

Assignment 3 - The free position facility location problem

Pranav Kasela 846965

Contents

The Problem	1
Distance function	1
The assignment and the solution	1

The Problem

In this assignment we will look at the **Facility Location Problem**. The problem can be described in general terms as follows:

We must decide the placement of a facility that distributes a certain good to a group of consumers that need it.

The placement must be chosen in order to **minimize** the total compound distance from the facility and the customers.

The following assumptions has to be taken into account:

1. The possible location for the facility is unknown, that is the problem is to find the right spot to build it.
2. The facility building costs are fixed and independent from the position of the building site.

Notice that in this scenario there is one possible decision to make:

- where to build the facility, that is find the position (χ, v) that minimises the compound distance of the facility with respect to all customers.

Data

A file with the locations of the consumers can be found in the **Data** folder.

Distance function

Given the position of the facility $f = (\chi, v)$ and of a consumer $p_i = (x_i, y_i)$ use the following formula to calculate the distance between them.

$$d(f, p_i) = \log((\chi - x_i)^2 + 1) + \log((v - y_i)^2 + 1)$$

The assignment and the solution

1. **Formulate the objective function to minimize for the described problem.**

Our Problem, in this case, is to minimize the function:

$$\min \quad d(\bar{f}) = \sum_{j=1}^n \text{distance}(\bar{f}, \bar{p}_j)$$

Where \bar{f} is the location of the facility while \bar{p}_j is the location of the j-th customer, with the distance as defined above.

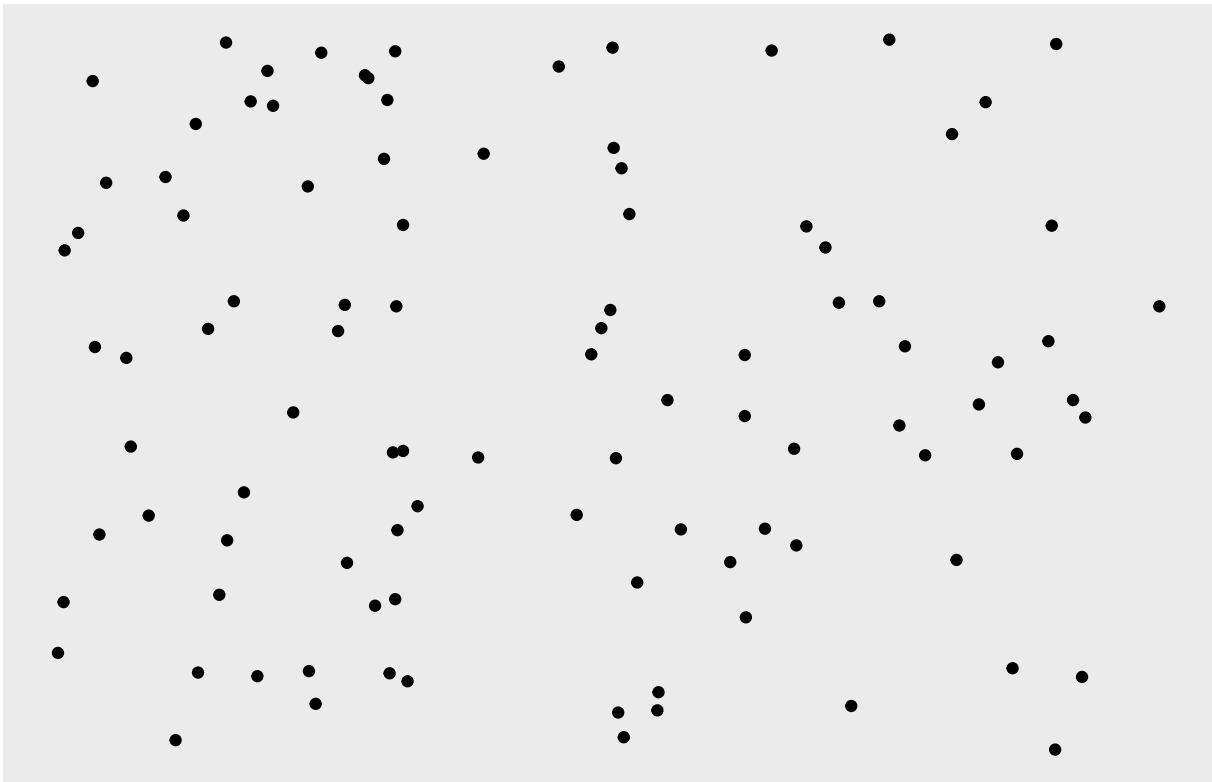
```
require(ggplot2)
require(gganimate)
require(curry)
require(optimx)
require(dplyr)
```

```
customer_locations <- read.csv("consumer_locations.csv")
```

```
plt <- ggplot(customer_locations, aes(x=x,y=y)) +
  geom_point() +
  ggtitle("Customers' Location Distribution") +
  theme(axis.ticks = element_blank(),
        axis.text = element_blank(),
        panel.grid = element_blank(),
        axis.title = element_blank(),
        plot.title = element_text(hjust = 0.5))
```

```
plt
```

Customers' Location Distribution



We see that the points are randomly distributed among the plane, this indicates the the minimum of the

function (the facility location) will not be near border otherwise the facility will be very far from the location on the opposite border.

2. Express in analytical form the gradient for the objective to minimize.

We start calculating the gradient of the function $d(\bar{f}) = d(\chi, v)$:

$$\nabla = \left(\sum_i \frac{2(\chi - x_i)}{(\chi - x_i)^2 + 1}, \sum_i \frac{2(v - y_i)}{(v - y_i)^2 + 1} \right)$$

This gradient will be used in the numerical methods to search for a local minimum.

