Convex Relaxation Approach to Appointment Scheduling with Piecewise Linear Cost Functions

Dongdong Ge∗ Guohua Wan† Zizhuo Wang‡ Jiawei Zhang§

November 6, 2011

Abstract: This note is concerned with the problem of scheduling a set of jobs with predetermined order on a single processor. The objective is to minimize the expected total cost incurred by job waiting and processor idling, where the job processing times are discrete random variables. We extend the scope of the discussion in Begen and Queyranne [2011] to piecewise linear cost functions. For piecewise linear functions with integer break points, we show the existence of an integer optimal solution. We also develop a convex programming relaxation and show that under mild conditions, the problem can be solved in polynomial time. In addition, we develop a direct sampling method when the distribution information is not known and only a set of samples is available and specify the bound on the number of samples required to obtain a provably near-optimal solution. The bound we obtain serves as an important complement to the results in Begen et al. [2010].

Subject Classifications: Health Care; Scheduling; Stochastic Programming; Convex Relaxation; Sampling; Duality

Area of Review: Optimization

1 Introduction

We study the appointment scheduling problem with discrete random durations, introduced by Begen and Queyranne [2011] and Begen et al. [2010], but with a broader class of cost functions. In such problems, there is a set of jobs to be processed on a single processor. Given the sequence of the jobs, the decision maker has to determine the starting time of each job, called the appointment date. A job can not be started before the previous job is finished or before its appointment date.

∗Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200052, China. Email:ddge@sjtu.edu.cn
†Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200052, China. Email:ghwan@sjtu.edu.cn
‡Department of Management Science and Engineering, Stanford University, Stanford, CA, 94305, USA. Email:zzwang@stanford.edu
§Stern School of Business, IOMS-Operations Management, New York University, New York, NY, 10012, USA. Email:jwzhang@stern.nyu.edu
Due to the randomness of processing durations, a job may be finished earlier or later than the appointment date of the next job. If it is finished earlier, an “underage” cost will be incurred due to the idling of the processor. On the other hand, if a job is finished later, the system will experience an “overage cost” due to the waiting of the next job. The objective of the appointment scheduling problem is to determine the appointment date for each job such that the total expected cost is minimized.

Appointment scheduling problem has many emerging applications in service management, particularly in health care industry. In health care problems, the jobs can be viewed as treatments (e.g., surgery, examination) for patients, and the processors are doctors or medical devices. Due to a high variability of the treatment durations and a high cost of the idling of doctors or devices and the waiting of patients, the appointment scheduling problem has become especially important and challenging. Other important applications appear in manufacturing industry (e.g., project scheduling), transportation problems (e.g., runway scheduling in airports) and etc.

Our work mainly aims to extend the seminal study in Begen and Queyranne [2011] and Begen et al. [2010]. In these two studies, the authors assumed that both the underage and overage cost are linear functions. Under this assumption, Begen and Queyranne [2011] showed the existence of an integer optimal solution and for the first time, characterized a class of problems that can be solved to optimality in polynomial time. Meanwhile, Begen et al. [2010] discussed a sampling method for the appointment scheduling problem in which only historical data are known. In this note, we consider a more general class of cost functions, namely the piecewise linear cost functions. Piecewise linear cost functions are of particular interest in practice since if one only considers integer decision variables and processing times, any cost function can be equivalently written as a piecewise linear function. In this sense, when one confines himself to a discrete sets of solutions, considering piecewise linear cost functions in fact allows arbitrary cost structures for this problem.

Our work extends the discussion of cost functions to a more general class, simplifies the proofs and provides an alternative sampling estimation analysis. More specifically, we extend and improve the results of Begen and Queyranne [2011] and Begen et al. [2010] in the following ways:

- In Begen and Queyranne [2011], the authors showed that when the cost functions are linear, there exists an integer optimal solution to the appointment scheduling problems, given that the process durations are integers. We generalize this result to piecewise linear cost functions, i.e., an integer optimal solution exists if the cost functions are piecewise linear with integer break points. Our analysis is much simpler by applying the unimodularity property of linear programming.

- In Begen and Queyranne [2011], the authors established a polynomial time algorithm when the linear cost coefficients are $\alpha$-monotone. We extend this result to a more general condition for piecewise linear cost functions which takes $\alpha$-monotonicity as a special case. Our condition also includes the case in which the underage cost functions are linear with homogeneous slopes and the overage cost functions are arbitrary convex piecewise linear. The latter case is of particular interest in practice because it captures the cost structures in many service
In Begen et al. [2010], the authors developed a sampling approach when the distribution of the processing time is not known and only a set of independent samples is available. They analyzed the subdifferentials of the cost function and established a bound on the number of samples required to obtain a near-optimal solution (with multiplicative error) with high probability. In our work, a simple analysis extends the sampling approach to a general class of functions and provides an alternative bound on the number of samples required to obtain a near-optimal solution (with additive error). It is observed that our bound serves as an important complement to theirs.

