Disruption Recovery at Airports: Integer Programming Formulations and Polynomial Time Algorithms

Prabhu Manyem
College of Science, Nanchang Institute of Technology
Nanchang, Jiangxi, P.R. China 330099

September 23, 2017

Abstract

We study disruptions at a major airport. Disruptions could be caused by bad weather, for example. Our study is from the perspective of the airport, the air services provider (such as air traffic control) and the travelling public, rather than from the perspective of a single airline. Disruptions cause flights to be subjected to ground holding or to be cancelled. We present polynomial time algorithms based on the primal-dual schema and show that the algorithms find an optimal solution if the problem is feasible. These algorithms return an optimal mix of which flights to be ground-held and which ones to be cancelled.

Keywords. Air traffic management, Disruption recovery, Discrete (Combinatorial) optimization, Integer programming, Primal-dual schema.

1 Introduction

Disruption Recovery at an airport involves recovery from perturbations to flight schedules caused by factors such as inclement weather. *Ground Holding* is a vital control strategy suggested by several authors for use in disruption recovery [11]. It consists of delaying the departure of a flight from its airport of origin, in anticipation of bad weather (or *reduced arrival capacity* in general) at the destination airport.

Types of disruptions. What types of disruptions to airport operations do we consider? Disruptions (to passenger flow) can be caused by various factors. We focus on disruptions that affect every airline. For example, staffing issues with airline ground staff are those suffered by a single airline, so we do not consider these.

A shortage of immigration officers[1] also causes airport disruption and affects all airlines equally. But we do not consider disruptions such as this. We only consider disruptions to landing and takeoff, such as those due to bad weather and runway(s) being closed for maintenance.

Ground holding is widely practised by the Federal Aviation Administration (FAA) in the United States. In Australia, it is mainly used for flights arriving at Sydney, Brisbane, Melbourne and Perth airports, implemented in a program called Metron [3].

Earlier models. The model developed by Navazio and Romanin-Jacur (the NRJ model) [15] is one of the earliest optimization models for ground holding. It minimizes the total cost of delaying flights, subject to restrictions on arrival capacities at airports. It assumes that the flying time of each flight is fixed, and hence any delay suffered by a flight is due to ground holding at the origin airport. For the sake of completeness the NRJ model is briefly reviewed in Section 2.

Unfortunately, the rather elegant model developed by Navazio and Romanin-Jacur (the NRJ model) does not account for some of the realities of operations at an airport such as Sydney. For example, it does not allow for flights to be cancelled, although such cancellations are not uncommon. Other realities include (a) curfew hours at airports during which time flights are not permitted to land; and (b) the non-linear nature of the costing of flight delays. For example, the total delay cost for the second 30-minute period is higher than that for the first 30-minute period.

We note that some of the publications that appeared in the 1990s on Ground Holding used a linear delay cost model [2, 4, 5, 15].

In recent years, there have been several studies that have studied disruptions from an airline's perspective. See Larsen et al. [6] for a survey. However, as in [9] and [10], we focus on a "common good perspective" that serves the interests of everyone: the airports, the airlines, the air service providers such as air traffic control, and last but certainly not the least, the travelling public.

In Chapter 3 of his Ph.D. thesis, White [19] discusses conditions under which various models (in particular, the linear relaxations of these integer programming models) in air traffic disruption recovery are guaranteed to return optimal solutions in polynomial time.

About a former publication of ours [9], the authors in [6] wrote: "A view which is seldom adopted in the recovery literature is the view of the airport. The paper by Filar et al. [9] describes techniques that enhance the utilization of airport capacities. The paper describes methods involving the traffic management, airport authorities and airlines".

Paper outline. After reviewing the NRJ model in Section 2, we rewrite their model with a different set of variables in Sec. 2.1. In Sec. 3, we discuss an enhanced version of the NRJ model that incorporates cancellations. For both these models, we provide polynomial

[1]For example, at an airport such as Singapore where every flight is an international flight.

time algorithms (in Sec. 2.4 and 3.3 respectively) based on the primal-dual schema and show that the algorithms return optimal solutions.

