

# Assignment 4

Pranav Kasela 846965

## Contents

Function definition	1
Table with 20 jobs for 5 machines for the performance checking	3
GA algorithm (from the GA package)	3
SA algorithm (manually implemented)	5
Final Results using 500 jobs with 20 machines.	7
For the Extra point, NEH algorithm for the suboptimal solution	9

```
require(GA)
require(ggplot2)
require(reshape2)
require(dplyr)
require(Rcpp)
```

## Function definition

Initially the functions that were being using to calculate the makespan and fitness were the written in R, but one of the colleagues: *Federico Moiraghi*, was kind enough to provide his code compiled in C++ for the time function to speed up the process. Some parameters are modified in the function before reusing his code.

Since the time is  $> 0$ , the inverse of the makespan can be used as the fitness function for **GA** algorithm, since maximizing the fitness is equivalent to minimizing the makespan.

```
##-- Function before C++
makespan <- function(perm,distMatrix){
  n_jobs      <- ncol(distMatrix)
  n_machines  <- nrow(distMatrix)
  dist        <- matrix(NA, nrow=n_machines, ncol=n_jobs)
  dist[1,]    <- cumsum(distMatrix[1,perm])
  dist[,1]    <- cumsum(distMatrix[,perm[1]])
  for (i in 2:n_machines){
    for (j in 2:n_jobs){
      dist[i,j] <- distMatrix[i,perm[j]] +
        max(dist[i,j-1],dist[i-1,j])
    }
  }
  makespan    <- dist[n_machines,n_jobs]
  return(makespan)
}

fitness <- function(perm,distMatrix){
  return(1/makespan(perm,distMatrix))
}
```

From hereon the time function used will be makespanCcpp, while the fitness code will be fitnessCcpp.

```

## function in C++
cppFunction('double fitnessCpp(NumericVector perm,
                             NumericMatrix distMatrix)
{
    int nrow = distMatrix.nrow();
    int ncol = distMatrix.ncol();
    int norder = perm.size();
    NumericVector order(norder);
    for (int i = 0; i < norder; i++) order[i] = perm[i]-1;
    NumericMatrix time_matrix(nrow, norder);
    time_matrix[0] = distMatrix[nrow * order[0]];
    for (int r = 1; r < nrow; r++)
        time_matrix[r] = time_matrix[r - 1] +
            distMatrix[nrow * order[0] + r];
    for (int c = 1; c < norder; c++)
        time_matrix[nrow * c] = time_matrix[nrow * (c - 1)] +
            distMatrix[nrow * order[c]];
    for (int r = 1; r < nrow; r++)
        for (int c = 1; c < norder; c++)
            if (time_matrix[nrow * c + (r - 1)] > time_matrix[nrow * (c - 1) + r])
                time_matrix[nrow * c + r] = time_matrix[nrow * c + (r - 1)] +
                    distMatrix[nrow * order[c] + r];
    else
        time_matrix[nrow * c + r] = time_matrix[nrow * (c - 1) + r] +
            distMatrix[nrow * order[c] + r];
    return 1/time_matrix[nrow * norder - 1];
}')

cppFunction('double makespanCpp(NumericVector perm,
                               NumericMatrix distMatrix)
{
    int nrow = distMatrix.nrow();
    int ncol = distMatrix.ncol();
    int norder = perm.size();
    NumericVector order(norder);
    for (int i = 0; i < norder; i++) order[i] = perm[i]-1;
    NumericMatrix time_matrix(nrow, norder);
    time_matrix[0] = distMatrix[nrow * order[0]];
    for (int r = 1; r < nrow; r++)
        time_matrix[r] = time_matrix[r - 1] +
            distMatrix[nrow * order[0] + r];
    for (int c = 1; c < norder; c++)
        time_matrix[nrow * c] = time_matrix[nrow * (c - 1)] +
            distMatrix[nrow * order[c]];
    for (int r = 1; r < nrow; r++)
        for (int c = 1; c < norder; c++)
            if (time_matrix[nrow * c + (r - 1)] > time_matrix[nrow * (c - 1) + r])
                time_matrix[nrow * c + r] = time_matrix[nrow * c + (r - 1)] +
                    distMatrix[nrow * order[c] + r];
    else
        time_matrix[nrow * c + r] = time_matrix[nrow * (c - 1) + r] +
            distMatrix[nrow * order[c] + r];
    return time_matrix[nrow * norder - 1];
}

