CSC2421: Online and other myopic algorithms
Spring 2021

Allan Borodin

January 14, 2021

Week 1

Course Organization

- This will be a reading course and will be run so that it is appropriate as a foundational course for anyone who has the equivalent of our undergraduate algorithms course CSC373.
- In addition, depending on what topics or papers students undertake, the course could also be appropriate for anyone interested in undertaking research relating to the scope of this course.
  - **Sources**: The main source is a text (with the same title as the course title) that I am now writing with Denis Pankratov.
  - In addition, there are course notes at other Universities, slides for this course and my previous courses, various textbooks, and relatively current research papers. See links posted on the course web page.
  - The precise reading requirements will depend on how many people are in the course (either enrolled or actively audting),

More organization

- This is a very active field. I will attempt give a sense of the different aspects of this field (according to the current planned Chapters) and then we will begin to have student presentations.
- I am thinking that every student (whether auditing or taking the course for credit) will give two presentations, perhaps an intiial overview of some topic, and then later a more detailed discussion of one or more specific papers in the chosen topic.
- It seems clear that we will be meeting online. The class will now meet Thursdays 4-6 to avoid the conflict wtih Nisarg Shah's Thursday 1-3 class. Rather than regularly scheduled office hours, I am happy to meet individually (say by zoom, or skype) whenever anyone wants to meet.
- There is also a piazza page which I encourage as a good way to share comments and questions. I particularly welcome students answering questions.
- My contact information : bor@cs.toronto.edu The course web page is www.cs.toronto.edu/~bor/2421s21

About the title and the course

An online algorithm receives its input as a sequence of *input items*. When a new input item arrives, the algorithm needs to make a decision about the input item. We think of the inputs arriving in discrete steps; that is, input item I_j arrives at step j. We may or may not also have a real time t associated with each input item.

When considering online algorithms, we use the term *competitive ratio* rather than approximation ratio. Informally, this is a "worst case" (over all possible input sequences) ratio for the performance (e.g, profit or cost) for an algorithm's solution relative to an optimal solution.

See Chapter 1 for a brief history and a couple of examples of online algorithms.

In Chapter 2, we formalize *deterministic online algorithms* in terms of *request answer games*. We then formalize the competitive ratio of a deterministic online algorithm.

Some variants of online algorithms

There are many variants of online algorithms. (We'll return to this when we consider a specific problem: Makespan.)

- The standard meaning of "online algorithm" within TCS is that the decision for each input item is *irrevocable* and must be made without knowledge of future input items. However, in some models, there can be limited ways in which decisions can be revoked,
- In the initial works on *competitive analysis*, the input item sequence can be completely *adversarial*. One variant is to assume that each item is drawn from some distribution where these distributions can be known or unknown in advance.
- Another example is when an adversary provides an input set but the input sequence is a random permutation of the items in the input set.
- Whenever there is randomness in the algorithm or randomness in the input sequence, we need to consider the expected performance; that is, the performance of an algorithm is a random variable and we consider its expected value. See Chapter 3 for the definition of a randomized online algorithm and its competitive ratio.

Myopic algorithms

I am using the term *myopic algorithm* to indiicate that an algorithm can have some ability to look at the input set but there is still information that is not yet revealed.
We consider every online algorithm to be a myopic algorithm. In addition:

- *Online algorithms with advice* refers to an online algorithm that can first obtain a limited amount of information about the input set. See Chapter 10.
- In the literature of scheduling algorithms, the term online algorithm often refers to what I would call a "real time algorithm" where decisions about an input item arriving at *time t* can be made at any time $t' \geq t$. See Chapter 13.
- In *probing algorithms*, some information about each input is revealed upon arrival but other information about the item is obtained by probing within say some budget. See Chapter 16.
- We consider greedy algorithms and some other one pass algorithms to be myopic. This is formalized in Chapter 17.

Why study online and other myopic algorithms

In many applications, there is an inherent online requirement. For example, *paging* (also called *caching*) algorithms have to decide what page to evict when the cache (or some level of the memory hierarchy) is full.