Our contributions (what is new in this publication): Publications [9] and [10] simply model the problem as an Integer Program and use the CPLEX software to find a solution. However, here, we have developed our own combinatorial algorithms to solve the two problems (single airport ground holding, with and without cancellations). To our knowledge, this is the first known primal dual algorithm for this particular class of disruption recovery problems.

1.1 Simplifications and assumptions

We make the following simplifying assumptions:

1. Linearity assumption: For each flight, the delay cost is linear with respect to time intervals. The cancellation cost is fixed. However in reality, delay and cancellation costs are non-linear with respect to time; see [7, 8, 17, 21].

   For instance, see Figures 7 and 8 in [8]; the authors there remark that after a five-hour delay, all passengers will switch (to a different flight or a different transportation mode). This concurs with our assumption in Sec. 3 that cancellation cost is equal to a delay cost of 4 hours and 30 minutes.

2. Cost coefficients for delay: In reality, the delay/cancellation costs depend on the mix of passengers in each flight and the connections between arriving and departing flights. However, we assume that the delay cost coefficients (delay cost per interval) are proportional[2] to aircraft capacity, and that cancellation cost for a flight is equal to a delay cost of 4 hours and 30 minutes of delay.

3. Cancellation cost: We assume that if a flight is cancelled, it incurs a fixed cost as above. But in reality, the actual cost will depend on the possibility of re-accommodating passengers in other flights.

4. We ignore "feedback" effects such as the total delay (and the costs associated) of a flight (say A) depending on the optimal decisions regarding later flights. For instance, if there is a flight B to the same destination at a later period, then the cost of cancelling A is likely to decrease.

[2]Of course, such costing treats regional airlines (those whose fleet consists **only** of small aircraft) badly all the time. Regional airlines will suffer at the expense of the big carriers. To ensure fairness among all airlines, we can raise the cost coefficient for the smaller aircraft (but still, keep it below the cost factor of the larger aircraft that carry more passengers). To ensure fairness among different airlines, delays should be distributed among the airlines in a random manner (randomised delays), so that the same airline does not suffer large delays on every single day.

5. Correlation between arrival and departure capacities. *Arrival capacity* for a time period (or interval) t is the maximum number of flights allowed to land during the interval. *Departure capacity* for period t is similarly defined. Naturally, there is a correlation between the two. For example, see [12, 13, 18]. However in this study, we only consider arrival capacities; we assume that departure capacities are not a barrier. Furthermore, in reality, fewer arrivals (departures) would allow for more departures (arrivals) respectively. However, we ignore such "feedback" (interaction) effects in this study.

The model and the algorithms can be more readily applied to practical situations once the above realities are incorporated.

Notation

The following notation is used in the NRJ model and in the enhanced model:

T	set of time periods, $T = \{1, \cdots, t, \cdots, \|T\|\}$
Z	set of airports, $Z = \{1, \cdots, z, \cdots, \|Z\|\}$
F	set of all flights considered, $F = \{1, \cdots, f, \cdots, \|F\|\}$
G	set of flights that are flown (not cancelled)
F_z	set of flights arriving at airport z
$F_{z,t}$	set of flights scheduled to arrive at airport z in period t
$K_{z,t}$	arrival capacity of airport z at period t
r_f	planned arrival period of flight f ($r_f \in T$)
θ_f	planned flying time of flight f, in periods
c_f	(linear) cost per period of delay of flight f
μ_f	cancellation cost of f
S_f	set of successors of flight f, $\|S_f\| = n_f$
P_f	set of predecessors of flight f, $\|P_f\| = m_f$
$\sigma_{f,g}$	service time (turn-around time) between flights f and g, in periods
Δ_f	delay experienced by flight f, in periods
A_f	Actual arrival period of flight $f = r_f + \Delta_f$
$\Delta_{\max}$	maximum allowed delay on any flight, in periods
L_f	$= r_f + \Delta_{\max}$
T_f	set of time intervals in which flight f may land $= [r_f, L_f]$
T_F	set of periods for which the capacity is fully utilised.
CS	Complementary slackness

If $t \in T_F$, then the number of flights arriving in period t is equal to the arrival capacity $K_{z,t}$ for that period.