```

```
}')
```

## Table with 20 jobs for 5 machines for the performance checking

The testing of the functions are done with the smallest table available on the website. The algorithm used are the Genetic Algorithm from the R package and the Simulated Annealing that will be implemented manually to check their performance.

### GA algorithm (from the GA package)

```
#This is the pre-test given in the assignment
time_matrix <- matrix(c(29,30,27,2,37,62,21,6,95,59,70,82,
                        85,11,62,80,65,55,67,57),nrow = 4) #used once to test and never again

time_matrix <- as.matrix(read.csv("j20-m5",sep=" ",header = FALSE))
n_jobs <- ncol(time_matrix)

time_taken_GA_20_5 <- microbenchmark::microbenchmark(
  GA.fit <- ga(type = "permutation",
    fitness = fitnessCpp,
    distMatrix = time_matrix,
    lower = 1,
    upper = n_jobs,
    popSize = 600,
    maxiter = 10000,
    run = 300,
    pmutation = 0.2,
    keepBest = TRUE,
    monitor = NULL,
    seed = 1234),
  times = 1
)

summary(GA.fit)

## -- Genetic Algorithm -----
##
## GA settings:
## Type = permutation
## Population size = 600
## Number of generations = 10000
## Elitism = 30
## Crossover probability = 0.8
## Mutation probability = 0.2
##
## GA results:
## Iterations = 359
## Fitness function value = 0.00077101
## Solutions =
##      x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 ... x19 x20
```

```
## [1,] 15 11 9 14 5 7 1 3 19 4 10 20
## [2,] 15 6 11 14 7 17 9 5 8 1 10 20
## [3,] 1 17 15 14 19 9 16 6 11 13 10 20
## [4,] 15 11 9 14 5 8 7 17 1 3 10 20
## [5,] 15 14 19 1 3 9 4 17 5 16 10 20
## [6,] 15 6 11 14 5 8 7 17 1 19 10 20
## [7,] 15 11 9 14 5 7 1 8 6 19 10 20
## [8,] 15 9 14 5 1 6 19 4 7 16 10 20
## [9,] 1 9 4 19 15 17 14 16 5 3 10 20
## [10,] 15 17 14 4 9 3 1 19 5 16 10 20
## ...
## [29,] 15 6 11 9 7 17 1 19 14 5 10 20
## [30,] 15 6 11 17 14 9 1 7 8 3 10 20
```

```
makespanCpp(GA.fit@solution[1,],time_matrix) #best time
```

```
## [1] 1297
```

```
out <- plot(GA.fit, main = "GA progression")
```

```
melt(out[,c(1:3,5)],id.var="iter") %>%
  mutate(inv.value=1/value) -> df1

ggplot(df1, aes(x = iter, y = inv.value,
                group = variable, colour = variable)) +
  xlab("Generation") + ylab("Makespan") +
  geom_line(aes(lty = variable)) +
  scale_colour_brewer(palette = "Set1") +
  labs(title = "GA Progression with 20 jobs in 5 machines")
```



The **GA** algorithm finds the best solution as 1297, with an initial population of 600, a mutation probability of 20% and a crossover probability of 80%. An initial trend of improvement can be seen

### SA algorithm (manually implemented)

The change that has been made the SA algorithm is the swapJobs function, in this case since the permutation will be done on a lot of elements (=number of jobs) - If the number of the jobs is greater than 10, the swap will be done on a random number of elements between 2 and 5; - If the number of jobs is smaller or equal to 10, the swap will be done on 2 elements.

