,y,b,lb,ub,factr=1e-7)$par
betamin2<-RStudio2020::boxls(X,y,b,lb,ub,factr=1e7)$par
betamin3<-RStudio2020::boxls(X,y,b,lb,ub,factr=1e17)$par

for(i in 1:1000){
  beta<-runif(20)
  ddiv1<-rbind(ddiv1,t(t(X)%*%(X%*%betamin1)-t(X)%*%y)%*%(beta-betamin1))
  ddiv2<-rbind(ddiv2,t(t(X)%*%(X%*%betamin2)-t(X)%*%y)%*%(beta-betamin2))
  ddiv3<-rbind(ddiv3,t(t(X)%*%(X%*%betamin3)-t(X)%*%y)%*%(beta-betamin3))
  }
summary(as.vector(ddiv1))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1059    2412    2825    2827    3253    4724
```

```
summary(as.vector(ddiv2))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##    1059    2412    2825    2827    3253    4724
```

```r
summary(as.vector(ddiv3))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1458.1   298.9   793.4   804.7  1301.3  3341.7
```

As more and more suboptimal $\hat{\beta}$'s are considered (where the factor is increased), the mean of the directional derivatives decreases, and the variance of the distribution increases, as noted by the increased spread of the distribution and more extreme maximum and minimum directional derivatives.