Advanced Algorithmic Techniques (COMP523)

Approximation Algorithms

Recap and plan

- Previously on COMP523:
 - We designed polynomial-time algorithms for several problems.
 - We saw that some problems do not have polynomial time algorithms (NP-hard problems).
- Next 4 lectures:
 - Approximation algorithms.
- This lecture:
 - Greedy approximation algorithms.
 - Load balancing on identical machines.

• For some problems (e.g., knapsack), we do not expect to find polynomial time algorithms.

- For some problems (e.g., knapsack), we do not expect to find polynomial time algorithms.
- How should we approach these problems?

- For some problems (e.g., knapsack), we do not expect to find polynomial time algorithms.
- How should we approach these problems?
- We can design approximation algorithms, which

- For some problems (e.g., knapsack), we do not expect to find polynomial time algorithms.
- How should we approach these problems?
- We can design approximation algorithms, which
 - Run in polynomial time.

- For some problems (e.g., knapsack), we do not expect to find polynomial time algorithms.
- How should we approach these problems?
- We can design approximation algorithms, which
 - Run in polynomial time.
 - Compute a solution that is "close" to the optimal.

 What does "close" to the optimal mean? How do we measure that?

- What does "close" to the optimal mean? How do we measure that?
- How do we make such an argument, if we cannot really find the optimal?

- What does "close" to the optimal mean? How do we measure that?
- How do we make such an argument, if we cannot really find the optimal?
- How do we know if our algorithm is the best possible? Can we get "closer" to the optimal?

Methods for approximation algorithms

- Greedy algorithms.
- Pricing method (also known as the Primal-Dual method).
- Linear Programming and Rounding.
- Dynamic Programming on rounded inputs.

Application: Load Balancing

- We have a set of m *identical* machines M_1, \ldots, M_m
- We have a set of n jobs, with job j having processing time tj.
- We want to assign every job to some machine.
- Let A(i) be the set of jobs assigned to machine i.
- The load of machine i is $T_i = \sum_{j \in A(i)} t_j$
- The goal is to minimise the makespan, i.e.,

 $T = max_i T_i$



 M_1

 M_2

M₃



































2

jobs

6

2

 M_1







jobs



jobs

makespan = 8

Load Balancing

- The load balancing problem on identical machines is NP-hard.
- We will design greedy approximation algorithms for it.

Greedy algorithm

Greedy algorithm

- Pick any job.
- Assign it to the machine with the smallest load so far.
- Remove it from the pile of jobs.

Greedy algorithm

- Pick any job.
- Assign it to the machine with the smallest load so far.
- Remove it from the pile of jobs.

```
Algorithm Greedy-Balance
```

```
Start with no jobs assigned

Set T_i = 0 and A(i) = \emptyset for all machines M_i

For j = 1, ..., n

Let M_i be the machine that achieves the minimum min<sub>k</sub> T_k

Assign job j to machine M_i

Set A(i) = A(i) \cup \{j\}

Set T_i = T_i + t_j

EndFor
```



 M_1

 M_2

M₃



































2

jobs

6

2

 M_1







jobs



jobs

makespan = 8



jobs

makespan = 8 A makespan of 7 is possible

Notation

- Let T be the makespan achieved by Greedy-Balance.
- Let T* be the optimal makespan.
• Challenge: We don't know T*! How are we supposed to argue about it?

- Challenge: We don't know T*! How are we supposed to argue about it?
 - We want to prove that **T** is not far from **T***.

- Challenge: We don't know T*! How are we supposed to argue about it?
 - We want to prove that **T** is not far from **T***.
 - We will show that T is not far from something which is smaller than T*.

- Challenge: We don't know T*! How are we supposed to argue about it?
 - We want to prove that **T** is not far from T*.
 - We will show that T is not far from *something which is smaller than* T*.
 - Then it is certainly not far from T*.

- Challenge: We don't know T*! How are we supposed to argue about it?
 - We want to prove that **T** is not far from T*.
 - We will show that T is not far from *something which is smaller than* T*.
 - Then it is certainly not far from T*.
- Fundamental technique in approximation algorithms analysis:

- Challenge: We don't know T*! How are we supposed to argue about it?
 - We want to prove that **T** is not far from T*.
 - We will show that T is not far from *something which is smaller than* T*.
 - Then it is certainly not far from T*.
- Fundamental technique in approximation algorithms analysis:
 - Bounding the optimal from below (for minimisation problems) and from above (for maximisation problems).

lower bound on T* optimal makespan T* algorithm makespan T





• What is a bound that we can use for the optimal?

- What is a bound that we can use for the optimal?
- Consider the total processing time of all the jobs (the sum of the processing times t_j).