The reminder of the paper is organized as follows. In Section 2, we describe the model and prove the existence of an integer optimal solution. In Section 3, we develop a convex relaxation to the problem and prove that under mild assumptions the problem can be solved in polynomial time. In Section 4, we discuss the situation in which the distribution of the processing durations is unknown and study a sampling approach. We conclude and make the final remarks in Section 5.

Before we proceed to the models, a brief literature review is in place. Due to the importance of the appointment scheduling problem, there is a rich literature on this subject. However, since the goal of this paper is mainly to serve as a note to complement the recent researches, we only review the literatures that directly relates to this work. We refer the readers to Pinedo [2001], Cardoen et al. [2010] and Cayirli and Veral [2003] for a more detailed review on this subject.

Based on the solution methods, the study of the appointment scheduling problem falls into different categories. One category uses the simulation based method, e.g., Klassen and Rohleder [1996], Rohleder and Klassen [2000], Liu and Liu [1998], and Yang et al. [1998]. These methods tend to find acceptable solutions to the problem or to find insights of scheduling policies by using simulation approach. Another approach, which we mostly take in this paper, is the stochastic programming method, e.g., Denton and Gupta [2003], Robinson and Chen [2003], Begen and Queyranne [2011] and Begen et al. [2010]. Denton and Gupta [2003] model the appointment scheduling problem as a two-stage stochastic program and solve it using a decomposition approach. Robinson and Chen [2003] formulate the problem as a stochastic program and solve it by Monte-Carlo integration. Recently, Begen and Queyranne [2011] study the appointment scheduling problem by assuming that the service duration information of each job is given by a joint discrete probability distribution. Under the assumption that the cost functions are linear and satisfy the \( \alpha \)-monotonicity, they prove two important properties of the objective function, namely, submodularity and L-convexity. Based on these properties, they develop a polynomial time algorithm to solve the problem. Furthermore, Begen et al. [2010] develop a sampling-based approach to a more realistic situation in which the distribution of service durations is not known in advance, while only a set of independent samples is available. They prove that solving a sampling problem can generate a near-optimal solution with high probability to the original problem and specify the number of samples needed for a given accuracy level and confidence level. Our note will mainly build on the studies by Begen and Queyranne [2011] and Begen et al. [2010] and hopefully shed light on the structure of the scheduling problem.
The Model and the Existence of Integer Optimal Solutions

We closely follow Begen and Queyranne [2011] to set up the model. Let \( S \) denote the set of jobs to be scheduled in a given session and let \( n = |S| \) be the total number of jobs. Each job \( i \in S \) has a random service duration denoted by \( p_i(\omega) \) where uncertainty is represented by the random outcome \( \omega \). The set of all outcomes is denoted by \( \Omega \). In our study, we assume that the service durations are integer-valued and bounded. More precisely, we make the following assumption about the random service durations.

**Assumption 1.** For each \( i \in S \), there exist a pair of nonnegative numbers, \( \underline{p}_i \) and \( \bar{p}_i \), such that \( p_i(\omega) \in [\underline{p}_i, \bar{p}_i] \cap \mathbb{Z} \) for all \( \omega \in \Omega \), where \( \mathbb{Z} \) is the set of integers.

Given \( n \) jobs, they must be scheduled sequentially in a predetermined order 1, 2, ..., \( n \). Let \( A_i \) denote the appointment time of job \( i \in S \). We also define an \((n+1)\)th job, a dummy job with service time 0, to compute the overage and underage cost of the \( n \)th job. The appointment time of the \((n+1)\)th job \( A_{n+1} \) can be viewed as the scheduled makespan of the whole session. We denote the decision variable in this problem by \( \mathbb{A} = \{A_1, A_2, \ldots, A_n, A_{n+1}\} \). A job can not start before its appointment time nor before its predecessors have been completed, and the start and completion times of a job depend on the appointment times as well as the random service durations. They can be computed recursively as follows. Let \( S_i(\omega) \) and \( C_i(\omega) \) denote the start time and completion time, respectively, of job \( i \) in scenario \( \omega \in \Omega \). Then

\[
S_i(\omega) = \max(C_{i-1}(\omega), A_i) \tag{1}
\]

\[
C_i(\omega) = S_i(\omega) + p_i(\omega) \tag{2}
\]

for \( i = 1, \ldots, n \), and \( C_0(\omega) = 0 \). If job \( i \) is completed before (after, respectively) the appointment time of job \( i+1 \), it incurs an underage (overage, respectively) cost captured by function \( u_i(\cdot) \) \((o_i(\cdot), \text{respectively})\): \( \mathbb{R}^+ \to \mathbb{R} \). More precisely, the underage cost and overage cost of job \( i \) are \( u_i((C_{i+1} - C_i(\omega))^+) \) and \( o_i((C_i(\omega) - A_{i+1})^+) \), respectively. Here \( x^+ = \max(x,0) \) is the positive part of real number \( x \). We make the following assumption about the overage and underage cost functions.

**Assumption 2.** The functions \( u_i(\cdot) \) and \( o_i(\cdot) \) are non-decreasing and satisfy \( o_i(0) = 0 \) and \( u_i(0) = 0 \).