3. Implement the Gradient Descent method and solve the problem with it.

In the Gradient Descent method, we decide to decay the learning rate (lr) with the number of iterations, based also on a coefficient (lr.decay) ≤ 0 , the decay is not drastic if the lr.decay is chosen accordingly, but if the lr.decay is wrongly chosen the method might not converge and slow down a lot and stop long before the zero. The smaller the value of lr.decay the smaller will be the decay, and for lr.decay = 0 there is no decay. It is similar to the concept of the stickiness, if stickiness coefficient = 0 then the movement is purely based on the gradient of the function, if it is > 0 the ball loses its ability to move after a while and sticks to the surface.

The idea behind it is to slow down the gradient descent after a while in order to achieve a greater precision in finding the zero and avoiding the situation of the estimated “bouncing” back and forth near the minimum (or maximum) of the function. Usually the learning rate or the step must be calculated at each step, as another minimization or maximization problem, but it can become really expensive computationally, another method is to use the Newton method calculating the inverse of the Hessian matrix and choosing it as the learning rate.

Another change is that we added a stopping criterion, in our case we decided to stop when the squared norm of the gradient vector $\|\nabla\|^2$ is smaller than a certain tolerance (which is by default 10^{-6}) the method will stop, this choice was made to make the code a little bit more efficient since it was useless to continue any further because the iteration step is defined as the previous value - lr*gradient, so if the gradient is too low the movement will be minimal and a gradient near to 0 indicated that a stationary point has been found.

The function returns a list consisting of all the points created with the iteration, a vector containing the change in value of the objective function (the distance), the optimal point and the value of the objective function in the optimal point and finally a boolean value indicating if the method converges or not (it is done by checking if the square of the norm of the gradient was smaller than the given tolerance).

```
distance <- function(data,f){
  sum(log((f[1]-data[,1])^2+1) + log((f[2]-data[,2])^2+1))
}

#fix the customer_location parameter in the distance function
fn.distance <- curry(distance, customer_locations)

gradient <- function(data,f){
  c(sum(2*(f[1]-data[,1])/((f[1]-data[,1])^2 +1)),
    sum(2*(f[2]-data[,2])/((f[2]-data[,2])^2 +1)))
}

#fix the customer_location parameter in the gradient function
fn.gr <- curry(gradient, customer_locations)

# gradient descent function
```

```

gradientDescent <- function(f.init, fn.gr, lr, max_iters=1000,
                           lr.decay = 0, toll=1e-6){
  f      <- matrix(NA, nrow = max_iters+1, ncol = 2)
  grad   <- matrix(NA, nrow = max_iters+1, ncol = 2)
  f[1,]  <- f.init
  lr.init <- lr
  grad[1,]<- fn.gr(f[1,])
  for(k in 1:max_iters){
    f[k+1,] <- f[k,] - lr*grad[k,]
    #decay the lr as the iteration increases
    #another solution is the Newton solution H.inverse(par[k,])
    lr      <- lr.init*exp(-k*lr.decay)
    grad[k+1,] <- fn.gr(f[k+1,])
    #stop if norm of gradient^2 is smaller than toll practically 0
    if(sum(grad[k+1,]^2) < toll)
      break
    iter <- k
  }
  f = f[!is.na(rowSums(f)),]
  distance <- apply(f,1,fn.distance)
  if (sum(grad[k+1,]^2) < toll)
    convergence = TRUE
  else
    convergence = FALSE
  return(list(f      = f,
             dist     = distance,
             opt_point = tail(f,1),
             min_dist  = tail(distance,1),
             convergence = convergence,
             lr        = lr,# only for testing and development
             iters     = iter))
}

```

*# A function to print the solution the same function will be used also for the
 # stochastic gradient method since the output list will have the same format*

```

print.solution <- function(result){
  P      <- result$opt_point
  P_dist <- result$min_dist
  f      <- result$f
  f_dist <- result$dist
  if (result$convergence)
    converge <- "The method converges to the minimum."
  else
    converge <- "The method doesn't converge to the minimum."

  msg <- paste0(converge, " ",
               "The minimum was the point of coordinates (",
               round(P[1],2), ",", round(P[2],2),
               ") and the value of the function is ",
               round(P_dist,2), " in ", result$iters,
               " iterations.")
  print(msg)
}

```

```

ggplot(data=data.frame(distance=f_dist)) +
  aes(x=distance) +
  geom_line(aes(x = 1:length(distance),y=distance)) +
  labs(title = "Convergence plot",
        x = "Iterations", y="Distance Value") +
  theme(plot.title = element_text(hjust = 0.5))
}

```

Here is a first attempt to find a solution using the gradient descent method.