Some auctions need to be run in an online fashion. One very important example is online advertising (i.e. selling ads deriving from online search to advertisers).

In other applications it may not be a requirement but rather a desireable property. Moreover, online and other myopic algorithms tend to be conceptually simple and very efficient.

Even when online algorithms do not provide solutions that are "competitive" in performance with more complex algorithms, they can serve as a benchmark or initial solution when an initial solution is needed quickly.

The difference between online/myopic algorithm analsyis and more standard algorithm analysis as in approximation algorithms and complexity theory.

The seminal work of Cook-Karp-Levin on *NP*-completeness strongly suggests that there are many related problems that cannot be solved efficiently (i.e., in polynomial time). But this is still a conjecture, an almost universally believed conjecture but still a conjecture.

Although we may strongly believe $P \neq NP$, we do not have explicit problems f in *NP* for which we can *prove* that f is not computable in linear time for a sufficiently general mlodel of computation.

In contrast, in the analysis of online and other myopic algorithms we restrict the class of algorithms and prove negative results without any complexity (or even computability) assumptions. We use what are called *information theoretic* arguments to establish negative results. That is, because the algorithm is working with incomplete information, there are limitations to what such an algorithm can do.

Comments and disclaimers on the course perspective

- I consider this course to be a "foundational graduate course" even though online and other myopic algorithms are a small part of the field of *algorithm design and analysis*.
- Although the topic seems focused, the analysis of these algorithms often introduces analysis methods used more generally in algorithmic analysis.
- Sushant Sachdeva is the instructor for a different section of CSC2421 "Topics in Algorithms: Graphs, Matrices and Optimization". The two course have some overlap but mainly they they are covering different topics. Most graduate algorithms courses are biased towards some research perspective. Given that CS might be considered (to some extent) The Science and Engineering of Algorithms, one cannot expect any comprehensive introduction to algorithm design and analysis in any course.

Rest of todays agenda

Just to motivate the course a little more I want to skim through what the different Chapters are about and then skim through some of the results in Chapters 1-3. I have posted the curent draft of the index and Chapters 1-3 on the web page.

For later Chapters, I may post them on BBCollaborate so that they are restricted to the class. So if you are interested in participating (even if auditing), I would lke to add you to the list of students on BbCollaborate. The easiest thing would be for you to audit the course. If you are on the course Quercus list you will get all course annoucements. If at any time you would no longer like to be on the list, I can delete you from the list so that you will not be bothered by these emails. We all get more than enough email.

So... before we begin, are there any questions or comments?

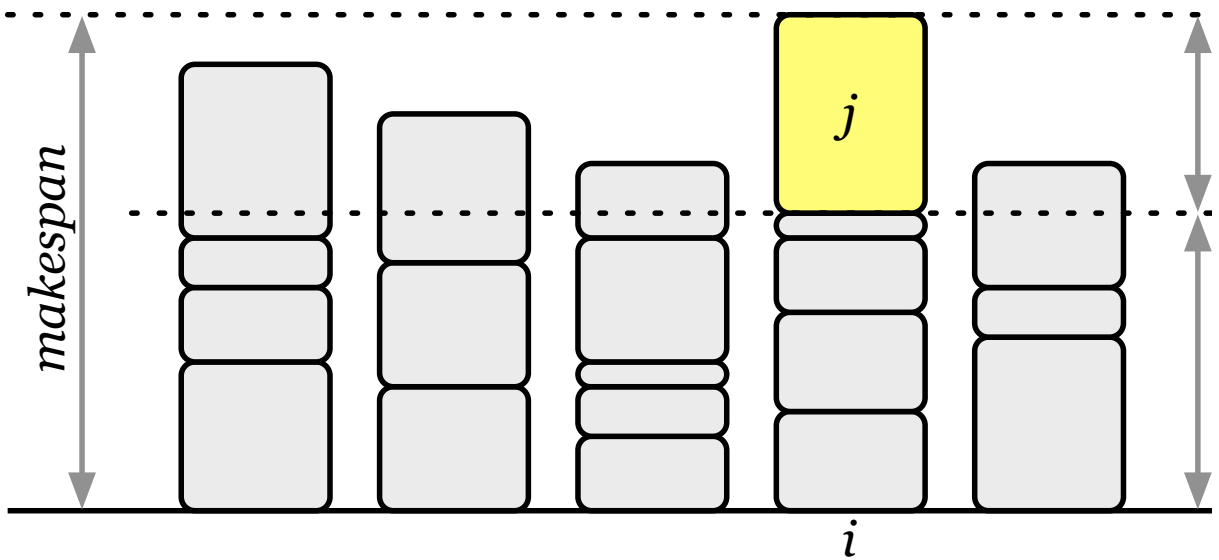
We start with a little history and the makespan problem.

