# Lab Part 0

Short Course

## Timing the spatialProcess function

Motivation for problems as problem size gets large

Note: loading **LatticeKrig** package automatically loads **fields**.

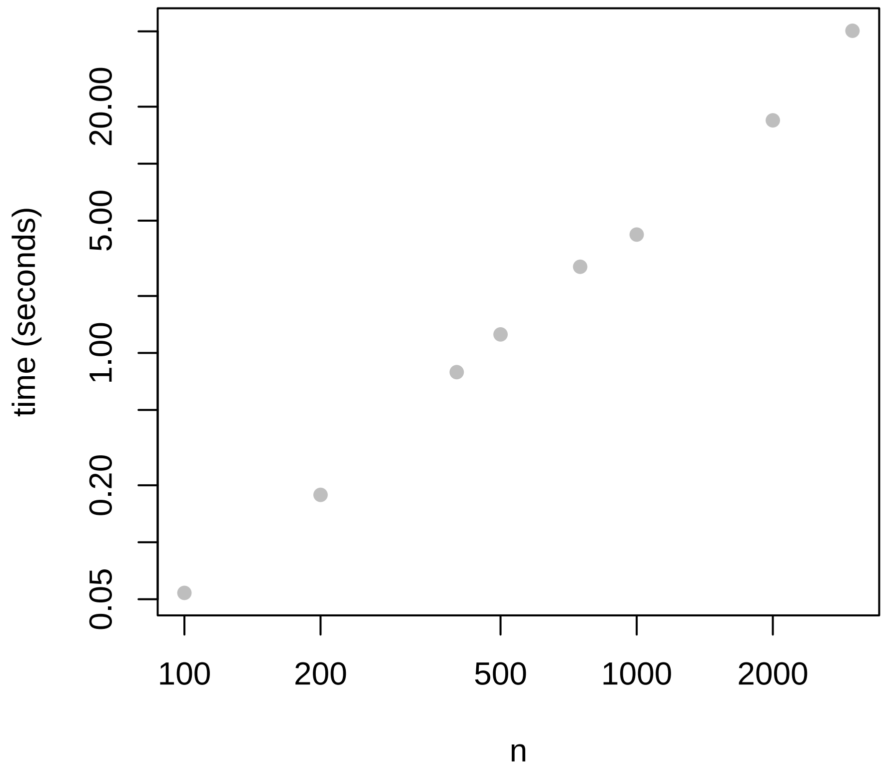
```
set.seed(222)
tabSP<- NULL
nObs<- c(100, 200, 400,  500, 750, 1000, 2000, 3000)
for( n in nObs ){
  x<- cbind( seq( 0,1,,n))
  Sigma<- Matern( rdist(x,x)/.15, smoothness = 1.0)
  U<- rnorm(n)
  E<- rnorm(n)
# add in a linear part too to match fitting
  y<-   (1 + x) + t(chol(Sigma))%*%U  + .05*E
  # fix the range to compare with example later on
  elapsedTime<- system.time(
              out<- spatialProcess( x,y, aRange=.15 )
        )
  #print(c( n,elapsedTime[3]) )
  tabSP<-  rbind( tabSP, c( n,elapsedTime[3]) )
}

print( tabSP)
```

```
##          elapsed
## [1,]  100   0.054
## [2,]  200   0.178
## [3,]  400   0.792
## [4,]  500   1.253
## [5,]  750   2.853
## [6,] 1000   4.221
## [7,] 2000  16.939
## [8,] 3000  50.368
```

Take a look at a log-log plot. Linear in log-log means a polynomial relationship.

```
plot( tabSP, log="xy", xlab="n", ylab="time (seconds)",
        col="grey", pch=16)
```




```r
y<- log10(tabSP[,2])
x<- log10(tabSP[,1])
lm( y~x)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##      -5.282        1.988
```