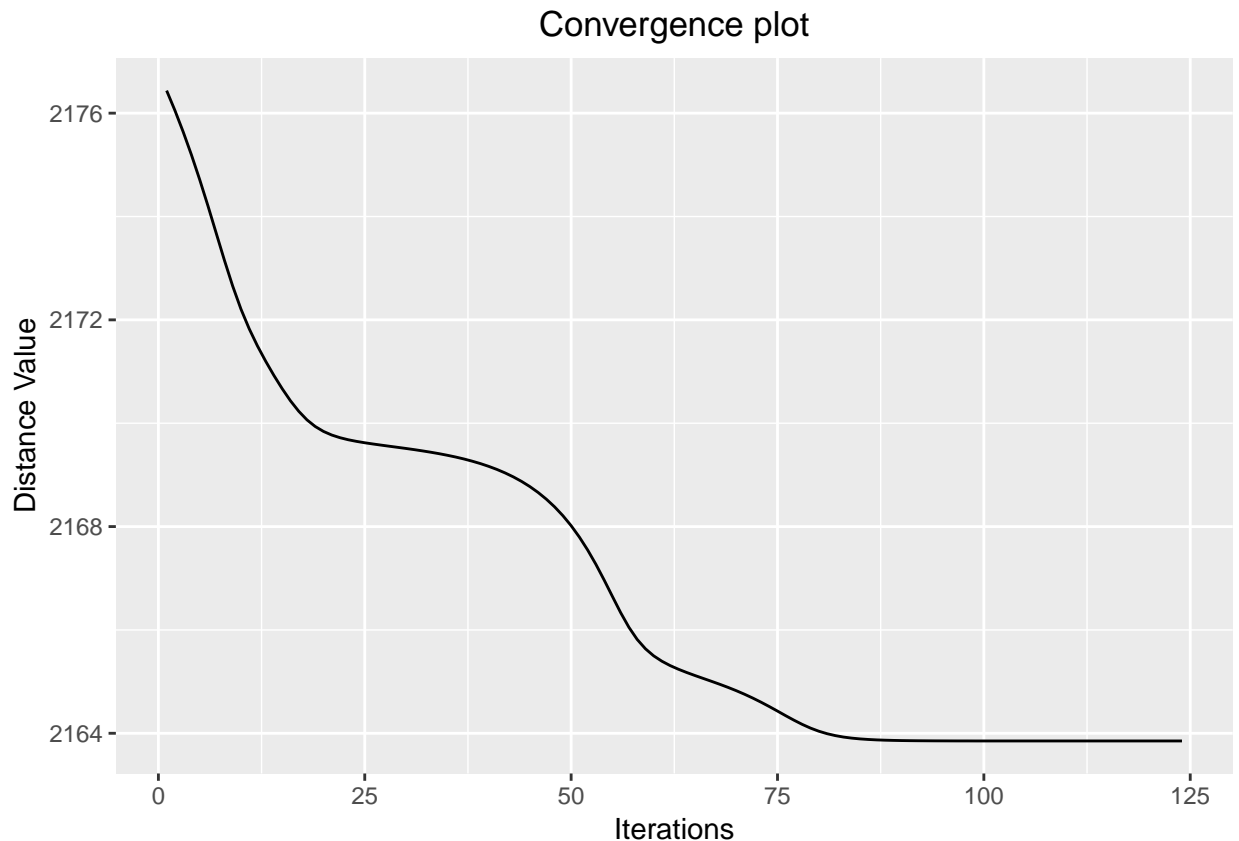
```

res <- gradientDescent(c(runif(1,0,1000),runif(1,0,1000)),
                      fn.gr=fn.gr, lr = 0.1,
                      max_iters = 1000, lr.decay = 0.001)

print.solution(res)

```

[1] "The method converges to the minimum. The minimum was the point of coordinates (121.32,127.71) and



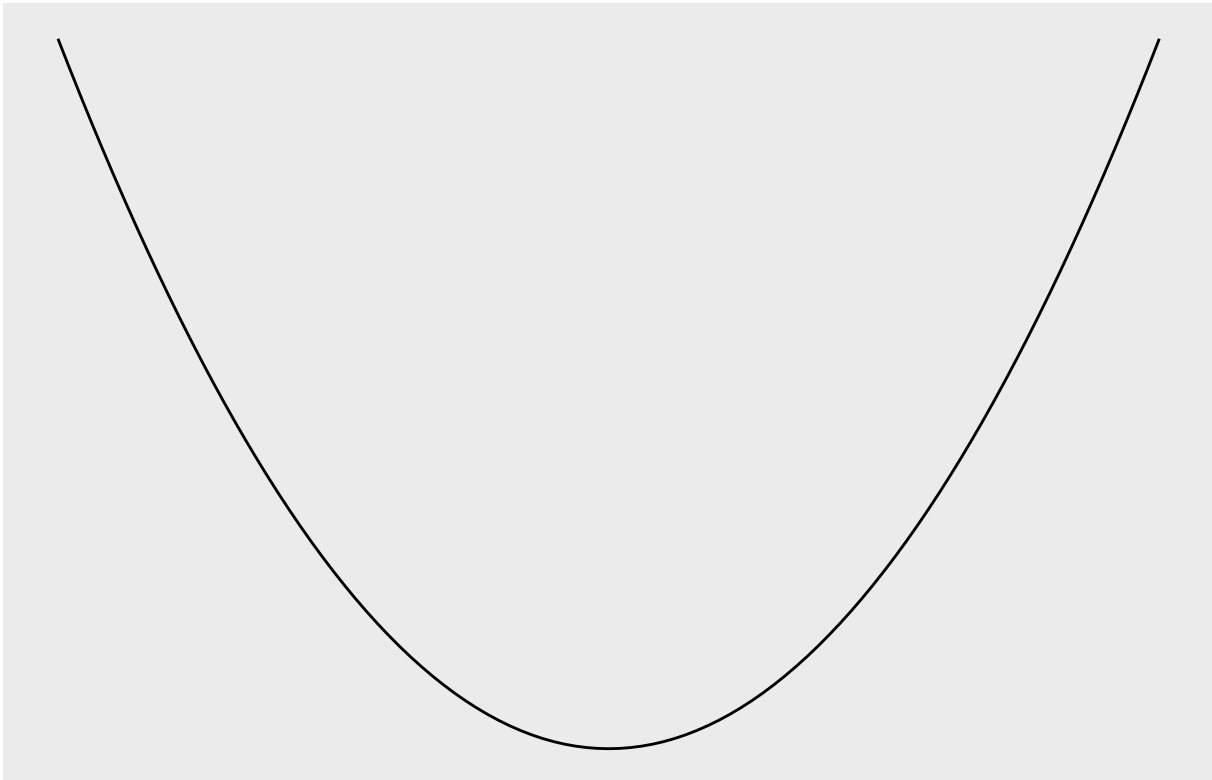
We notice that the solution is heavily influenced by the initial point (we noticed it on different run of the latter code), to verify if we have a local minimum problem we try to do a 3D plot and a contour plot of the distance function, we expect it to have a lot of minimum function since the function distance from single point has only 1 minimum which is the point itself it has distance 0 from it self and distance for defintion is positive, but as we sum all these distances and it creates a lot of minimum between the point.