Greedy and online algorithms: Graham's *online* and LPT makespan algorithms

- Let's start with these two greedy algorithms (one online and one "semi-online") that date back to 1966 and 1969 papers Graham.
- These are good starting points since (preceding NP-completeness) Graham conjectured that these are hard (requiring exponential time) problems to compute optimally but for which there were worst case approximation ratios (although he didn't use that terminology).
- This might then be called the start of worst case approximation algorithms. One could also even consider this to be the start of online algorithms and competitive analysis (although one usually refers to a 1985 paper by Sleator and Tarjan as the seminal paper in this regard). As pointed out in Chapter 1, there are other works that precede even the Graham paper.
- There are some general concepts to be observed in Graham's work and (even after more than 50 years) still open questions concerning the many variants of makespan problems.

The makespan problem for identical machines

- The input consists of n jobs $\mathcal{J} = J_1 \ldots, J_n$ that are to be scheduled on m identical machines.
- Each job J_k is described by a processing time (or load) p_k.
- The goal is to minimize the latest finishing time (maximum load) over all machines.
- That is, the goal is a mapping $\sigma : \{1, \ldots, n\} \to \{1, \ldots, m\}$ that minimizes $\max_k \left(\sum_{\ell:\sigma(\ell)=k} p_\ell \right)$.



[picture taken from Jeff Erickson's lecture notes]

Redux: The Many Variants of Online Algorithms

As I indicated, Graham's algorithm could be viewed as the first example of what has become known as *competitive analysis* (as named in a paper by Manasse, McGeoch and Sleator) following the paper by Sleator and Tarjan which explicitly advocated for this type of analysis. Another early (pre Sleator and Tarjan) example of such analysis was Yao's analysis of online bin packing algorithms.

In competitive analysis we compare the performance of an online algorithm against that of an optimal solution. The meaning of *online algorithm* here is that input items arrive sequentially and the algorithm must make an irrevocable decision concerning each item. (For makespan, an item is a job and the decision is to choose a machine on which the item is scheduled.)

But what determines the order of input item arrivals?

The Many Variants of Online Algorithms continued

- In the "standard" meaning of online algorithms (for CS theory), we think of an adversary as creating a nemesis input set and the ordering of the input items in that set. So this is traditional worst case analysis as in approximation algorithms applied to online algorithms. If not otherwise stated, we will assume this as the meaning of an online algorithm and if we need to be more precise we can say *online adversarial model*.
- We will also sometimes consider an *online stochastic model* where an adversary defines an input distribution and then input items are sequentially generated. There can be more general stochastic models (e.g., a Markov process) but the i.d. and i.i.d models are common in analysis. *Stochastic analysis* is well studied in OR.
- In the i.d. and i.i.d models, we can assume that the distributions are *known* by the algorithm or *unknown*.
- In the *random order model* (ROM), an adversary creates a size n nemesis input set and then the items from that set are given in a uniform random order (i.e. uniform over the $n!$ permutations)

More general online frameworks

In the standard online model (and the variants we just mentioned), we are considering a one pass algorithm that makes one irrevocable decision for each input item.

There are many extensions of this one pass paradigm. For example:

- An algorithm is allowed some limited ability to revoke previous decisions.
- There may be some forms of lookahead (e.g. buffering of inputs).
- The algorithm may maintain a "small' number of solutions and then (say) take the best of the final solutions.
- The algorithm may do several passes over the input items.
- The algorithm may be given (in advance) some *advice bits* based on the entire input.