The *service time* $\sigma_{f,g}$ between flights f and g needs some explanation. The arrival of f and the departure of g occurs at the same airport. Flight g is a *successor* of f (and hence f is a *predecessor* of g). Flight g cannot depart before f arrives. A certain amount of service time is necessary between f's arrival and g's departure; for example, some passengers who arrive in f may transfer to g. Or, some crew members from f may transfer to g. Furthermore, if flights f and g use the same aircraft, then an airline requires a certain amount of time after f's arrival to prepare the aircraft for boarding for flight g (typically 15 minutes for a small aircraft and close to an hour for a large one).

2 The NRJ Integer Programming Model

The NRJ model [15], on which our model will be based, is presented below. Time is discretised into intervals (or periods). Each period is typically of 15 or 30 minutes.

We define that $x_{f,t} = 1$ if flight f arrives at or before time period t, and 0 otherwise.

The objective function is

$$\text{minimize} \quad \sum_{f \in F} c_f \left[\sum_{t \in T_f} t(x_{f,t} - x_{f,t-1}) - r_f \right] \tag{1}$$

subject to:

$$\begin{aligned}
-\sum_{f \in F_z} (x_{f,t} - x_{f,t-1}) &\geq -K_{z,t} \;\; \forall z \in Z, \;\; \forall t \in T && (2)\\
x_{f,t} &= 0 \;\; \forall f \in F, \;\; \forall t \in \{1, \cdots, (r_f - 1)\} && (3)\\
x_{f,t} - x_{f,t-1} &\geq 0 \;\; \forall f \in F, \;\; \forall t \in T_f && (4)\\
x_{f,t} &= 1 \;\; \forall f \in F, \;\; \forall t \in \{(r_f + \Delta_{max}), \cdots, |T|\} && (5)\\
x_{f,t} - x_{g,u} &\geq 0 \;\; \forall f \in F, \;\; \forall t \in T_f, \;\; \forall g \in S_f \\
&\quad\;\; \text{and } \forall u \in T_g \;\text{ such that} \\
&\quad\;\; u = t + \sigma_{f,g} + \theta_g && (6)\\
x_{f,t} &\in \{0, 1\} \;\; \forall f \in F, \;\; \forall t \in T && (7)
\end{aligned}$$

Interpretation. The constituent components of the model should be interpreted as follows:

- The objective function (1) is the sum of the delay costs of all flights. For each flight, the delay cost is the product of the number of delay periods and the unit delay cost c_f. We want to minimize the total cost of flight delays.

- The constraints define the feasible region in which the optimal solution is to be found.
- Capacity constraints (2) determine a limit on the number of flights that can land at an airport in each time-period.
- Assignment constraints (3-5) respectively prevent a flight from arriving early, force each flight to land at the airport of destination in only one time-interval, and force a flight to land before the maximum delay for that flight expires.
- Coupling constraints (6) force the flight to land at a certain time prior to the departure time of its successor.
- Constraint (7) forces the $x_{f,t}$ to be binary variables.

2.1 The NRJ model with different variables

Let us modify the decision variables. We now define that $x_{f,t} = 1$ if flight f arrives during time period t, and 0 otherwise. The modified model is presented below.

$$\text{Minimize} \quad \sum_{f \in F} c_f \Delta_f = \sum_{f \in F} c_f \left[\sum_{t \in T_f} t x_{f,t} - r_f \right] = \sum_{f \in F} c_f \sum_{t \in T_f} t x_{f,t} - \sum_{f \in F} c_f r_f. \tag{8}$$

subject to these constraints:

$$\sum_{f \in F_z} -x_{f,t} \geq -K_{z,t} \quad \forall z \in Z, \ \forall t \in T \quad [\alpha_{\mathbf{z},\mathbf{t}}] \tag{9}$$

$$\sum_{t=1}^{r_f - 1} x_{f,t} = 0 \quad \forall f \in F \quad [\beta(\mathbf{f})] \tag{10}$$