It is more intuitive to think a point a heavy mass which deforms the plane and the deformation is based on the distance itself, in the case of a euclidean distance the deformation is a paraboloid, in this case we have a logarithmic tranformation of the euclidean distance traslated by +1, thus changing the concavity of the distance function. Now as we put more points on the plane we have that the deformation of this plane creates a lot of local minimum and also a lot of maximum.

```
quadratic <- function(x) x^2

ggplot(data = data.frame(x=0), mapping = aes(x=x)) +
  stat_function(fun = quadratic) + xlim(-5,5) +
  labs(title = "Section of Euclidean distance") +
  theme(axis.ticks = element_blank(),
        axis.text = element_blank(),
        panel.grid = element_blank(),
        axis.title = element_blank(),
        plot.title = element_text(hjust = 0.5))
```

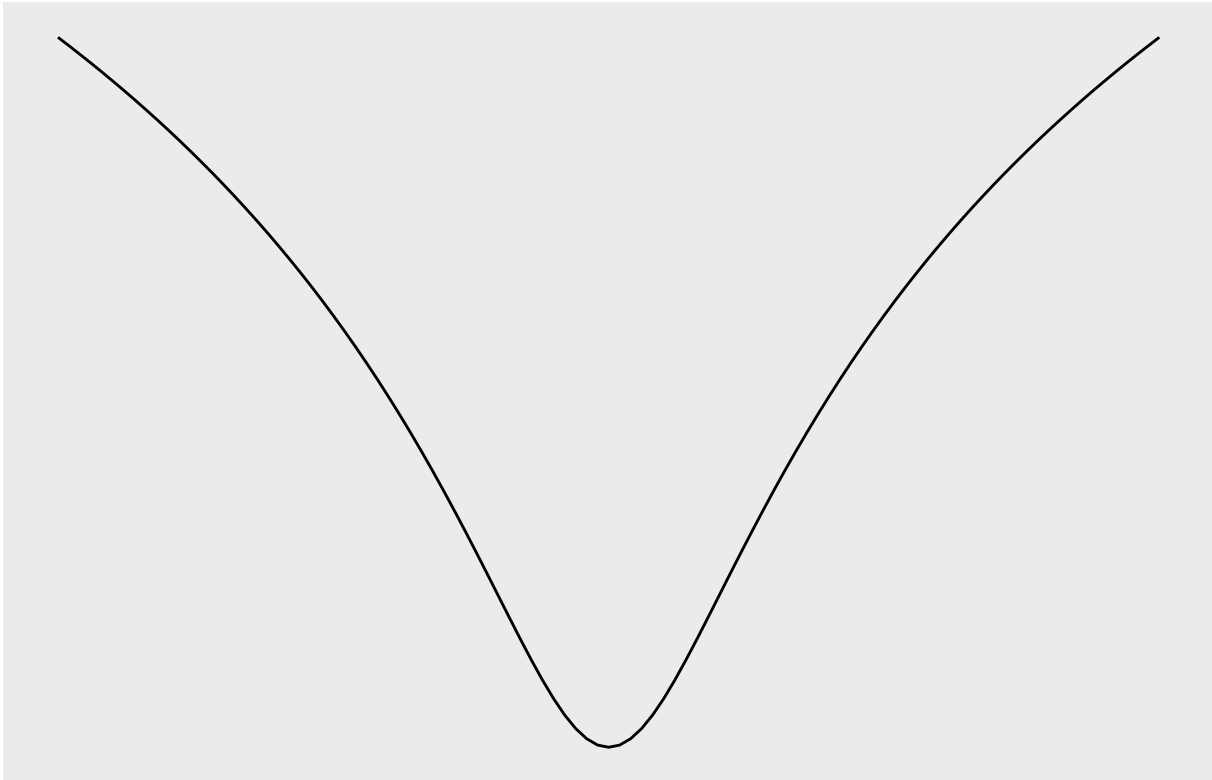
Section of Euclidean distance



```
logarithmic <- function(x) log(x^2+1)

ggplot(data = data.frame(x=0), mapping = aes(x=x)) +
  stat_function(fun = logarithmic) + xlim(-5,5) +
  labs(title = "Section of log transformation of Euclidean dist.") +
  theme(axis.ticks = element_blank(),
        axis.text = element_blank(),
        panel.grid = element_blank(),
        axis.title = element_blank(),
        plot.title = element_text(hjust = 0.5))
```

Section of log transformation of Euclidean dist.

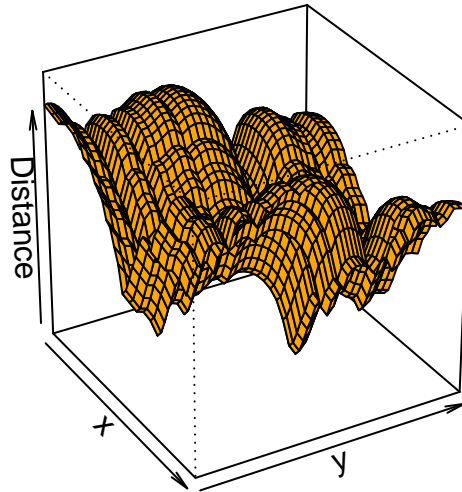


Let's try to see it with a 3D plot helped by a contour plot, in order to avoid complex and black plot each point is distant 5 units from each other and we plot only between: $400 < x < 600$ and $300 < y < 500$ to show the problem of a lot a local minima and maxima.