- What is a bound that we can use for the optimal?
- Consider the total processing time of all the jobs (the sum of the processing times t_j).
- One of the m machines must be allocated at least an 1/m fraction of the total work.

- What is a bound that we can use for the optimal?
- Consider the total processing time of all the jobs (the sum of the processing times t_j).
- One of the m machines must be allocated at least an 1/m fraction of the total work.
- We have that:

$$T^* \ge \frac{1}{m} \sum_{j=1}^n t_j$$

Taking the average sum of processing times assumes that this job can be split.

Taking the average sum of processingtimes assumes that this job can be split.

In reality, OPT will assign this job to some machine and the makespan will be its processing time.

Taking the average sum of processingtimes assumes that this job can be split.

In reality, OPT will assign this job to some machine and the makespan will be its processing time.

Our bound assumes that OPT is approximately m times better than it is.

The bound can be good in situation where jobs have fairly similar processing times.

Taking the average sum of processingtimes assumes that this job can be split.

In reality, OPT will assign this job to some machine and the makespan will be its processing time.

Our bound assumes that OPT is approximately m times better than it is.

• Can we think of another bound?

- Can we think of another bound?
- Every job must be scheduled to some machine.

- Can we think of another bound?
- Every job must be scheduled to some machine.
- The makespan is certainly at least the largest processing time t_j of any job.

- Can we think of another bound?
- Every job must be scheduled to some machine.
- The makespan is certainly at least the largest processing time t_j of any job.
- We have that:

$$T^* \ge \max_j t_j$$

In this example, this is a good bound as the maximum processing time is very large.

In this example, this is a good bound as the maximum processing time is very large.

In other cases, it might not be such a good bound.

In this example, this is a good bound as the maximum processing time is very large.

In other cases, it might not be such a good bound.

But we will actually use both bounds!

• Two lower bounds:





The performance of Greedy-Balance

• Theorem: Algorithm Greedy-Balance produces an assignment of jobs to machines with makespan $T \le 2T^*$.

 Consider a job j that was assigned to machine M_i by Greedy-Balance.

- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.

- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine j was $T_i t_j$.

- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine j was T_i t_j.
 - This was the smallest load among all machines (why?)


 M_{i}





- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine **j** was **T**_i **t**_j.
 - This was the smallest load among all machines (why?)

- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine **j** was **T**_i **t**_j.
 - This was the smallest load among all machines (why?)
 - Every other machine has load at least T_i t_j.

Every other machine has load at least T_i - t_j.

- Every other machine has load at least T_i t_j.
- Summing up over all machines we get:

$$\sum_{k} T_{k} \ge m(T_{i} - t_{j}) \Rightarrow T_{i} - t_{j} \le \frac{1}{m} \sum_{k} T_{k}$$

Lower bounding the optimal

• Two lower bounds:





- Every other machine has load at least T_i t_j.
- Summing up over all machines we get:

$$\sum_{k} T_{k} \ge m(T_{i} - t_{j}) \Rightarrow T_{i} - t_{j} \le \frac{1}{m} \sum_{k} T_{k}$$

 $T_i - t_j \le T^*$ (first lower bound)

- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine **j** was **T**_i **t**_j.
 - This was the smallest load among all machines (why?)
 - Every other machine has load at least T_i t_j.

- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine **j** was T_i t_j.
 - This was before we added the job.
 - After we add the job, the load is $T_i t_j + t_{j.}$

Lower bounding the optimal

• Two lower bounds:





- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine j was $T_i t_j$.
 - This was before we added the job.
 - After we add the *final* job, the load is $T_i t_j + t_{j.}$
 - Obviously $t_j \le \max_k t_k \le T^*$

- Every other machine has load at least T_i t_j.
- Summing up over all machines we get:

$$\sum_{k} T_{k} \ge m(T_{i} - t_{j}) \Rightarrow T_{i} - t_{j} \le \frac{1}{m} \sum_{k} T_{k}$$

 $T_i - t_j \le T^*$ (first lower bound)

$$T_i - t_j \le T^*$$
 (first lower bound)

 $t_j \leq T^*$ (second lower bound)

$$T_i - t_j \le T^*$$
 (first lower bound)

 $t_j \leq T^*$ (second lower bound)

$$T_i \leq 2T^*$$

$$T_i - t_j \le T^*$$
 (first lower bound)

 $t_j \leq T^*$ (second lower bound)

$$T_i \leq 2T^*$$

 $T \leq 2T^*$ (since *j* was the final job)

• We have shown that the makespan of the solution of Greedy-Balance is at most a 2 factor away from the optimal makespan.

- We have shown that the makespan of the solution of Greedy-Balance is at most a 2 factor away from the optimal makespan.
- Can we show that it is also at least a 2 factor away in the worst case?