In the appointment scheduling problem, the objective is to minimize the expected total overage and underage cost. We can formulate the problem as follows:

\[
\min_{\mathbb{A}} \mathbb{E}_\omega \left\{ \sum_{i=1}^n \{o_i((C_i(\omega) - A_{i+1})^+) + u_i((A_{i+1} - C_i(\omega))^+)\} \right\} \tag{3}
\]
s.t. \( C_i(\omega) = \max(A_i, C_{i-1}(\omega)) + p_i(\omega) \quad \forall i \in S, \forall \omega \in \Omega. \)

It is easy to see that (3) can be equivalently written as the following form.

\[
(P1) \quad \min_{\mathbb{A},C} \mathbb{E}_\omega \left[ \sum_{i=1}^n \{o_i((C_i(\omega) - A_{i+1})^+) + u_i((A_{i+1} - C_i(\omega))^+)\} \right] \tag{4}
\]
s.t. \( C_i(\omega) = \max(A_i, C_{i-1}(\omega)) + p_i(\omega) \quad \forall i \in S, \forall \omega \in \Omega. \tag{5} \]
The next theorem establishes the conditions under which \( (P1) \) has an integer optimal solution.

**Theorem 1.** In addition to Assumptions 1 and 2, if functions \( u_i(\cdot) \) and \( o_i(\cdot) \) are both piecewise linear and the break points are integers, then \( (P1) \) has an integer optimal solution.

**Proof.** By the assumption that \( o_i(\cdot) \) and \( u_i(\cdot) \) are piecewise linear with integer break points, for each \( i \in S \) and each nonnegative integer \( l \), there exist \( o^l_i, r^l_i, u^l_i, q^l_i \) such that

\[
o_i(x) = o^l_i x + r^l_i \quad \text{if} \quad x \in [l, l+1)
\]

and

\[
u_i(x) = u^l_i x + q^l_i \quad \text{if} \quad x \in [l, l+1).
\]

Let \( A^* \) be any optimal solution to problem \( (P1) \) and \( C^* \) be the corresponding job completion times. For each \( \omega \in \Omega \), let

\[
I(\omega) = \{ i \in S : A^*_i + 1 \geq C^*_i(\omega) \}.
\]

It follows that, for \( i \in S \setminus I(\omega) \),

\[
C^*_i(\omega) > A^*_i + 1
\]

Therefore, the optimal objective value of \( (P1) \) can be written as

\[
E_\omega \left[ \sum_{i \in S \setminus I(\omega)} \left( o^l_i(\omega) (C^*_i(\omega) - A^*_i + 1) + r^l_i(\omega) \right) + \sum_{i \in I(\omega)} \left( u^l_i(\omega) (A^*_i + 1 - C^*_i(\omega)) + q^l_i(\omega) \right) \right]
\]

where \( l^u_i(\omega) = |C^*_i(\omega) - A^*_i + 1| \) and \( l^l_i(\omega) = |A^*_i + 1 - C^*_i(\omega)| \). Now we define a linear program as follows (we define \( p_{n+1}(\omega) = 0 \):

\[
\begin{align*}
\text{(LP)} & \quad \min_{X,Y} \quad E_\omega \left[ \sum_{i \in S \setminus I(\omega)} \left( o^l_i(\omega) (Y_i(\omega) - X_{i+1}) + r^l_i(\omega) \right) + \sum_{i \in I(\omega)} \left( u^l_i(\omega) (X_{i+1} - Y_i(\omega)) + q^l_i(\omega) \right) \right] \\
\quad \text{s.t.} & \quad Y_{i+1}(\omega) = X_{i+1} + p_{i+1}(\omega) \quad \forall \omega \in \Omega, \quad i \in I(\omega) \\
& \quad Y_{i+1}(\omega) \geq Y_i(\omega) + p_{i+1}(\omega) \quad \forall \omega \in \Omega, \quad i \in I(\omega) \\
& \quad Y_{i+1}(\omega) = Y_i(\omega) + p_{i+1}(\omega) \quad \forall \omega \in \Omega, \quad i \in S \setminus I(\omega) \\
& \quad Y_{i+1}(\omega) \geq X_{i+1} + p_{i+1}(\omega) \quad \forall \omega \in \Omega, \quad i \in S \setminus I(\omega) \\
& \quad l^u_i(\omega) \leq Y_i(\omega) - X_{i+1} \leq l^l_i(\omega) + 1 \quad \forall \omega \in \Omega, \quad i \in S \setminus I(\omega) \\
& \quad l^l_i(\omega) \leq X_{i+1} - Y_i(\omega) \leq l^u_i(\omega) + 1 \quad \forall \omega \in \Omega, \quad i \in I(\omega).
\end{align*}
\]

By Assumption 1, the set \( \Omega \) is finite, therefore \( (LP) \) is a finite dimension linear program. Note that any feasible solution \( (X, Y) \) to \( (LP) \) is a feasible solution to \( (P1) \) by setting \( (A, C) = (X, Y) \). Furthermore, \( (X, Y) = (A^*, C^*) \) is a feasible solution to \( (LP) \). Now notice that each constraint of \( (LP) \) contains exactly one 1 and one -1, thus the constraint matrix of \( (LP) \) is totally unimodular. Therefore, problem \( (LP) \) has an integer optimal solution (see, e.g., Heller and Tompkins [1956]).
By the construction of $\text{LP}$, this integer solution has the same objective value as $\textbf{(A)} \ast, \textbf{C} \ast$ and thus is optimal to $\textbf{(P1)}$. This completes the proof.  \[\square\]

We note that the piecewise linear cost functions are important because all cost functions can be represented by piecewise linear functions if only integer grid points are defined. Therefore, Theorem 1 justifies the use of $\textbf{(P1)}$ to find the optimal integer schedules regardless of the cost functions.

