27 Load Balancing, Center Selection

- Approximation Algorithms
  - Load balancing
  - Center selection

27.1 Load Balancing

Lemma 27.1
The optimal value for the load balancing problem is at least $\max\{\frac{1}{m}\sum_i t_i, \max_i t_i\}$.

Proof. (Proof of the theorem)
Let machine J achieve the max load.
Let job i be the last job assigned to machine j.

Let T_j be the final load on machine j.
Before job i was assigned, the load on machine j was $T_j - t_i$, which was the smallest load on any machine.

So the total load is $\sum_k T_k \geq m(T_j - t_i)$.
$\implies T_j - t_i \leq \frac{1}{m}\sum_k T_k = \frac{1}{m}\sum_l t_l$

So the optimal value T^* satisfies $T^* \geq T_j - t_i$.

But machine j also has job i, so

$$T_j = (T_j - t_i) + t_i \leq T^* + T^* \leq 2T^*$$

Where the first inequality is achieved from the above lemma. □

We can do better by sorting the jobs and considering them in order from longest to shortest.

Lemma 27.2
If there are more than m jobs, then $T^* \geq 2t_{m+1}$

(Here, we order the jobs so $t_1 \geq t_2 \geq \cdots \geq t_n$)

Proof. Some machine gets 2 jobs, each of which have duration at least t_{m+1}. □

Proof. (Proof that we get approximation ratio $\frac{3}{2}$)
As before, suppose machine j gets the max load. If it only has one job, the schedule is optimal.
If it has at least 2 jobs, and t_i is the last assigned, then $t_i \leq t_{m+1} \leq \frac{T^*}{2}$ (by lemma)

But as in the proof of the previous theorem,

$$T_j \leq (T_j - t_i) + t_i \leq T^* + \frac{T^*}{2} \leq \frac{3}{2}T^*$$

□

27.2 Center Selection

Given n sites (specified by their coordinates), find k centers so that any site is within distance r of some center, with r as small as possible.

Definition 27.3
The **covering radius** of a set of centers C for a set of sites S is the smallest r such that $\forall s \in S, \exists c \in C$ with $\text{dist}(s, c) \leq r$.

Possible greedy algorithm: repeatedly add the center that minimizes the covering radius.

In the case that we have 2 sites and 2 centers, the greedy algorithm would first put a center in the halfway between the 2 sites. After that, no matter where you put the second center, the covering radius can not decrease. But, there was a way to place the centers to make the radius 0!
So, this means the greedy approach does not get any finite approximation ratio.

For simplicity, suppose we know the optimal radius. Say there exists a set of centers C^* with radius r. So for any site s, there is a center $c \in C^*$ within distance r of s.
Then everything within distance r of c is within $2r$ of s.
So this can give us a $2r$-cover.

```
GreedyCover(S, r)
    let S' = S (sites yet to be covered)
    let C = null (centers)
    while S' != null
        choose any s in S'
        add s to C
        delete all sites from S' within distance 2r of s
    endwhile
    return C
```

We claim that if this algorithm returns a cover C with $|C| > k$, then no set of k centers can have covering radius $\leq r$.

Proof. Suppose (for contradiction) there exists some C^* with $|C^*| = k$ and covering radius at most r. Since C (output by the greedy algorithm) satisfies $C \subseteq S$, there must be a $c^* \in C^*$ within distance r of any $c \in C$.

Claim: No $c^* \in C^*$ can be within distance r of two centers in C.

Assuming the claim, each $c \in C$ has a unique $c^* \in C^*$, so $|C^*| \geq |C| > k$, a contradiction.

To see the claim: Suppose there is a $c, c' \in C$ with $\text{dist}(c, c^*) \leq r$, and $\text{dist}(c', c^*) \leq r$.
But then $\text{dist}(c, c^*) + \text{dist}(c^*, c') \geq \text{dist}(c, c') > 2r$
Which contradicts the triangle inequality. □