- We have shown that the makespan of the solution of Greedy-Balance is at most a 2 factor away from the optimal makespan.
- Can we show that it is also at least a 2 factor away in the worst case?
 - In other words, is there an example (an *instance*) of the load balancing problem for which the algorithm actually produces a makespan which is *twice as much* as the optimal makespan?

- We have shown that the makespan of the solution of Greedy-Balance is at most a 2 factor away from the optimal makespan.
- Can we show that it is also at least a 2 factor away in the worst case?
 - In other words, is there an example (an *instance*) of the load balancing problem for which the algorithm actually produces a makespan which is *twice as much* as the optimal makespan?
 - In other words, is our analysis of the algorithm *tight*?

Tight example for Greedy-Balance



Tight example for Greedy-Balance



Tight example for Greedy-Balance



Greedy-Balance assigns *m-1* "small" jobs to each machine and then finally assigns the "large" job to one machine.

Makespan: 2m-1

The optimal assigns the "large" job to one machine, and evenly spread the "small" jobs over the remaining *m*-1 machines.

Makespan: m

 M_1 M_2 ... M_m

- Consider a minimisation problem P and an objective obj.
 - Here: Load Balancing on identical machines and makespan.
 - Consider an approximation algorithm A.
 - Consider an input x to the problem P.
 - Let obj(A(x)) be the value of the objective from the solution of A on x.
 - Let opt(x) be the minimum possible value of the objective on x.

• The approximation ratio of A is defined as

max_x obj(A(x)) / opt(x)

 i.e., the worst case ratio of the objective achieved by the algorithm over the optimal value of the objective, over all possible inputs to the problem.

All inputs to the problem P

Algorithm A

Optimal









All inputs to the problem P

Algorithm A

Optimal









All inputs to the problem P

Algorithm A

Optimal

Ratio on Input A Ratio on Input B


Ratio on Input A Ratio on Input B









- That means that:
 - In order to prove an upper bound on the approximation ratio, we have to somehow argue about *all* inputs to the problem.
 - In order to prove a lower bound on the approximation ratio, we have to argue about one input to the problem.

• For maximisation problems, we define

max_x opt(x) / obj(A(x))

- i.e., the worst case ratio of the optimal value of the objective over the value of the objective achieved by the algorithm, over all possible inputs to the problem.
- Convention, to have approximation ratios always be ≥ 1 .

- What does "close" to the optimal mean? How do we measure that?
- How do we make such an argument, if we cannot really find the optimal?
- How do we know if our algorithm is the best possible? Can we get "closer" to the optimal?

- What does "close" to the optimal mean? How do we measure that? Approximation ratio.
- How do we make such an argument, if we cannot really find the optimal?
- How do we know if our algorithm is the best possible? Can we get "closer" to the optimal?

- What does "close" to the optimal mean? How do we measure that? Approximation ratio.
- How do we make such an argument, if we cannot really find the optimal? We lower or upper bound the optimal.
- How do we know if our algorithm is the best possible? Can we get "closer" to the optimal?

A better greedy algorithm for load balancing

- Greedy-Balanced was:
 - Pick any job.
 - Assign it to the machine with the smallest load so far.
 - Remove it from the pile of jobs.

A better greedy algorithm for load balancing

- Greedy-Balanced was:
 - Pick any job.
 - Assign it to the machine with the smallest load so far.
 - Remove it from the pile of jobs.

We did not really take into account the order in which we consider the jobs.

A better greedy algorithm for load balancing

- Sorted-Balance:
 - Sort the jobs in non-increasing order of processing times.
 - Pick a job according to this order.
 - Assign it to the machine with the smallest load so far.
 - Remove it from the pile of jobs.

• Assume that we have at most m jobs.

- Assume that we have at most m jobs.
- What is the approximation ratio of **Sorted-Balance**?

- Assume that we have at most m jobs.
- What is the approximation ratio of **Sorted-Balance**?
 - Each job goes to a different machine.

- Assume that we have at most m jobs.
- What is the approximation ratio of **Sorted-Balance**?
 - Each job goes to a different machine.
 - **Sorted-Balance** produces an optimal allocation.

- Assume that we have at most m jobs.
- What is the approximation ratio of **Sorted-Balance**?
 - Each job goes to a different machine.
 - **Sorted-Balance** produces an optimal allocation.
 - The same was actually true for **Greedy-Balance**.

• Assume that we have more than m jobs.

- Assume that we have more than m jobs.
- Then, it holds that $T^* \leq 2t_{m+1}$

- Assume that we have more than m jobs.
- Then, it holds that $T^* \leq 2t_{m+1}$
 - Consider the first m+1 jobs in sorted order.