## 3 Convex Relaxation and Polynomial Time Algorithm

In this section, we study how we can solve $\textbf{(P1)}$ efficiently. In Begen and Queyranne [2011], it is shown that for linear cost functions, the scheduling problem can be solved in polynomial time if the cost coefficients $(u, o)$ satisfy the $\alpha$-monotonicity property. In this work, we extend this result to a more general condition. For this purpose, it is useful to reformulate problem $\textbf{(P1)}$ as a two-stage stochastic program. We first rewrite the objective function based on the following observation. From constraint (5), we have

$$C_{i+1}(\omega) = \max(C_i(\omega), A_{i+1}) + p_{i+1}(\omega) = (C_i(\omega) - A_{i+1})^+ + A_{i+1} + p_{i+1}(\omega).$$

It follows that

$$(C_i(\omega) - A_{i+1})^+ = C_{i+1}(\omega) - A_{i+1} - p_{i+1}(\omega).$$  \(6\)

Similarly, we have

$$(A_{i+1} - C_i(\omega))^+ = C_{i+1}(\omega) - C_i(\omega) - p_{i+1}(\omega).$$  \(7\)

Using (6) and (7), we can replace the objective function of $\textbf{(P1)}$ with

$$\min_{A,C} E_{\omega} \left[ \sum_{i=1}^{n} \{ o_i(C_{i+1}(\omega) - A_{i+1} - p_{i+1}(\omega)) + u_i(C_{i+1}(\omega) - C_i(\omega) - p_{i+1}(\omega)) \} \right].$$  \(8\)

For any $A, C$, and $\omega$, we define

$$h_{\omega}(C,A) = \sum_{i=1}^{n} \{ o_i(C_{i+1}(\omega) - A_{i+1} - p_{i+1}(\omega)) + u_i(C_{i+1}(\omega) - C_i(\omega) - p_{i+1}(\omega)) \}. $$ \(9\)

Now we are ready to present a two-stage stochastic programming formulation of $\textbf{(P1)}$ as follows:

$$\textbf{(P2)} \quad \min_{A} \quad g(A) := E_{\omega}[f_{\omega}(A)]$$

where for each $\omega \in \Omega$ and $A$,

$$f_{\omega}(A) := \min_{C} \quad h_{\omega}(C,A) \quad \text{s.t.} \quad C_{i+1}(\omega) = \max(A_{i+1}, C_i(\omega)) + p_{i+1}(\omega) \quad \forall i \in S$$ 

$$C_1(\omega) = p_1(\omega).$$  \(12\)  \(13\)
Here $p_{n+1}(\omega) = 0$ and $C = (C_i(\omega) : i = 1, 2, \cdots, n + 1)$ are the second stage (recourse) variables. Notice that with constraint (12), there is only one feasible solution to each second stage problem so that the minimization is actually not necessary. We keep using this two-stage stochastic programming formulation since it will become useful when we relax constraint (12).

In the following, we assume that the cost functions $o_i(\cdot)$ and $u_i(\cdot)$ are convex. Although the objective function of the second stage problem $h_\omega(C, A)$ is convex in $C$ for any given $\omega$ and $A$, the feasible set is in general not. This motivates us to relax constraint (12) by linear inequalities. By doing this, we get the following two-stage stochastic programming problem as a convex relaxation to (10)-(13):

$$\min_A G(A) := E_\omega[F_\omega(A)]$$  \hspace{1cm} (14)

where for each $\omega \in \Omega$ and $A$,

$$F_\omega(A) := \min_C h_\omega(C, A)$$  \hspace{1cm} (15)

s.t. \hspace{0.5cm} \begin{align*}
C_{i+1}(\omega) &\geq A_{i+1} + p_{i+1}(\omega) \quad \forall i \in S \\
C_{i+1}(\omega) &\geq C_i(\omega) + p_{i+1}(\omega) \quad \forall i \in S \\
C_1(\omega) &= p_1(\omega) \hspace{1cm} (17)
\end{align*}

It is clear that constraints (16)-(18) are relaxations of (12)-(13), thus problem (14)-(18) is a convex relaxation of problem (10)-(13). Therefore, $F_\omega(A) \leq f_\omega(A)$ for each $\omega \in \Omega$ and $A$, and $G(A) \leq g(A)$.

In the following, we study the conditions under which the convex relaxation is exact. We have the following theorem.