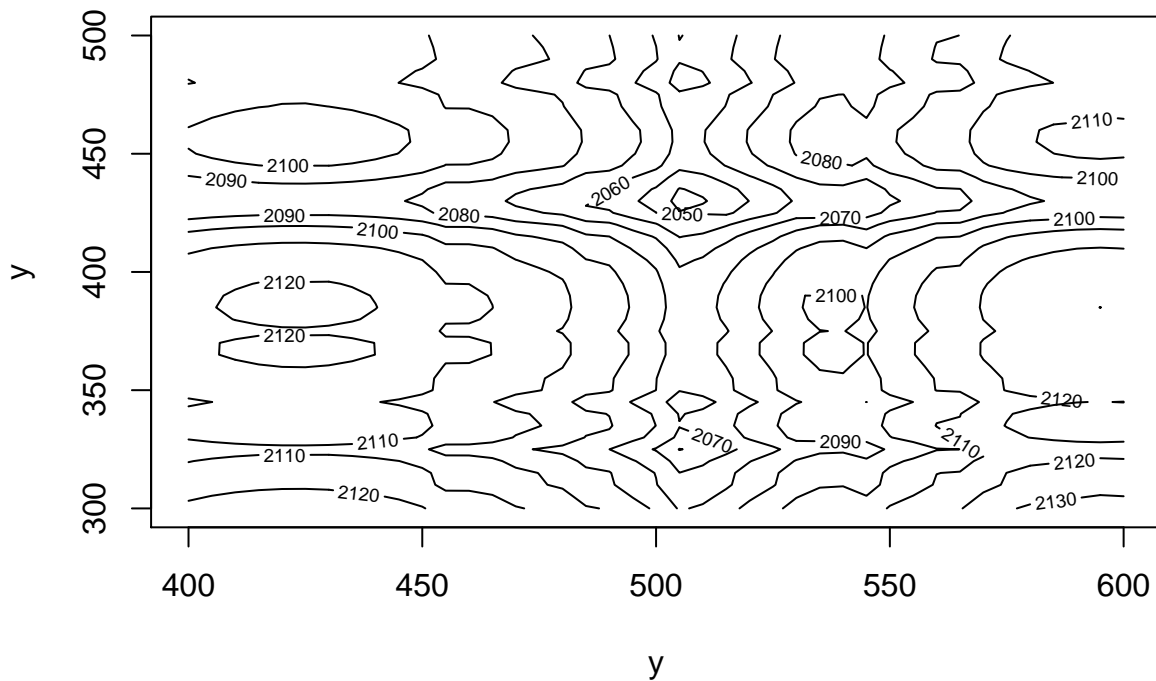
```
helper.dist <- function(x,y){ #m and c are vectors
dist = matrix(nrow = length(x),ncol = length(y))
for (i in seq(1, length(x), by = 1)) {
  for (j in seq(1, length(y), by = 1)) {
    dist[i,j] = fn.distance(c(x[i], y[j]))
  }
}
dist
}

x=seq(400, 600, 5)
y=seq(300, 500, 5)
z <- helper.dist(x,y)

persp(x, y, z, phi = 30, theta = 60,col = "orange",xlab = "x",ylab = "y", zlab = "Distance", r=10, d=5)
```



```
contour(x,y,z, xlab="y",ylab="y")
```



A great method to solve this problem, in this less complex case, is to initiate a lot of random points from which we start the solution, one would be tempted to use a uniform distribution of the point.

```
set.seed(123456789)
x <- runif(1000,0,1000)
y <- runif(1000,0,1000)
random.points <- cbind(x,y)

ris_temp <- lapply(1:length(y), function(x) {
  gradientDescent(random.points[x,],fn.gr,lr = 0.1,max_iters = 1000,lr.decay = 0.0001)})

opt_points <- matrix(NA,ncol=2, nrow=length(x))
distances <- rep(NA,length(x))

for (i in 1:nrow(random.points)){
```