Throughout our discussion of algorithms, we can consider deterministic or randomized algorithms. In the online models, the randomization is in terms of the decisions being made. (Of course, the ROM model is an example of where the ordering of the inputs is randomized.)

Other measures of performance

The above variants address the issues of alternative input models, and relaxed versions of the online paradigm.

Competitive analysis is really just asymptotic approximation ratio analysis applied to online algorithms. Given the number of papers devoted to online competitive analysis, it is the standard measure of performance.

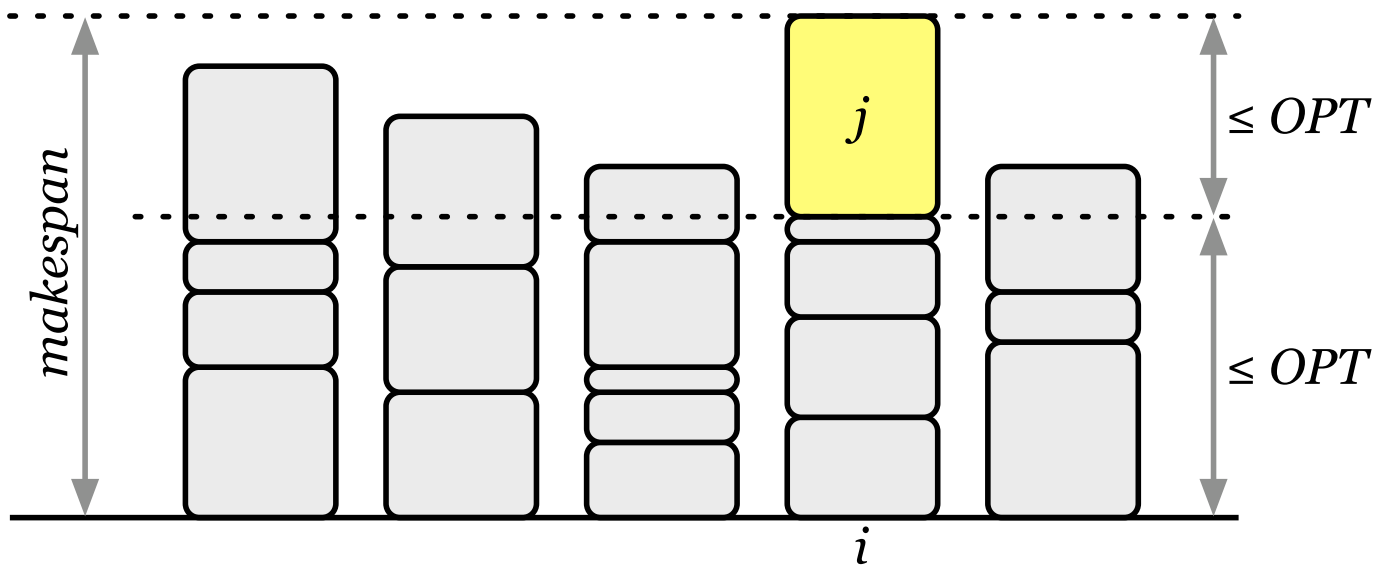
However, it has long been recognized that as a measure of performance, competitive analysis is often at odds with what seems to be observable in practice. Therefore, many alternative measures have been proposed. An overview of a more systematic study of alternative measures (as well as relaxed versions of the online paradigm) for online algorithms is provided in Kim Larsen's lecture slides that I have placed on the course web site.

See, for example, the discussion of the *accommodating function* measure (for the dual bin packing problem) and the *relative worst order* meaure for the bin packing coloring problem.

Returning to Graham's online greedy algorithm

Consider input jobs in any order (e.g. as they arrive in an *online* setting) and schedule each job J_j on any machine having the least load thus far.

- We will see that the **approximation ratio** for this algorithm is $2 - \frac{1}{m}$; that is, for any set of jobs $\mathcal{J}$, $C_{Greedy}(\mathcal{J}) \leq (2 - \frac{1}{m}) C_{OPT}(\mathcal{J})$.
  - C_A denotes the cost (or makespan) of a schedule A.
  - OPT stands for any optimum schedule.
- **Basic proof idea:** $OPT \geq (\sum_j p_j)/m$; $OPT \geq max_j p_j$
  What is C_{Greedy} in terms of these requirements for any schedule?