**Theorem 2.** If $h_\omega(C, A)$ is non-decreasing in $(C_2(\omega), C_3(\omega), \cdots, C_n(\omega))$ for each $\omega \in \Omega$ and $A$, then the convex relaxation (14)-(18) is exact, i.e., $F_\omega(A) = f_\omega(A)$ for each $\omega \in \Omega$ and $A$, and $G(A) = g(A)$.

**Proof.** Since problem (14)-(18) is a relaxation of problem (10)-(13) with the same objective function. It suffices to show that there exists an optimal solution to problem (14)-(18) that is feasible to problem (10)-(13).

Suppose that for each $\omega \in \Omega$ and $A$, $C^*$ is the least optimal solution to problem (15)-(18). In the following, we show that $C^*$ satisfies constraint (12)-(13). If not, there exists $j$ such that $C^*_{j+1} > \max(A_{j+1}, C_j^*) + p_{j+1}(\omega)$. Without loss of generality, we assume that $j$ is the smallest index such that constraint (12) is violated for this $A$ and $\omega$. Now we consider a new solution $\bar{C}$ with all the values the same as $C^*$ except that $\bar{C}_{j+1} = \max(A_{j+1}, C_j^*) + p_{j+1}(\omega)$. It is clear that $\bar{C}$ is feasible to problem (15)-(18). But since $\bar{C} \leq C^*$ ($\bar{C}_{j+1} < C_j^* + p_{j+1}(\omega)$), and by assumption that $h_\omega(C, A)$ is non-decreasing in $C$, we must have $h_\omega(\bar{C}, A) \leq h_\omega(C^*, A)$. This contradicts the assumption that $C^*$ is the least optimal solution to problem (15)-(18). \hspace{1cm} \Box

We now specify a condition on $u(\cdot)$ and $o(\cdot)$ such that the conditions in Theorem 2 is satisfied.
**Proposition 1.** Assume \( u(\cdot) \) and \( o(\cdot) \) are both convex, non-decreasing piecewise linear functions. Define \( u_i \) and \( \bar{u}_i \) to be the smallest and largest slopes for \( u_i(\cdot) \); \( a_i \) and \( \bar{a}_i \) to be the smallest and largest slopes for \( o_i(\cdot) \), respectively. If \( u_i, a_i \) and \( \bar{u}_i \) satisfies

\[
\begin{align*}
u_i + a_i & \geq \bar{u}_{i+1} \quad \text{for } i = 1, \ldots, n - 1, \\
\end{align*}
\]

then the convex relaxation (14)-(18) is exact.

Before we prove Proposition 1, we remark that the condition specified in Proposition 1 is quite general and a broad class of cost functions fit into this frame. In particular, we would like to specify two important classes of cost functions that satisfy condition (19):

**P-uniform:** In this case, there exists \( u \geq 0 \), such that \( u_i(x) = u \cdot x \) for all \( i \in S \) while \( o_i(x) \) can be any convex non-decreasing piecewise linear function. This describes a very important and practical case, which is the case with uniform idle cost rate. In real applications, the idle cost is usually generated by the same processor (e.g., doctors) while the overage cost are generated by different jobs (e.g., different patients). Therefore, this case has very broad applications in real problems.

**Generalized \( \alpha \)-monotonicity:** In this case, it is assumed that both underage cost and overage cost functions are linear, i.e., \( o_i(x) = o_i \cdot x \) and \( u_i(x) = u_i \cdot x \) and \( u_i + o_i \geq u_{i+1} \) for \( i = 1, \ldots, n - 1 \). One can verify that if the cost function satisfies the \( \alpha \)-monotonicity defined in Begen and Queyranne [2011], then it must satisfy generalized \( \alpha \)-monotonicity. Therefore, this case is a generalization of the \( \alpha \)-monotonicity property defined in Begen and Queyranne [2011].

**Proof of Proposition 1:** By Theorem 2, it suffices to verify that for each \( i, i = 1, \ldots, n - 1 \), \( h_\omega(\mathbb{C}, \mathbb{A}) \) is non-decreasing in \( C_{i+1}(\omega) \) when (19) holds.

To show this, note that in (9), the derivative of \( h_\omega(\mathbb{C}, \mathbb{A}) \) to \( C_{i+1} \) is \( a_i + u_i - \bar{u}_{i+1} \), for \( i = 1, \ldots, n - 1 \). Therefore, when (19) holds, \( h_\omega(\mathbb{C}, \mathbb{A}) \) must be non-decreasing and therefore the convex relaxation is exact. \( \square \)

Next, we study the characteristics of the objective function of the scheduling problem based on the two-stage stochastic program formulation. First, we have the following theorem showing the convexity and submodularity of the objective function \( G(\mathbb{A}) \):

**Theorem 3.** \( G(\mathbb{A}) \) is convex and submodular in \( \mathbb{A} \) on \( \mathbb{R}^{n+1} \).

**Proof.** It suffices to show that \( F_\omega(\mathbb{A}) \) is convex and submodular in \( \mathbb{A} \) for every \( \omega \in \Omega \). By assumption, \( o_i(\cdot) \) is non-decreasing and convex for each \( i \in S \). It follows that \( o_i(C_{i+1}(\omega) - A_{i+1} - p_{i+1}(\omega)) \) is jointly convex and submodular in \((C_{i+1}(\omega), A_{i+1})\). Also, \( u_i(C_{i+1}(\omega) - C_i(\omega) - p_{i+1}(\omega)) \) is jointly convex and submodular in \((C_i(\omega), C_{i+1}(\omega))\). Therefore, \( h(\mathbb{C}, \mathbb{A}) \) is jointly convex and submodular in \((\mathbb{C}, \mathbb{A})\).