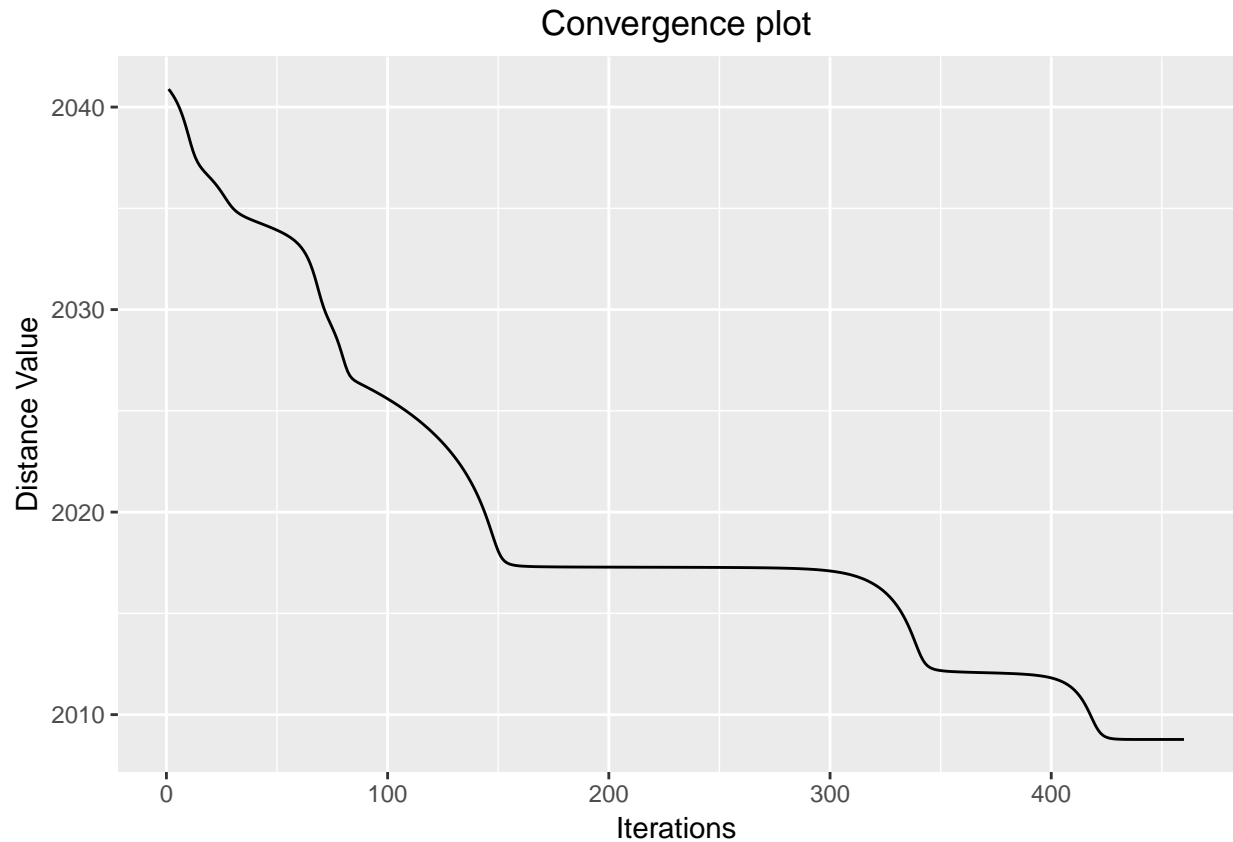
```

    opt_points[i,] <- ris_temp[[i]]$opt_point
    distances[i]   <- ris_temp[[i]]$min_dist
  }

print.solution(ris_temp[[which.min(distances)]])

```

```
## [1] "The method converges to the minimum. The minimum was the point of coordinates (310.16,565.48) and
```



Another choice could have been to generate the random point using a normal distribution with the standard deviations and means of the x,y components of the customer locations. This way the points are concentrated in the middle where we suspect the global minimum to be (more density \rightarrow less distance between points).

```

set.seed(123456789)
sd.x <- sd(customer_locations$x)
mean.x <- mean(customer_locations$x)
sd.y <- sd(customer_locations$y)
mean.y <- mean(customer_locations$y)
x <- rnorm(500,mean.x,sd.x)
y <- rnorm(500,mean.y,sd.y)
random.points <- cbind(x,y)

ris_temp <- lapply(1:length(y), function(x) {
  gradientDescent(random.points[x,],fn.gr,lr = 0.1,max_iters = 1000,lr.decay = 0.0001)})

opt_points <- matrix(NA,ncol=2, nrow=length(x))
distances <- rep(NA,length(x))

for (i in 1:nrow(random.points)){

```

```

    opt_points[i,] <- ris_temp[[i]]$opt_point
    distances[i]   <- ris_temp[[i]]$min_dist
  }

```

```

print.solution(ris_temp[[which.min(distances)]])

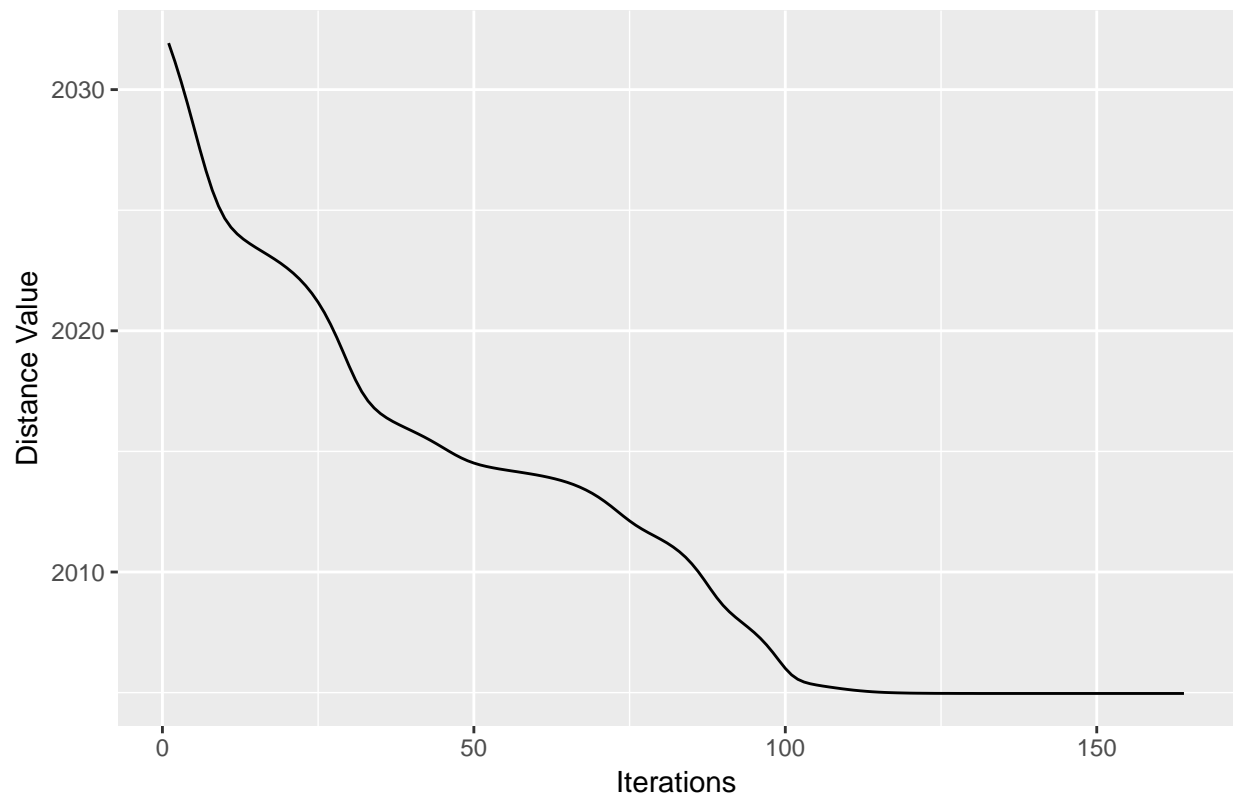
```

```

## [1] "The method converges to the minimum. The minimum was the point of coordinates (310.16,430.87) and