The algorithm, since it's computationally easy to compute, will be done 5 time, to avoid heavy dependencies from it's probabistic nature.

```
swapJobs <- function(perm){
  perm <- as.numeric(perm)
  n <- length(perm)
  if(n>10)
    n_change <- min(sample(2:ceiling(n/5),1),5) #no more than 5 changes at a time
  else
    n_change <- 2
  change <- sort(sample.int(n,n_change))
  newperm <- replace(perm,change,perm[sort(change,decreasing = TRUE)])
  return(as.numeric(newperm))
}

SA <- function(tour, distMatrix, maxIterNoChange = 2000, T_ini = 50, T_min = 1){
  path <- tour
```

```

n <- length(path)
tmin <- T_min      # minimum temperature
alpha <- 0.999     # update factor
T <- T_ini
tini <- T_ini      # starting temperature
dist <- makespanCpp(path, distMatrix)
bestLength <- dist
traceBest <- c(dist)
traceCurrentLength <- c(dist)
iterNoChange = 0
while(T >= tmin){   # if the temperature is not at its minimum
  iterNoChange = iterNoChange+1
  newpath <- swapJobs(path) #swap
  dist_new <- makespanCpp(newpath, distMatrix)
  if(dist_new <= bestLength){
    path <- newpath
    dist <- dist_new
    bestLength <- dist
    iterNoChange <- 0
  }
  else {

    if (exp((dist-dist_new)/T)>runif(1, 0, 1)){
      dist <- dist_new
      path <- newpath
      iterNoChange <- 0
    }

  }
  traceBest <- append(traceBest, bestLength)
  traceCurrentLength <- append(traceCurrentLength, dist)
  T <- T*alpha # the temperature is updated
  if(iterNoChange >= maxIterNoChange){ break}
}
res = list(route=path, traceBest = traceBest, trace = traceCurrentLength)
return(res)
}

start <- as.numeric(sample(1:n_jobs,n_jobs)) #start randomly
best_res <- SA(start, time_matrix, maxIterNoChange = 10000)

for (i in 1:4){
  start <- as.numeric(sample(1:n_jobs,n_jobs)) #start randomly
  res <- SA(start, time_matrix, maxIterNoChange = 10000)
  if (tail(res$traceBest,1) < tail(best_res$traceBest,1))
    best_res <- res
}
tail(best_res$traceBest,1) #best value found

```

```
## [1] 1297
```

When trying with the 100x20 table, the **GA** algorithm finished in approximately 5 minutes, while **SA** algorithm finished in a few seconds giving more or less the same result (GA best makespan was 6557 while

SA best makespan was 6594). The result for the 100x20 table is not provided here for the simplicity of the report.

## Final Results using 500 jobs with 20 machines.

```
time_matrix <- as.matrix(read.csv("j500-m20",sep=" ",header = FALSE))
n_jobs <- ncol(time_matrix)
```

```
time_taken_GA_500_20 <- microbenchmark::microbenchmark(
  GA.fit <- ga(type = "permutation",
    fitness = fitnessCpp,
    distMatrix = time_matrix,
    lower = 1,
    upper = n_jobs,
    popSize = 300,
    maxiter = 10000,
    run = 100,
    pmutation = 0.2,
    keepBest = TRUE,
    monitor = NULL,
    seed = 1234),
  times = 1
)
```

```
makespanCpp(GA.fit@solution[1,],time_matrix) #best time
```