$$\sum_{t \in T_f} x_{f,t} = 1 \quad \forall f \in F \quad [\gamma_{\mathbf{f}}] \tag{11}$$

$$\sum_{t=r_f}^{\tau} x_{f,t} - \sum_{u=r_g}^{U} x_{g,u} \geq 0 \quad \forall f \in F, \ \forall \tau \in T_f, \ \forall g \in S_f \text{ and } \forall U \in T_g \text{ such that } U = \tau + \sigma_{f,g} + \theta_g \quad [\mathbf{\Phi}(\mathbf{f}, \tau, \mathbf{g})] \tag{12}$$

$$x_{f,t} \in \{0,1\} \quad \forall f \in F, \ \forall t \in T \tag{13}$$

In the above model, the dual variables are written in brackets besides the corresponding primal constraints. Note that the delay Δ_f of flight f is given by

$$\Delta_f = \sum_{t \in T_f} t x_{f,t} - r_f. \tag{14}$$

If the left side of (12) is positive, this means that $\sum_{t=r_f}^{\tau} x_{f,t} = 1$ and $\sum_{u=r_g}^{U} x_{g,u} = 0$. This means that f arrived in some interval t where $r_f \le t \le \tau$, whereas up to and including interval U, flight g did *not* arrive. Also note that if h is the predecessor of f, then the last constraint (12) can be modified as:

$$\sum_{v=r_h}^{\tau} x_{h,v} - \sum_{t=r_f}^{U} x_{f,t} \ \ge\ 0, \ \ \forall f \in F, \ \ \forall U \in T_f, \ \ \forall h : f \in S_h$$
$$\text{and } \ \forall \tau \in T_h \ \text{ such that}$$
$$\tau = U - \sigma_{h,f} - \theta_f \quad [\mathbf{\Phi}(\mathbf{h}, \tau, \mathbf{f})]. \tag{15}$$

We will use this to model the third constraint (18) in the dual problem. Recall that θ_g is the planned flying time (in periods) of flight g.

Remark 1
(a) The variables in Constraint (10) don't appear in the objective function, nor do they appear in any of the other constraints. Also, their values are meant to be zero. Hence we remove this constraint as well as these variables.

(b) For computational purposes, it is beneficial to use fewer variables. Hence we define $x_{f,t}$ only for $t \in T_f$. For $t \notin T_f$, we leave $x_{f,t}$ undefined. □

2.2 Number of predecessors

Predecessors have more flexibility in terms of arrival times than successors. This is due to the slackness ("buffer time") built into the scheduling. The higher this slackness, the greater the flexibility for the predecessor.

2.2.1 The case of a single predecessor

Let A_f be the actual arrival time of f. Observe that $slack(f, g) = r_g - r_f - \sigma_{f,g} - \theta_g$ is the slack time between f's arrival and g's departure. That is, if f's arrival is delayed by an amount of time $\le slack(f, g)$, the departure of g will *not* be affected.

When a flight g departing from the main airport has only one predecessor f, it is easy to show that the LHS of (12) is always zero. If $r_f \le A_f < r_f + slack(f, g)$, then $U < r_g$; hence (12) is undefined.

If $A_f = r_f + slack(f, g) + i$ for $i \ge 0$, then $A_g = r_g + i$. (Note that i is the delay of f beyond the slack allowance; hence i is also the delay experienced by g.) If $\tau = A_f + j$, then $U = A_g + j$, that is, τ and U differ from the arrival periods of their respective flights by the same amount. Thus the LHS of (12) = 1−1 = 0.

2.2.2 Two predecessors

However if g has two predecessors, say h and f, and suppose h arrives before its slack allowance expires; that is, $A_h \leq r_h + slack(h,g)$. If g's departure period is higher than $t_1 = r_h + slack(h,g) + \sigma_{h,g}$, then for $\tau = r_h + slack(h,g)$ (and correspondingly, $U = \tau + \sigma_{h,g} + \theta_g$), h has arrived by τ, but g has *not* arrived by U, hence the LHS of (12) = $1 - 0 = 1$.

This occurs because, the arrival period of f is such that $A_f + \sigma_{f,g} > t_1$.