Also inequalities (16) and (17) are linear and each contains at most two entries in \((\mathbb{C}, \mathbb{A})\) with opposite signs. Therefore, the set of \((\mathbb{C}, \mathbb{A})\) that satisfy the inequalities (16) and (17) is a lattice and is convex.
It is well-known that the projection of a convex (submodular, respectively) function is still convex (submodular, respectively), thus the theorem follows. □

Furthermore, one can prove that the function \( G(\mathbb{A}) \) also satisfies L-convexity, an important discrete convexity property.

**Definition 1.** (Murota [2003a]) \( f : \mathbb{Z}^q \to \mathbb{R} \cup \infty \) is L-convex if and only if
\[
 f(z) + f(y) \geq f(z \lor y) + f(z \land y), \quad \forall z, \forall y \in \mathbb{Z}^q \text{ and } \exists r \in \mathbb{R} : f(z + 1) = f(z) + r, \quad \forall z \in \mathbb{Z}^q.
\]

**Theorem 4.** \( G(\mathbb{A}) \) is L-convex in \( \mathbb{A} \) on \( \mathbb{R}^{n+1} \).

**Proof.** Due to the submodularity by Theorem 3, we only need to prove that there exists \( r \in \mathbb{R} \), such that \( G(\mathbb{A} + 1) = G(\mathbb{A}) + r \), for all \( \mathbb{A} \in \mathbb{Z}^{n+1} \).

We claim \( G(\mathbb{A} + 1) = G(\mathbb{A}) \). Given \( \mathbb{A} \in \mathbb{R}^{n+1} \) and a realization \( \omega \in \Omega \), for any feasible solution \( C \) to problem \( F_\omega(\mathbb{A}) \), \( C + 1 \) is also a feasible solution to problem \( F_\omega(\mathbb{A} + 1) \) and \( h_\omega(C, \mathbb{A}) = h_\omega(C + 1, \mathbb{A} + 1) \). Therefore \( F_\omega(\mathbb{A} + 1) \leq F_\omega(\mathbb{A}) \). Similarly, we can prove that \( F_\omega(\mathbb{A}) \leq F_\omega(\mathbb{A} + 1) \). Therefore \( F_\omega(\mathbb{A}) = F_\omega(\mathbb{A} + 1) \) for any \( \omega \in \Omega \). Thus we conclude that \( G(\mathbb{A} + 1) = G(\mathbb{A}) \). Combining with Theorem 3, we conclude that \( G(\mathbb{A}) \) is L-convex. □

Now we prove the theorems regarding the complexity to solve (P1). The proofs are similar to Theorem 7.1 - 7.3 in Begen and Queyranne [2011].

**Theorem 5.** Assume \( u_i(\cdot) \) and \( o_i(\cdot) \) are piecewise linear functions with integer break points satisfying Assumption 2 and condition (19). Then there exists an algorithm which solves (P1) using polynomial time and a polynomial number of expected cost evaluations.

**Proof.** By Proposition 1, it suffices to show that under such conditions on \( u_i(\cdot) \) and \( o_i(\cdot) \), there exists an algorithm to minimize \( G(\mathbb{A}) \) using polynomial time and a polynomial number of expected cost evaluations. According to Theorem 1 and Proposition 1, there exists an optimal integer vector solution \( \mathbb{A}^* \) to minimize \( G(\mathbb{A}) \). Thus we only need to consider minimizing \( G(\mathbb{A}) \) over integer vectors. We have proved in Theorem 5 that \( G(\mathbb{A}) \) is L-convex. Therefore, by using Iwata’s steepest descent scaling algorithm (see e.g., Murota [2003b]), \( G(\mathbb{A}) \) can be minimized in \( O(\sigma(n) EOn^2 \log \tilde{p}) \) time where \( \sigma(n) \) is the number of function evaluations required to minimize a submodular set function over an \( n \)-element ground set and \( E \) is the time needed for an expected cost evaluation for \( G(\mathbb{A}) \). □

**Theorem 6.** Assume that the processing durations are stochastically independent and the assumptions in Theorem 5 hold. Then (P1) can be solved in \( O(n^9 \tilde{p}^2 \log \tilde{p}) \) time.