```

Convergence plot



```

# P      <- ris_temp[[which.min(distances)]]$f
# final  <- ris_temp[[which.min(distances)]]$opt_point
# z      <- seq(1,nrow(P),2)
#
# ggif_minimal <- data.frame(x=P[z,1],y=P[z,2],z=z)%>%
#   ggplot(aes(x = x, y = y))+
#   geom_point(col="magenta",alpha=0.5) +
#   labs(title = "Iteration of the point towards the minimum")+
#   geom_point(data=data.frame(x=final[1],y=final[2]),col="blue")+
#   geom_text(data=data.frame(z=z),
#             mapping = aes(x = 300, y = 435,
#                           label = paste0("ITERATION N. ",z)))+
#   theme(plot.title = element_text(hjust = 0.5))+
#   transition_reveal(z)+
#   ease_aes("linear")+
#   enter_appear()+
#   exit_fade()
#
# animate(ggif_minimal, fps=90)

```

The best solution found is (310.16,430.87) with a value of 2004.96.

4. Solve the problem with a package provided by R (for instance, using the function `optimr` within the package `optimx`). Note that it is not required to use the gradient descent algorithm to solve the problem, other algorithms can be used as well.

In this case there is nothing to comment, we just use the function provided by `optimr` on the random points generated randomly using a normal distribution.

```
ris_temp <- lapply(1:length(y), function(x) {
  optimr(random.points[x,],fn.distance, fn.gr,method = "CG")})

opt_points <- matrix(NA,ncol=2, nrow=length(x))
distances <- rep(NA,length(x))

for (i in 1:nrow(random.points)){
  opt_points[i,] <- ris_temp[[i]]$par
  distances[i] <- ris_temp[[i]]$value
}

proptimr(ris_temp[[which.min(distances)]])
```

```
## Result ris_temp[[which.min(distances)]] proposes optimum function value = 2004.964 at parameters
##      x      y
## 310.1627 430.8702
## After 76 fn evals, and 33 gr evals
## Termination code is 0 :
## -----
```

5. Implement the Stochastic gradient descent algorithm with mini-batches and use it to solve the problem.

In the stochastic method, since our data has the same scale there is no need to normalize. For the stopping criterion we choose to stop if the $\|\nabla\|^2 < \text{toll}$ or if the improvement of the approximation is very low. The subset of the points is chosen as the 20% of the initial data randomly sampled with replacement. Also in this case we adapt the decaying learning rate.

```
stoch_gradDescent <- function(xy,f.init, fun.gr, lr,
                              max_iters=1000, lr.decay=0){
  f <- f.init
  # Initialize a matrix to store values of the point
  # and distance for each iteration
  P <- matrix(NA, nrow = max_iters + 1, ncol = 2)
  dist <- rep(NA,times=max_iters)
  P[1,] <- f
  dist[1] <- fn.distance(P[1,])
  # set seed value for random sampling
  set.seed(42)
  lr.init <- lr
  # now iterate using mini batches of randomly sampled data,
  for (i in 1:max_iters) {
    # randomly sample 20% of data from the xy data frame
    xysamp <- xy[sample(nrow(xy), floor(nrow(xy)*0.2),
                        replace = TRUE), ]
```

```

# update point using mini batches
f <- f - lr * fun.gr(xysamp, f)
lr <- lr.init * exp(-i*lr.decay)
# save the beta values for iteration i to a matrix for future plotting
P[i+1,] <- f
iter <- i
dist[i+1] <- fn.distance(P[i+1,])
} # end for loop
P <- P[!rowSums(is.na(P)),]
dist <- dist[!is.na(dist)]
if (sum((P[i+1,]-P[i,])^2)<1e-6)
  convergence <- TRUE
else
  convergence <- FALSE
return(list( f      = P,
             dist    = dist,
             opt_point = P[which.min(dist),],
             min_dist  = dist[which.min(dist)],
             convergence = convergence,
             lr        = lr, # only for testing and development
             iters     = iter))
}