2.3 Dual of the LP relaxation to the primal problem (8-13)

As per Remark 1(a), the dual variable $\beta(f)$ is no longer required. For every $f \in F$ and every $t \in T_f$, corresponding to primal variable $x_{f,t}$, we have the following dual constraints:

$$\gamma_f - \alpha_{z,t} \leq tc_f \tag{16}$$

$$\sum_{g \in S_f} \sum_{\tau:(\tau \in [t, L_f],\ U \in T_g)} \Phi(f,\tau,g) + \gamma_f - \alpha_{z,t} \leq tc_f \tag{17}$$

$$-\sum_{h: f \in S_h} \sum_{\tau:(U \in [t, L_f], \tau \in T_h)} \Phi(h,\tau,f) + \gamma_f - \alpha_{z,t} \leq tc_f \tag{18}$$

$$\gamma_f \text{ is unrestricted; } \alpha_{z,t},\ \Phi() \geq 0. \tag{19}$$

where U and τ are related by $U = \tau + \sigma_{f,g} + \theta_g$ in (17) and
$U = \tau + \sigma_{h,f} + \theta_f$ in (18).

Constraint (16): For flights $f \in F$ with neither successors nor predecessors.
Constraint (17): For flights with successors, but no predecessors.
Constraint (18): For flights with predecessors, but no successors. (See Remark 2 below.)

We consider connections at only one airport; hence we ignore cases where a flight can have both successors and predecessors.

Remark 2 *Flights arriving at airports other than the main airport ($z = 1$) are not subject to capacity constraints. Hence the dual variable $\alpha_{z,t}$ does not mean anything for the airports z where such flights arrive. Thus the dual constraints corresponding to variables $x_{f,t}$ for such flights can be rewritten as*

$$-\sum_{h: f \in S_h} \sum_{\tau:(U \in [t, L_f], \tau \in T_h)} \Phi(h,\tau,f) + \gamma_f \leq tc_f \tag{20}$$

An example of such flights would be the flights referred to as "f" in (18). □

2.3.1 The dual objective function

Dual objective function: Maximize $\sum_{z\in Z,\ t\in T}(-K_{z,t}\alpha_{z,t}) + \sum_{f\in F}\gamma_f - \sum_{f\in F} c_f r_f$.

In the process of writing the Lagrangean dual of the primal problem (8-13), the constant term at the end of (8), which is $(-\sum_{f\in F} c_f r_f)$, will be lost. Hence we have manually added this term to the dual objective function above. Thus if both the primal and dual linear programs have optimal solutions, their objective function values will be equal.

By a *single airport problem*, we mean that only at one airport (which we refer to as the "main airport", for which the index z is set to one), capacity constraints are considered.

For a single airport problem, for airports other than the main airport, that is, if $z \neq 1$, we set $\alpha_{z,t} = 0$ since capacity is not an issue at those airports. Hence we can simplify the dual objective as:

$$\text{Maximize} \quad \sum_{t\in T}(-K_{1,t}\alpha_{1,t}) + \sum_{f\in F}(\gamma_f - c_f r_f). \tag{21}$$

Remark 3 *(a) Note that γ_f is unrestricted. It can assume either positive or negative values or zero. The coefficient of γ_f is positive; hence we should increase γ_f as high as possible.*

(b) The coefficient of $\alpha_{z,t}$ is negative; so we should keep the value of these variables as low as possible.

*(c) The variables Φ don't appear in the dual objective function. This means that we can set these variables to any non-negative value we like, at **no** cost. In practice, when we need to set the dual constraints (16-18) to equality, we increase γ_f and decrease $\alpha_{z,t}$ as much as possible before increasing Φ.* □

2.3.2 Complementary slackness (CS)

We will check both primal and dual complementary slackness (CS) conditions. (However, in classical primal-dual algorithms in general, the solution only needs to obey primal feasibility, not primal CS.)

Dual CS:

(a) (Dual constraint is *not* tight) $\Rightarrow$ (Primal variable = 0), and

(b) (Primal variable = 1) $\Rightarrow$ (Dual constraint is tight).