[picture taken from Jeff Erickson's lecture notes]

Graham's online greedy algorithm

Consider input jobs in any order (e.g. as they arrive in an online setting) and schedule each job J_j on any machine having the least load thus far.

- In the online "competitive analysis" literature the ratio $\frac{C_A}{C_{OPT}}$ is called the **competitive ratio** and it allows for this ratio to just hold in the limit as C_{OPT} increases. This is the analogy of *asymptotic approximation ratios*.

NOTE: Often, we will not provide proofs in the lecture notes but rather will do or sketch proofs in class (or leave a proof as an exercise).

- The approximation ratio for the online greedy is "tight" in that there is a sequence of jobs forcing this ratio.
- The negative result (i.e. there is "bad" input sequence for the algorithm) is not an asymptotic result. Is there a "meaningful asymptotic result?
- This bad input sequence suggests a better algorithm, namely the LPT (offline or sometimes called semi-online) algorithm.

Graham's LPT algorithm

Sort the jobs so that $p_1 \geq p_2 \ldots \geq p_n$ and then greedily schedule jobs on the least loaded machine.

- The (tight) approximation ratio of LPT is $\left(\frac{4}{3} - \frac{1}{3m}\right)$.
- It is believed that this is the **best** "greedy" algorithm but how would one prove such a result? This of course raises the question as to what is a greedy algorithm.
- We will present the priority model for greedy (and greedy-like) algorithms. I claim that all the algorithms mentioned on the next slide can be formulated within the priority model.
- Assuming we maintain a priority queue for the least loaded machine,
  - the online greedy algorithm would have time complexity $O(n \log m)$ which is $(n \log n)$ since we can assume $n \geq m$.
  - the LPT algorithm would have time complexity $O(n \log n)$.

Greedy algorithms in CSC373

Some of the greedy algorithms we study in different offerings of CSC 373

- The optimal algorithm for the fractional knapsack problem and the approximate algorithm for the proportional profit knapsack problem.
- The optimal unit profit interval scheduling algorithm and 3-approximation algorithm for proportional profit interval scheduling.
- The 2-approximate algorithm for the unweighted job interval scheduling problem and similar approximation for unweighted throughput maximization.
- Kruskal and Prim optimal algorithms for minimum spanning tree.
- Huffman's algorithm for optimal prefix codes.
- Graham's online and LPT approximation algorithms for makespan minimization on identical machines.
- The 2-approximation for unweighted vertex cover via maximal matching.
- The "natural greedy" $\ln(m)$ approximation algorithm for set cover.

Makespan: Some additional comments

- There are many refinements and variants of the makespan problem.
- There was significant interest in the best competitive ratio (in the online setting) that can be achieved for the identical machines makespan problem.
- The online greedy gives the best online ratio for m $= 2,3$ but better bounds are known for $m \geq 4$. For arbitrary m, as far as I know, following a series of previous results, the best known approximation ratio is 1.9201 (Fleischer and Wahl) and there is 1.88 inapproximation bound (Rudin). **Basic idea:** leave some room for a possible large job; this forces the online algorithm to be non-greedy in some sense but still within the online model.
- Randomization can provide somewhat better competitive ratios.
- Makespan has been actively studied with respect to three other machine models.

The uniformly related machine model

- Each machine i has a speed s_i
- As in the identical machines model, job J_j is described by a processing time or load p_j.
- The processing time to schedule job J_j on machine i is p_j/s_i.
- There is an online algorithm that achieves a constant competitive ratio.
- I think the best known online ratio is 5.828 due to Berman et al following the first constant ratio by Aspnes et al.
- Ebenlendr and Sgall establish an online inapproximation of 2.564 following the 2.438 inapproximation of Berman et al.