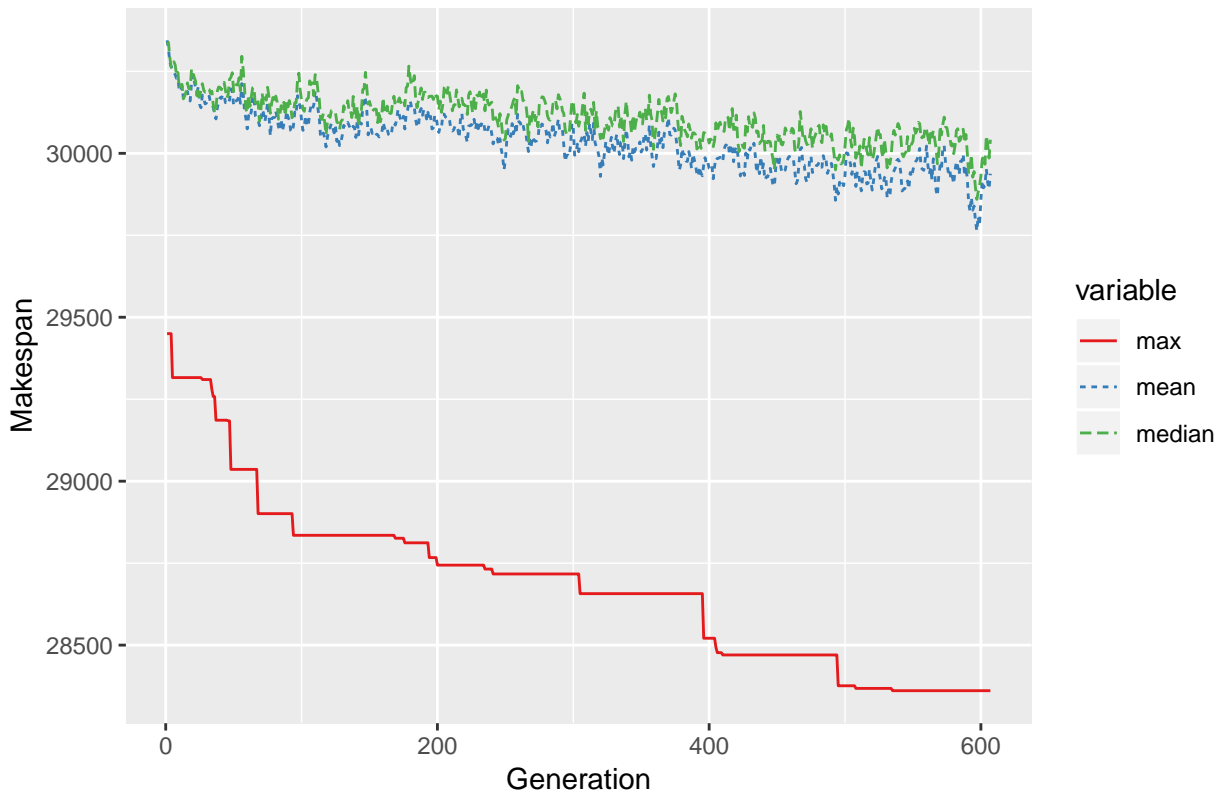
```
## [1] 28361
```

```
out <- plot(GA.fit, main = "GA progression")
```

```
melt(out[,c(1:3,5)],id.var="iter") %>%
  mutate(inv.value=1/value) -> df1

ggplot(df1, aes(x = iter, y = inv.value,
  group = variable, colour = variable)) +
  xlab("Generation") + ylab("Makespan") +
  geom_line(aes(lty = variable)) +
  scale_colour_brewer(palette = "Set1") +
  labs(title = "GA Progression with 500 jobs in 20 machines")
```

## GA Progression with 500 jobs in 20 machines



The **SA** algorithm solution is better than the one given by the **GA** algorithm.