Primal CS:

(c) $(\alpha_{z,t} > 0) \Rightarrow (\sum_{f\in F_z} -x_{f,t} = -K_{z,t})$;

(d) $(\sum_{f\in F_z} -x_{f,t} > -K_{z,t}) \Rightarrow (\alpha_{z,t} = 0)$;

(e) $(\sum_{t=r_f}^{\tau} x_{f,t} - \sum_{u=r_g}^{U} x_{g,u} > 0) \Rightarrow (\Phi(f,\tau,g) = 0)$; and

(f) $(\Phi(f,\tau,g) > 0) \Rightarrow (\sum_{t=r_f}^{\tau} x_{f,t} - \sum_{u=r_g}^{U} x_{g,u} = 0)$.

2.3.3 Flights with neither successors nor predecessors

Let us first consider this simplest case — flights with neither successors nor predecessors. The dual constraint for these flights is (16), which we repeat below:

$$-\alpha_{z,t} + \gamma_f \leq tc_f \;\; \forall f \in F, \;\; \forall t \in T_f. \tag{22}$$

First choose time period 1 ($t = 1$) and airport $z = 1$.

Remember that $\alpha_{z,t}$ should be non-negative and that γ_f is unrestricted. Increasing $\alpha_{z,t}$ will only have the undesirable effect of making (22) a strict inequality. So let us keep $\alpha_{1,t} = \alpha_{1,1} = 0$ and set $\gamma_f = tc_f = c_f$.

To minimize the costs, first choose the flight f (say f_1) with the highest cost coefficient c_{f_1} and set $\gamma_{f_1} = (1)(c_{f_1}) = c_{f_1}$. Next, choose the flight with the second highest cost coefficient (say f_2) and set $\gamma_{f_2} = (1)(c_{f_2}) = c_{f_2}$.

Continue this until all flights scheduled to arrive at $t = 1$ at airport $z = 1$ have been scheduled, or, if the arrival capacity $K_{1,1}$ has been reached, whichever occurs first.

2.3.4 Flights with successors and no predecessors

The constraint is:

$$-\alpha_{z,t} + \gamma_f + \sum_{g \in S_f} \sum_{\tau:(\tau \in [t, L_f], \; U \in T_g)} \Phi(f, \tau, g) \leq tc_f \;\; \forall f \in F, \;\; \forall t \in T_f. \tag{23}$$

If scheduling the arrival of f at t does not affect the scheduling of the arrival of any of its successors, then we treat this case the same as that in Sec. 2.3.3, and hence set $\Phi(f, \tau, g) = 0$.

2.3.5 Flights with predecessors and no successors

Remark 2 applies to these flights, since these arrive at airports other than the main airport (airport $z = 1$). Since there are no limits on arrival capacities at those airports, their scheduling depends only on the arrival of their predecessors at airport 1. This case is handled in Lines 27-35 of Algorithm 1 below.

Considering Remark 3, we should be careful about increasing $\Phi(f, \tau, g)$; it can act as a double-edged sword. The variable appears with a negative sign in the dual constraints such as (18), corresponding to $x_{g,u}$ related to the scheduling of flight g, thus requiring γ_g to be raised to an even higher value.

2.4 The Primal dual algorithm

If we set one of the $x_{f,t}$'s to one, then all other $x_{f,t}$ for this flight f should be set to zero.

Initially set all dual variables to zero. In the classical primal-dual method, one picks the dual constraint with the least value on the right hand side, sets it to an equality, and sets the corresponding primal 0-1 variable to one.

So which $x_{f,t}$ should be set to one first? That is, at first, which flight's arrival is to be scheduled, and during which interval? We should start filling up the airport capacity (the "arrival slots") from the earliest time slot. If we fill Time Period 2 first and Period 1 later, a flight f_1 scheduled to arrive in Period 1 could become delayed, in which case f_1 may have to be delayed even further, to Period 3 or beyond! Hence we start filling the arrival slots from the earliest period.

Note that for a given period t, since t remains fixed, comparing tc_f for different flights is the same as comparing their c_f.