The restricted machines model

- Every job J_j is described by a pair (p_j, S_j) where $S_j \subseteq \{1, \ldots, m\}$ is the set of machines on which J_j can be scheduled.
- This (and the next model) have been the focus of a number of papers (for both online and offline) and there has been some relatively recent progress in the offline restricted machines case.
- Even for the case of two allowable machines per job (i.e. the *graph orientation problem*), this is an interesting problem and we will look at some recent work later.
- Azar et al show that $\log_2(m)$ (resp. $\ln(m)$) is (up to ± 1) the best competitive ratio for deterministic (resp. randomized) online algorithms with the upper bounds obtained by the "natural greedy algorithm".
- It is not known if there is an offline greedy-like algorithm for this problem that achieves a constant approximation ratio. Regev [IPL 2002] shows an $\Omega(\frac{\log m}{\log \log m})$ inapproximation for "fixed order priority algorithms" for the restricted case when every job has 2 allowable machines.

The unrelated machines model

- This is the most general of the makespan machine models.
- Now a job J_j is represented by a vector $(p_{j,1}, \ldots, p_{j,m})$ where $p_{j,i}$ is the time to process job J_j on machine i.
- A classic result of Lenstra, Shmoys and Tardos [1990] shows how to solve the (offline) makespan problem in the unrelated machine model with approximation ratio 2 using LP rounding.
- There is an online algorithm with approximation $O(\log m)$. Currently, this is the best approximation known for greedy-like (e.g. priority) algorithms even for the restricted machines model although there has been some progress made in this regard (which we will discuss later).
- NOTE: All statements about what we will do later should be understood as intentions and not promises.

Makespan with precedence constraints; how much should we trust our intuition

Graham also considered the makespan problem on identical machines for jobs satisfying a precedence constraint. Suppose $\prec$ is a partial ordering on jobs meaning that if $J_i \prec J_k$ then J_i must complete before J_k can be started. Assuming jobs are ordered so as to respect the partial order (i.e., can be reordered within the priority model) Graham showed that the ratio $2 - \frac{1}{m}$ is achieved by "the natural greedy algorithm", call it $\mathcal{G}_{\prec}$.

Makespan with precedence constraints; how much should we trust our intuition

Graham also considered the makespan problem on identical machines for jobs satisfying a precedence constraint. Suppose $\prec$ is a partial ordering on jobs meaning that if $J_i \prec J_k$ then J_i must complete before J_k can be started. Assuming jobs are ordered so as to respect the partial order (i.e., can be reordered within the priority model) Graham showed that the ratio $2 - \frac{1}{m}$ is achieved by "the natural greedy algorithm", call it $\mathcal{G}_{\prec}$.

Graham's 1969 paper is entitled "Bounds on Multiprocessing Timing Anomalies" pointing out some very non-intuitive anomalies that can occur.

Consider $\mathcal{G}_{\prec}$ and suppose we have a given an input instance of the *makespan with precedence* problem. Which of the following should never lead to an increase in the makepan objective for the instance?

- Relaxing the precedence $\prec$
- Decreasing the processing time of some jobs
- Adding more machines

Makespan with precedence constraints; how much should we trust our intuition

Graham also considered the makespan problem on identical machines for jobs satisfying a precedence constraint. Suppose $\prec$ is a partial ordering on jobs meaning that if $J_i \prec J_k$ then J_i must complete before J_k can be started. Assuming jobs are ordered so as to respect the partial order (i.e., can be reordered within the priority model) Graham showed that the ratio $2 - \frac{1}{m}$ is achieved by "the natural greedy algorithm", call it $\mathcal{G}_\prec$.

Graham's 1969 paper is entitled "Bounds on Multiprocessing Timing Anomalies" pointing out some very non-intuitive anomalies that can occur.

Consider $\mathcal{G}_\prec$ and suppose we have a given an input instance of the *makespan with precedence* problem. Which of the following should never lead to an increase in the makepan objective for the instance?

- Relaxing the precedence $\prec$
- Decreasing the processing time of some jobs
- Adding more machines

In fact, all of these changes could increase the makespan value.

Index of cbapters

My intention is to cover selected chapters in theb text we are considering. We end today meeting with a quick llok at the index for the text.