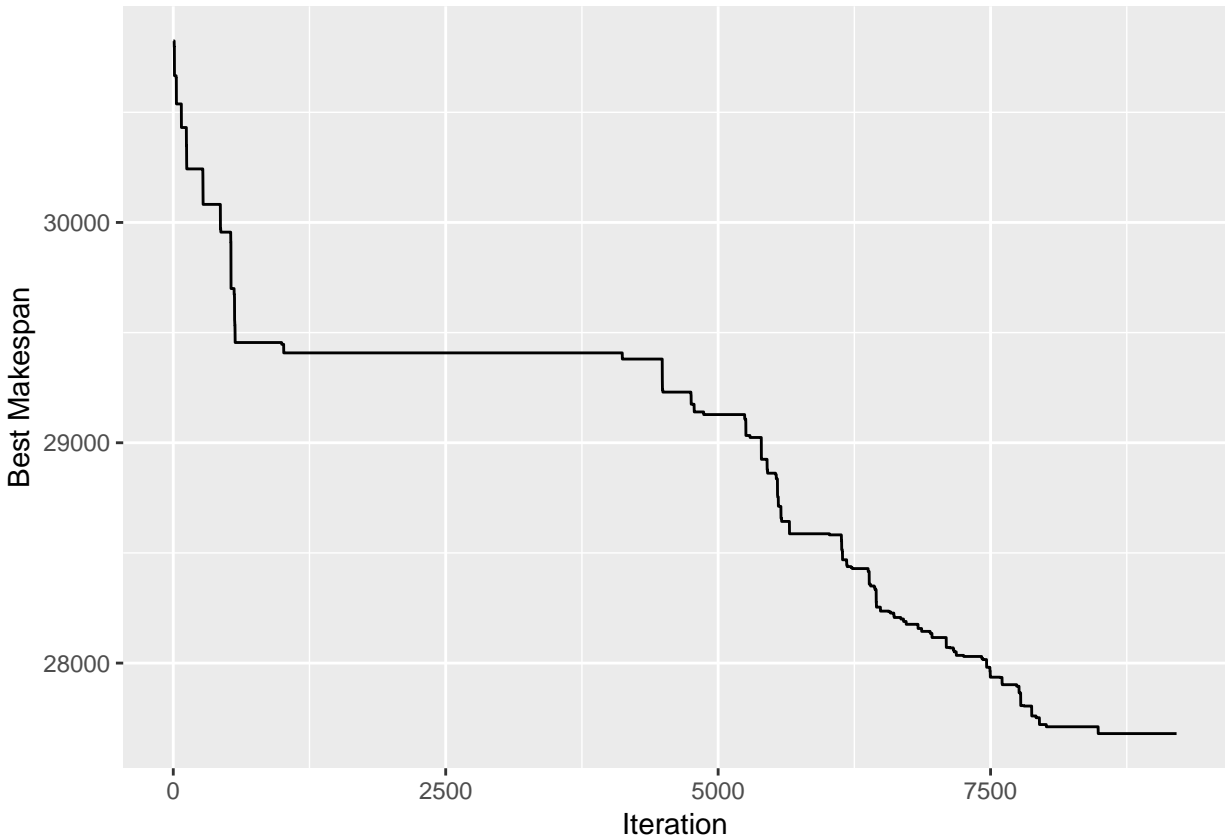
```
start <- as.numeric(sample(1:n_jobs,n_jobs)) #start randomly
best_res <- SA(start, time_matrix, maxIterNoChange = 30000,
               T_ini = 10000,T_min = 1)

for (i in 1:4){
  start <- as.numeric(sample(1:n_jobs,n_jobs)) #start randomly
  res <- SA(start, time_matrix, maxIterNoChange = 30000,
            T_ini = 10000,T_min = 1)
  if (tail(res$traceBest,1) < tail(best_res$traceBest,1))
    best_res <- res
}
tail(best_res$traceBest,1) #best value found

## [1] 27680

ggplot(data = data.frame(iteration=1:length(best_res$traceBest),
                        optimal=best_res$traceBest),
       aes(iteration,optimal)) +
  geom_line() +
  xlab("Iteration") + ylab("Best Makespan") +
  scale_colour_brewer(palette = "Set1")
```





For the Extra point, NEH algorithm for the suboptimal solution

```
insert_at <- function(x,pos,val){
  "inserts at a given position (pos) the value (val) in the array x"
  if (pos==1)
    return(c(val,x))
  len <- length(x)
  if (pos==(len+1))
    return(c(x,val))
  return(c(x[1:pos-1],val,x[pos:len]))
}

NEH <- function(distMatrix,fun.obj,REPORT=0){
  sum <- rbind(colSums(distMatrix),1:ncol(distMatrix))
  sum_order <- as.numeric(sum[2,order(sum[1,])])

  job <- sum_order[1] #the job that takes less time in all machines

  for (i in 2:ncol(distMatrix)){
    temp_ris <- lapply(1:(length(job)+1), function(x) {
      temp_job = insert_at(job,x,sum_order[i])
      t <- fun.obj(1:(length(job)+1),distMatrix[,temp_job])
      return(list(temp_job=temp_job, t=t))
    })
    t <- unlist(lapply(temp_ris, '[[', 't'))
  }
}
```

```

    job <- lapply(temp_ris, '[','temp_job')[[which.min(t)]]
    if(REPORT!=0 && i%%REPORT==0)
      print(paste0("Done ",as.character(i)," jobs"))
  }
  return(list(sol = job,
             value = fun.obj(job,distMatrix) ))
}