As for dual variables Φ, we follow the suggestion in Remark 3. We initially set them to zero, and then increase them beyond zero only when other options (such as increasing γ_f and decreasing $\alpha_{1,t}$) have been exhausted. (However, it appears that there is a never a need to raise Φ beyond zero.)

See the end of Section 1 for notation. The complete method is provided in Algorithm 1 on Page 11.

Marginal costs. In this algorithm, we have used the marginal cost approach to choose which flights to land and which flights to delay, in a given time period t. Marginal cost considers the following: if a flight is delayed by one time period, what is the effect on the objective function? We choose the flight with a higher cost coefficient over a flight with a lower cost coefficient, even if the total accumulated delay cost of the latter (up to t) has exceeded that of the former.

Algorithm 1 Primal dual algorithm for ground holding for flights arriving at a single airport

procedure MAIN
    for $t = 1$ to T **do** $\text{Flt}[t] = F_{1,t}$;
    for $f = 1$ to $|F|$, $\tau = 1$ to T and $g = 1$ to $|F|$ **do** $\Phi(f, \tau, g) = 0$;
    for $t = 1$ to T **do**
        RemCap = Cap[t]; // Remcap is the remaining capacity
        while (Flt[t] $\neq \emptyset$ & RemCap > 0) **do**
            Let f = the most expensive flight in Flt[t], with cost c_f;
            $x_{f,t} = 1$; // Schedule f to arrive at t
            $A_f = t$;
            RemCap = RemCap -1;
            Flt[t] = Flt[t] $\setminus \{f\}$;
            $\gamma_f = tc_f$;

13: **If** $(t > r_f)$ **then** goBackAdjustGamma$(f,\, t)$;
14: **end while**
15: **if** (RemCap $= 0$ & Flt$[t] \neq \emptyset$ & $t = T$) **then**
16: declare infeasibility and exit;
17: // Flights in Flt$[t]$ can never be scheduled to land, hence the infeasibility
18: **end if**
19: **if** (RemCap $= 0$ & $\exists f \in$ Flt$[t]$ with $t = L_f$) **then**
20: declare infeasibility and exit; // (since f can never be scheduled to land)
21: **end if**
22: **if** (RemCap $= 0$ & Flt$[t] \neq \emptyset$ & $t < T$) **then**
23: // Notice that for every $f \in$ Flt$[t]$, $t < L_f$
24: Flt$[t+1]$ = Flt$[t+1] \cup$ Flt$[t]$;
25: **end if**
26: **end for**
27: // Now schedule arrivals of flights g at airports other than the main airport:
28: **for** every flight $f \in F_z$ where $z \neq 1$ **do**
29: **if** any of f's predecessors have not been scheduled to arrive **then**
30: declare infeasibility and exit;
31: **end if**
32: Let $t_{\min}$ = earliest time that f could arrive, considering the actual arrival times of all of f's predecessors, as per successor constraint (12).
33: **If** $t_{\min} > L_f$, declare infeasibility and exit;
34: $x_{f,t_{\min}} = 1$; // Schedule f to arrive at $t_{\min}$
35: **end for**
36: **end procedure**
37: **procedure** GOBACKADJUSTGAMMA$(\phi,\, \tau)$
38: **while** $\tau > r_\phi$ **do**
39: $\tau = \tau - 1$;
40: Flt$[\tau] = F_{1,\tau}$;
41: $\alpha_{1,\tau} = \max[\gamma_\phi - (\tau - r_\phi)c_\phi,\ 0]$;
42: // The dual constraint corresponding to $x_{\phi,\tau}$ need not be tight.
43: **for** each flight $f \in (Flt[\tau] \ \setminus \ \{\phi\})$ **do**
44: **if** $x_{f,\tau} = 1$ **then**
45: // Flight f has been scheduled to arrive in τ.
46: // The dual constraint should be tight.
47: $\gamma_f = (\tau - r_\phi)c_\phi + \alpha_{1,\tau}$;
48: **end if**
49: **end for**
50: **end while**
51: **end procedure**

Rolledover flight. In Line 22 of Alg. 1, if Flt$[t]$ is non-empty, these flights are considered for arrival in the next time period. Such flights are known as *rolledover* flights of period t.