## Timing just the Cholesky

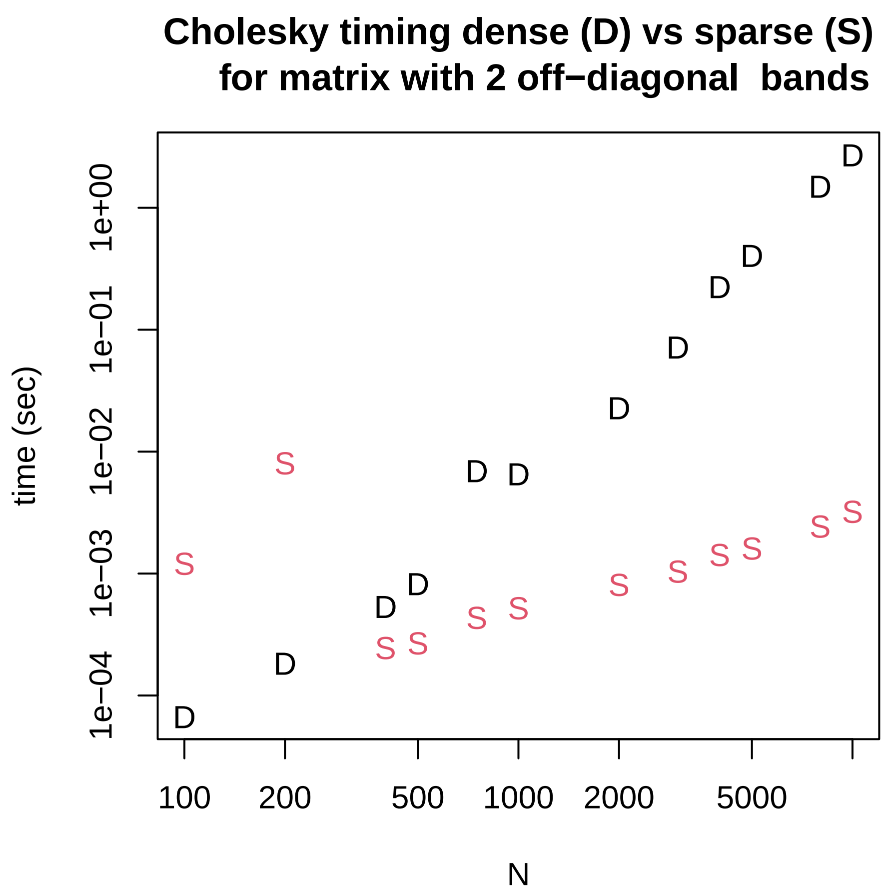
Focusing just on the Cholesky decoposition – theoretically the most time consuming step. Use a test matrix that has lots of zeroes and compare timing to sparse Cholesky decomposition.

```
sizes<- c(100, 200, 400,  500, 750, 1000, 2000, 3000, 4000, 5000, 8000, 10000)
NTotal<- length( sizes)
tabChol<- matrix( NA, nrow= NTotal, ncol=4)
dimnames(tabChol)<- list( NULL, c("N","Dense",
                              "Sparse","speedup"))
for(k in 1:NTotal) {
 N<- sizes[k]
#weights are a 4th differece
# sparse matrix construction using LatticeKrig utility
SMat <- LKDiag( c(1,  -10,   27,  -10,    1), N)
# convert to full ( now the zeroes are consider real values)
FMat <- spam2full(SMat)
# dense matrix Cholesky
startTime <- Sys.time() #
FChol <- chol(FMat)
deltaF<- as.numeric(Sys.time() - startTime) #
# sparse matrix Cholesky
startTime <- Sys.time()
SChol <- chol(SMat)
deltaS<- as.numeric(Sys.time() - startTime )
tabChol[k,]<- c(N,deltaF, deltaS, deltaF/deltaS )
}
print( tabChol)
```

```
##           N        Dense       Sparse      speedup
##  [1,]   100 6.699562e-05 0.0012130737   0.05522799
##  [2,]   200 1.838207e-04 0.0080890656   0.02272459
##  [3,]   400 5.369186e-04 0.0002470016   2.17374517
##  [4,]   500 8.258820e-04 0.0002698898   3.06007067
##  [5,]   750 6.976128e-03 0.0004367828  15.97161572
##  [6,]  1000 6.571054e-03 0.0005238056  12.54483386
##  [7,]  2000 2.288318e-02 0.0008139610  28.11335677
##  [8,]  3000 7.202697e-02 0.0010461807  68.84753874
##  [9,]  4000 2.255261e-01 0.0014319420 157.49667000
## [10,]  5000 4.066441e-01 0.0016179085 251.33937518
## [11,]  8000 1.501122e+00 0.0024490356 612.94411994
## [12,] 10000 2.714866e+00 0.0032370090 838.69580909
```

Log- log plot to look for polynomial dependence

```
matplot( tabChol[,1], tabChol[,2:3], type="p", pch=c( "D","S"),
xlab="N", ylab="time (sec)", log="xy")
title("Cholesky timing dense (D) vs sparse (S)
      for matrix with 2 off-diagonal  bands ")
```




fitting line by OLS

```
y<- log10(tabChol[6:12,2])
x<- log10(tabChol[6:12,1])
lm( y~x)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##     -10.492        2.724
```

```
y<- log10(tabChol[6:12,3])
x<- log10(tabChol[6:12,1])
lm( y~x)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##     -5.6685       0.7843
```

## On your own . . .

1. Extrapolating from the smaller sample results in **tabSP** estimate the time for **spatialProcess** to handle a problem of size 26000 ( about the size of the CO2 data set.)

2. Is the time for spatialProcess ( **tabSP[,2]**) linearly related to the time for the Cholesky decomposition ( **tabChol[1:8, 2]**)?