**Proof.** According to Theorem 2 in Begen and Queyranne [2011]), if the processing durations are stochastically independent, then \( G(\mathbb{A}) \) may be computed in \( O(n^5 \tilde{p}^2) \) time. Thus the proof follows from Theorem 5 and the fact that \( \sigma(n) = O(n^5) \) by Orlin [2001]. □
4 A Sampling Approach

In the previous sections, we assume that the complete distribution information of the job durations is known. However, in practice, the true distribution of processing time is often not explicitly given, but a set of independent samples from the past data following the true underlying distribution is available. We study such cases in this section. We use the well-known sample average approximation (SAA) approach, i.e., to minimize the empirical cost function of $N$ independent samples drawn from the true distribution, and compare the sampled optimal solution to the optimal solution of the original problem. In this section, we develop a bound on the required number of independent samples to guarantee that the solution obtained by SAA is near-optimal with high probability.

Following the previous discussion, we make the following assumption to the cost functions in this section:

**Assumption 3.** For every $i$, the cost functions $u_i(\cdot)$ and $o_i(\cdot)$ are convex, non-decreasing, and piecewise linear with integer break points. Furthermore, they all satisfy condition (19).

Under Assumption 3, the convex relaxation in (15)-(18) is exact. Suppose we have $N$ samples denoted by $p(\omega_k) = (p_1(\omega_k), p_2(\omega_k), ..., p_n(\omega_k))$ and let $\hat{\omega} = \hat{\omega}(N)$ be the empirical distribution of $\omega$, i.e., $\text{Prob}(\hat{\omega} = \omega_k) = \frac{1}{N}$ for $k = 1, 2, ..., N$. We denote the objective function obtained from the empirical distribution by $\hat{G}(\hat{A}) = E_{\hat{\omega}}[F_{\omega}(\hat{A})]$ and denote by $\hat{A}$ an optimal solution that minimizes $\hat{G}(\hat{A})$. We also define $A^*$ to be the true optimal appointment vector, i.e., $A^*$ is a minimizer of $G(A)$.

Given $N$ samples, if the cost functions satisfy Assumption 3, the time to evaluate an expected cost is $N^2$. Thus we can find an optimal solution $\hat{A}$ for $\hat{G}(\hat{A})$ efficiently.

**Corollary 1.** Assume the cost functions satisfy Assumption 3. Then given $N$ samples, $\hat{G}(\hat{A})$ can be minimized in $O(n^2N \log \bar{p})$ time.

For the case in which the piecewise linear cost functions don’t satisfy condition (19), we remark that the sampling problem with the convex relaxation can still be solved efficiently. By applying the network programming algorithms in Ahuja et al. [1999], we can show that for general piecewise linear cost functions with integer break points, $\hat{G}(\cdot)$ can be minimized in $O(n^9N \log([h/2n]) \log n \log n^2 \bar{p})$). Thus the sampling approach can still provide a useful approximation to the original problem in this case.

Next we specify the bound on the required number of samples such that the SAA solution $\hat{A}$ will be near-optimal. We introduce the following form of the Hoeffding inequality:

**Lemma 1.** (van der Vaart and Wellner [1996]) Let $x_1, ..., x_r$ be samples drawn from a random variable $c$. Then

$$P(|\frac{1}{r} \sum_{i=1}^{r} x_i - \bar{c}| \geq \epsilon) \leq 2 \exp\left(\frac{-r \epsilon^2}{C^2}\right)$$

where $C = \sup c$, $\bar{c}$ is the expectation of $c$.

We have the following theorem regarding the bound in our sampling approach.
Theorem 7. Let $\epsilon > 0$ (accuracy level) and $0 < \delta < 1$ (confidence level) be given. If $N \geq 4(1/\epsilon)^2 n^5 (o_{\text{max}} + u_{\text{max}})^2 \bar{p}^2 \log (n\bar{p}/\delta)$ then $G(\hat{A}) \leq G(A^*) + \epsilon$ with probability at least $1 - \delta$.

Before we prove Theorem 7, we would like to compare it to the result in Begen et al. [2010]. For the linear cost functions satisfying the $\alpha$-monotonicity, they obtained a first bound on the samples, $(4.5(1/\epsilon)^2 (n^2 (n+1) (9o_{\text{max}} + 4u_{\text{max}})/\eta)^2 \log (2(5n^2 + 5)/\delta))$, to guarantee that $G(\hat{A}) \leq (1 + \epsilon)G(A^*)$ with probability at least $1 - \delta$, where $\eta = \min\{u_{\text{max}}, o_{\text{max}}\}$. Firstly, our result applies to a more general class of cost functions defined in Assumption 3, including linear cost functions satisfying $\alpha$-monotonicity as a special case. Secondly, the bound in Theorem 7 is in terms of an additive error while the bounds in Begen et al. [2010] is in terms of a multiplicative error. This difference makes it hard to directly compare the two bounds in general. Besides the difference in the logarithm term (which is usually negligible in practice), the ratio between our bound to theirs is approximately $\bar{p}^2 \eta^2 / 190n$ (assume $o_{\text{max}}$ is approximately equal to $u_{\text{max}}$). Indeed, $\bar{p}$ could be large in practice. However, when it is large, it is likely that the optimal objective value is large as well, which makes the additive bound stronger and potentially compensates the deteriorations of the bound. On the other hand, our bound does not involve $\eta$, which is the smallest slopes for all the cost functions. In the cases where a small amount of waiting time is not costly, $\eta$ could be very small (or even be zero). Also, our bound involves a smaller constant and one order less in $n$. Although we can not directly compare these two bounds in general, it is clear to us that the bound stated in Theorem 7 will serve as an important complement to the bound in Begen et al. [2010] and it applies to a more general class of problems as described.