```

The best choice is always to start randomly, so we choose the random values already created, for the Point 3. Here we see that the method does not converge exactly at the desired point, one of the reasons is how the stochastic method is defined, but also the shape of the function is so that the gradient rapidly becomes zero very near the minimum (as we can see in the plot in point 3 of the section of the log transformation), so even if the solution is near to the minimum it can have a gradient not even close to zero. Infact we can see that the difference between the given solution from the stochastic method and the gradient method is small (the difference is of an order of 10^{-2}) but the gradient is not small enough to be considered zero.

```

# We take randoms from the point 3.
ris_temp <- lapply(1:nrow(random.points), function(x) {
  stoch_gradDescent(customer_locations,random.points[x,],
                    gradient,lr=0.1,max_iters=2000,
                    lr.decay=0.0015)})

opt_points <- matrix(NA,ncol=2, nrow=length(x))
distances <- rep(NA,length(x))

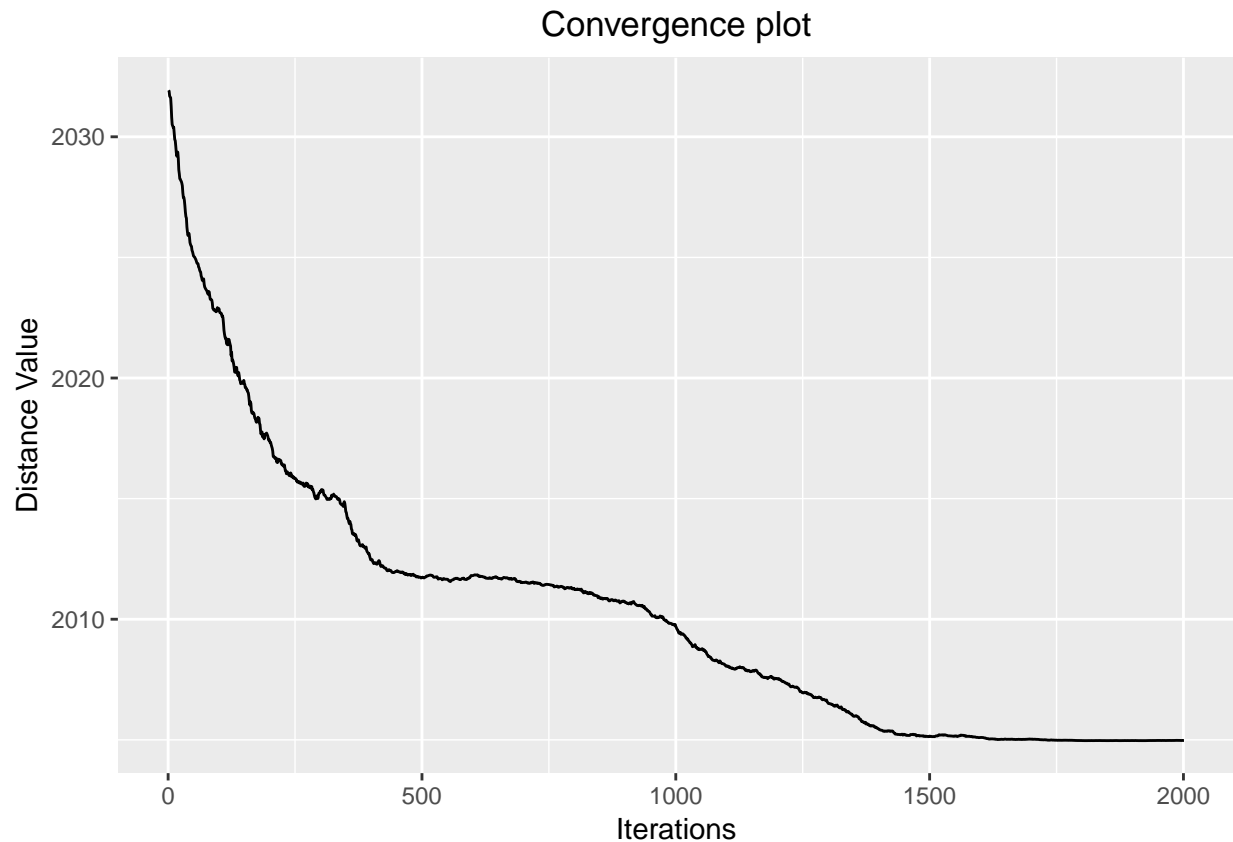
for (i in 1:nrow(random.points)){
  opt_points[i,] <- ris_temp[[i]]$opt_point
  distances[i] <- ris_temp[[i]]$min_dist
}

f <- ris_temp[[which.min(distances)]]$opt_point
f_dist <- ris_temp[[which.min(distances)]]$min_dist
P_dist <- ris_temp[[which.min(distances)]]$dist
msg <- paste0("The minimum is the point of coordinates (",
              round(f[1],2), ",",round(f[2],2),
              ") and the value of the function is ",
              round(f_dist,2))
print(msg)

```

```
## [1] "The minimum is the point of coordinates (310.13,430.93) and the value of the function is 2004.9"
```

```
ggplot(data=data.frame(distance=P_dist)) +
  aes(x=distance) +
  geom_line(aes(x = 1:length(distance),y=distance)) +
  labs(title = "Convergence plot",
        x = "Iterations", y="Distance Value") +
  theme(plot.title = element_text(hjust = 0.5))
```



In the following plot we can see the chaotic movement of the iteration reaching finally the solution where it start to stablize, here the decaying learing rate is helping the method to stablize the solution after a certain number of iterations.

```
# P      <- ris_temp[[which.min(distances)]]$f
# final  <- ris_temp[[which.min(distances)]]$opt_point
# z      <- seq(1,nrow(P),2)
#
# ggif_minimal <- data.frame(x=P[z,1],y=P[z,2],z=z)%>%
#   ggplot(aes(x = x, y = y))+
#   geom_point(col="magenta",alpha=0.5) +
#   labs(title = "Iteration of the point towards the minimum")+
#   geom_point(data=data.frame(x=final[1],y=final[2]),col="blue")+
#   geom_text(data=data.frame(z=z),
#             mapping = aes(x = 306, y = 430,
#                           label = paste0("ITERATION N.  ",z)))+
#   theme(plot.title = element_text(hjust = 0.5))+
#   transition_reveal(z)+
#   ease_aes("linear")+
#   enter_appear()+
```

```
#  exit_fade()  
#  
#  animate(ggif_minimal, fps=90)
```