```

```

time_taken_NEH_500_20 <- microbenchmark::microbenchmark(
  ris <- NEH(distMatrix = time_matrix, fun.obj = makespanCpp),
  times = 1
)

```

```
ris$value
```

```
## [1] 26936
```

```
#ris$sol
```

```
SA(ris$sol,time_matrix,maxIterNoChange = 10000,T_ini = 10,T_min = 1)
```

```
## $route
```

```

## [1] 485 183 10 475 157 176 326 288 46 169 31 51 166 203 135 494 378
## [18] 207 303 487 104 186 275 362 108 421 355 283 248 100 484 56 201 337
## [35] 96 460 285 29 373 97 36 48 270 358 254 145 429 499 256 81 225
## [52] 418 78 286 87 252 229 220 76 441 305 129 223 146 102 213 74 120
## [69] 344 404 444 15 474 296 356 149 397 68 193 150 500 243 301 142 319
## [86] 371 446 11 384 402 442 455 124 412 114 2 44 63 492 38 329 479
## [103] 461 405 163 302 450 281 14 289 264 438 401 192 65 407 85 433 369
## [120] 247 443 19 291 23 34 440 208 251 13 379 282 178 482 389 478 86
## [137] 352 312 159 25 341 436 315 314 49 180 468 388 128 55 347 240 342
## [154] 354 423 143 255 134 382 457 308 141 377 21 107 428 339 27 198 374
## [171] 189 99 392 349 237 28 413 16 214 453 151 140 219 1 325 411 336
## [188] 470 139 263 161 221 61 451 267 381 332 257 330 8 249 498 121 37
## [205] 304 196 26 18 238 83 57 122 88 398 62 269 268 298 22 424 233
## [222] 113 184 79 119 95 278 393 165 59 473 416 306 245 400 321 408 493
## [239] 331 241 80 299 24 297 7 177 41 432 236 365 70 370 462 217 110
## [256] 162 447 153 191 181 160 84 215 335 276 338 230 73 272 258 224 360
## [273] 477 53 350 261 399 190 194 295 60 259 91 480 395 188 168 116 242
## [290] 5 202 340 287 167 131 52 394 72 483 222 496 469 422 58 144 231
## [307] 410 154 476 33 174 92 437 69 324 127 449 328 182 123 253 50 372
## [324] 171 172 75 45 216 94 101 40 458 246 218 490 67 4 106 434 126
## [341] 132 197 467 20 497 47 456 152 205 125 227 173 368 293 148 98 232
## [358] 310 206 234 54 376 179 495 415 385 431 380 472 117 130 435 327 486
## [375] 42 452 417 320 226 345 489 481 459 147 409 235 209 109 204 9 71
## [392] 274 364 396 280 105 3 32 158 318 390 322 195 77 187 448 406 307
## [409] 39 244 112 133 309 425 491 351 391 439 115 323 403 212 313 427 317
## [426] 43 387 239 156 90 164 353 12 316 359 292 170 488 200 137 273 185
## [443] 386 279 346 284 35 17 471 414 260 111 375 210 175 30 445 419 290
## [460] 383 366 6 228 265 426 334 250 311 64 211 82 89 277 138 361 136
## [477] 199 93 155 103 66 343 271 367 266 357 454 420 463 333 348 430 118
## [494] 466 363 294 465 464 262 300

```

```
##
```

```
## $traceBest
```

```
## [1] 26936 26936 26936 26936 26936 26936 26936 26936 26936 26936 26936
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]





[illegible]