Proof of Theorem 7: The proof follows the idea introduced in Kleywegt et al. [2001]. First note that for any set of samples $\tilde{\Omega}$, the optimal solution $A$ must satisfy $A_i \leq \sum_{j<i} \bar{p}_j$ (See Lemma 4.3 in Begen and Queyranne [2011]). We claim that, for any cost functions satisfying Assumption 3, when the number of samples $N \geq 4(1/\epsilon)^2 n \log (n\bar{p}/\delta) C^2$, where $C = n^2 (u_{\text{max}} + o_{\text{max}}) \bar{p} \geq \sum_{i=1}^n (u_{\text{max}} + o_{\text{max}}) A_i \geq \sup \hat{G}(A)$ is an upper bound of the objective value for any solution satisfying $A_i \leq \sum_{j<i} \bar{p}_j$ and for all samples, with probability at least $1 - \delta$, we have both

$$|\hat{G}(A^*) - G(A^*)| \leq \frac{\epsilon}{2}$$

and

$$|\hat{G}(\hat{A}) - G(\hat{A})| \leq \frac{\epsilon}{2}.$$  

In order to prove this claim, we note that there always exists an integer optimal solution according to Theorem 1. Thus, given the upper bound on $A_i$ as above, there are at most $(n\bar{p})^n$ possible integer optimal solutions (either for a sampled problem or the stochastic problem itself). And by Lemma 1, for each fixed $A$, the probability that $|\hat{G}(A) - G(A)| \geq \epsilon$ is less than $2 \exp(-\frac{N\epsilon^2}{C^2})$. Then we take a union bound over all the integer solutions and the claim holds.

Therefore, given $N = 4(1/\epsilon)^2 n^5 (o_{\text{max}} + u_{\text{max}})^2 \bar{p}^2 \log (n\bar{p}/\delta)$ samples, with probability at least $1 - \delta$,

$$G(A^*) + \frac{\epsilon}{2} \geq \hat{G}(A^*) \geq \hat{G}(\hat{A}) \geq G(\hat{A}) - \frac{\epsilon}{2},$$

i.e., Theorem 7 holds. □
5 Final Remarks

In this note, we study the appointment scheduling problem with discrete random durations. We extend the study in Begen and Queyranne [2011] to piecewise linear cost functions. The discussion on piecewise linear functions is important since it provides a good approximation (or exact representation in many general cases) to general cost functions. We prove that there exists an integer optimal solution in this case and present a simplified proof using totally unimodularity properties of linear programs. We show that under mild conditions, we can solve this problem in polynomial time. We also discuss a sampling approach for the cases when only a set of samples of the duration times is available. The bound on the number of required samples to achieve a near-optimal solution is competitive.

We believe that the convex programming approach potentially inhabits more merits to help us further understand the structure of the appointment scheduling problem. For example, by the duality theory of the two-stage stochastic programming, we can characterize the subdifferential set of $G(A)$ for piecewise linear cost functions by a simple closed form. For the simplicity of the discussion, we discuss here the linear cost function case. For each given $\omega \in \Omega$ and $A$, the dual form of the first stage problem (15)-(18) is:

$$\max \sum_{i=1}^{n} \{(A_{i+1} + p_{i+1}(\omega))\alpha_i(\omega) + p_{i+1}(\omega)\beta_i(\omega) - (a_i + u_i)p_{i+1}(\omega)\} + p_1\alpha_0(\omega)$$

s.t. $\alpha_i(\omega) + \beta_i(\omega) - \beta_{i+1}(\omega) = a_i + u_i - u_{i+1}$ \forall $i = 1, 2, \ldots, n - 1$ \hspace{1cm} (21)

$\alpha_n(\omega) + \beta_n(\omega) = a_n + u_n$, \hspace{1cm} (22)

$\alpha_0(\omega) - \beta_1(\omega) = -u_1$, \hspace{1cm} (23)

$\alpha_i(\omega), \beta_i(\omega) \geq 0$, \hspace{1cm} $\forall i = 1, 2, \ldots, n$, \hspace{1cm} (24)

Then, by applying Proposition 2.2 and 2.3 in Shapiro et al. [2009], we characterize the subdifferential of $G(A)$ as

$$\partial G(A) = \{\nabla G(A) : \nabla_i G(A) = E_\omega [\alpha_{i-1}^*(\omega)] - \alpha_{i-1}, i = 1, 2, \ldots, n + 1\}$$

where $\alpha_i^*(\omega)$ belongs to the optimal solution set of the dual problem (20)-(24) and $\alpha_0 = 0$.

A simple argument shows that each subgradient is bounded. And all the analysis can naturally extend to the piecewise linear cost function case. These observations may be of interest for further studies of this problem.

References