- Assume that we have more than m jobs.
- Then, it holds that $T^* \leq 2t_{m+1}$
 - Consider the first m+1 jobs in sorted order.
 - Each one of them takes at least t_{m+1} time.

- Assume that we have more than m jobs.
- Then, it holds that $T^* \leq 2t_{m+1}$
 - Consider the first m+1 jobs in sorted order.
 - Each one of them takes at least t_{m+1} time.
 - Since there are m machines, there must be one machine that receives at least two of these jobs.

- Assume that we have more than m jobs.
- Then, it holds that $T^* \leq 2t_{m+1}$
 - Consider the first m+1 jobs in sorted order.
 - Each one of them takes at least t_{m+1} time.
 - Since there are m machines, there must be one machine that receives at least two of these jobs.
 - The load on this machine will be at least $2t_{m+1}$.

Lower bounding the optimal

• Two lower bounds:





Lower bounding the optimal

Three lower bounds:

 $T^* \ge \frac{1}{m} \sum_{j=1}^n t_j \qquad T^* \ge \max_j t_j$



 $T^* \ge 2t_{m+1}$

The performance of Sorted-Balance

• Theorem: Algorithm Sorted-Balance produces an assignment of jobs to machines with makespan $T \le (3/2)T^*$.

- Let M_i be the machine with the maximum load according to the assignment of Sorted-Balance.
- If M_i is assigned a single job, the outcome is optimal.
- Assume M_i that is assigned at least two jobs and let j be the last job assigned to the machine.
 - Note that $j \ge m+1$
 - Therefore, $t_j \le t_{m+1} \le (1/2)T^*$

The proof (still the same argument)

- Every other machine has load at least T_i t_j.
- Summing up over all machines we get:

$$\sum_{k} T_{k} \ge m(T_{i} - t_{j}) \Rightarrow T_{i} - t_{j} \le \frac{1}{m} \sum_{k} T_{k}$$

$$T_i - t_j \le T^*$$
 (first lower bound)

The proof (previous argument)

- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine j was $T_i t_j$.
 - This was before we added the job.
 - After we add the *final* job, the load is $T_i t_j + t_j$.
 - Obviously $t_j \le \max_k t_k \le T^*$

The proof (new argument)

- Consider a job j that was assigned to machine M_i by Greedy-Balance.
- Consider the time when this assignment took place.
 - The load of machine j was T_i t_j.
 - This was before we added the job.
 - After we add the *final* job, the load is $T_i t_j + t_j$.
 - We established that $t_j \le t_{m+1} \le (1/2)T^*$

$$T_i - t_j \le T^*$$
 (first lower bound)

$$t_j \leq \frac{1}{2}T^*$$
 (third lower bound)

$$T_i - t_j \le T^*$$
 (first lower bound)

$$t_j \leq \frac{1}{2}T^*$$
 (third lower bound)

$$T_i \le \frac{3}{2}T^*$$

$$T_i - t_j \le T^*$$
 (first lower bound)

$$t_j \leq \frac{1}{2}T^*$$
 (third lower bound)

$$T_i \leq \frac{3}{2}T^*$$
$$T \leq \frac{3}{2}T^*$$

- What does "close" to the optimal mean? How do we measure that? Approximation ratio.
- How do we make such an argument, if we cannot really find the optimal? We lower or upper bound the optimal.
- How do we know if our algorithm is the best possible? Can we get "closer" to the optimal?
• The Sorted-Balance algorithm actually gives a 4/3 approximation ratio, with a better analysis.

- The Sorted-Balance algorithm actually gives a 4/3 approximation ratio, with a better analysis.
- For the load balancing problem on identical machines, there is a Polynomial Time Approximation Scheme (PTAS).

- The Sorted-Balance algorithm actually gives a 4/3 approximation ratio, with a better analysis.
- For the load balancing problem on identical machines, there is a Polynomial Time Approximation Scheme (PTAS).
 - An algorithm which, given an input and a constant parameter ε, runs in polynomial time and produces an outcome which is (1+ε) far from the optimal.

• Generally:

- Generally:
 - A PTAS (or an FPTAS, more about that later) is the best approximation we can hope for, for an NP-hard problem.

- Generally:
 - A PTAS (or an FPTAS, more about that later) is the best approximation we can hope for, for an NP-hard problem.
 - Sometimes it is impossible to get that close.

- Generally:
 - A PTAS (or an FPTAS, more about that later) is the best approximation we can hope for, for an NP-hard problem.
 - Sometimes it is impossible to get that close.
 - Inapproximability **α** of problem P:

- Generally:
 - A PTAS (or an FPTAS, more about that later) is the best approximation we can hope for, for an NP-hard problem.
 - Sometimes it is impossible to get that close.
 - Inapproximability α of problem P:
 - There is no polynomial time algorithm that achieves an approximation ratio better than α.