Complexity of the algorithm. We examine $|F||T|$ constraints as we schedule flights.

For each flight $f \in F$, we may need to go back and adjust $\alpha_{z,t}$ and γ_f at at most $|F||T|$ constraints.

Hence the complexity of the algorithm is $O(|F||T|) + O(|F|^2|T|) = O(|F|^2|T|)$, which is polynomial in $|F|$ and $|T|$.

2.5 Feasibility of Algorithm 1

Line 6 (the *while* condition) ensures that the capacity constraint (9) is always obeyed.

In Lines 22-25, flights that are not yet scheduled in the current interval t are rolled over to the next interval $(t + 1)$. But in Lines 19-21, we ensure that
(a) flights whose arrivals are yet to be scheduled and,
(b) flights that have reached their maximum permitted delay
are *not* rolled over. Thus the algorithm never allows a flight to arrive outside its set T_f of permitted arrival periods.

Furthermore, we declare infeasibility if any flight is unable to be scheduled within T_f (Lines 19-21).

The last two paragraphs ensure that if the algorithm returns a solution, then constraint (11) is obeyed.

Lines 2 and 24 ensure that flights are considered for scheduling only from their published arrival times r_f, not earlier.

2.6 Optimality of Algorithm 1

Using complementary slackness (CS), we now show that Algorithm 1 (Alg. 1) produces an optimal solution.

Dual CS. The *while* loop in the main procedure ensures that whenever $x_{f,t} > 0$, that is, whenever $x_{f,t}$ is set to one (in Line 8), the dual constraint is set to be tight (in Line 12). This satisfies complementary slackness for the dual constraints.

Primal CS. Note that $\alpha_{z,t}$ is raised to a value above zero only in the procedure goBack-AdjustGamma. The procedure sets $\alpha_{z,t}$ to be more than zero only when the corresponding flight f was unable to arrive during period t; this can happen only when the capacity at t $(K_{z,t})$ has been fully utilised.

Thus concerning primal constraint (9), $\alpha_{z,t} > 0 \Rightarrow \sum_{f \in F_z} x_{f,t} = K_{z,t}$.

Hence CS for the primal constraints in (12) can be satisfied.

Theorem 4 *([20], p.23 and pp 163-164). If a solution satisfies primal and dual CS conditions, then the primal (dual) solution is optimal to the primal (dual) optimization problems respectively.* □

Proof of Theorem 4. Theorem 4 is fairly easy to show for a pair of primal and dual problems with linear objective functions and linear constraints. The proof can be found in several books on Mathematical Programming. We outline the proof here. Consider a primal-dual problem pair as follows:

Primal: $Max\ Z_1 = \sum_{j=1}^{n} c_j x_j,\ \sum_{j=1}^{n} a_{ij} x_j \leq b_i\ (1 \leq i \leq m),\ x_j \geq 0\ (1 \leq j \leq n)$.

Dual: $Min\ Z_2 = \sum_{i=1}^{m} b_i y_i,\ \sum_{i=1}^{m} a_{ij} y_i \leq c_j\ (1 \leq j \leq n),\ y_i \geq 0\ (1 \leq i \leq m)$.

CS conditions (x_j^* and y_i^* are primal and dual feasible solutions respectively):

(a) $y_i^*(b_i - \sum_{j=1}^{n} a_{ij} x_j^*) = S_i = 0,\ (1 \leq i \leq m)$ and

(b) $x_j^*(c_j - \sum_{i=1}^{m} a_{ij} y_i^*) = T_i = 0,\ (1 \leq j \leq n)$.

Then we obtain the desired result (that $\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$) by setting $\sum_{i=1}^{m} S_i = 0 = \sum_{j=1}^{n} T_j$. QED.

Corollary. If x_j^* (the primal optimal solution) happens to be an integer for every $j \in [1, n]$, then we have an integer optimal solution to the primal problem.

Since the solution produced by the algorithm satisfies primal and dual CS conditions, from Theorem 4, we conclude that:

Theorem 5 *If a feasible solution exists, then Algorithm 1 returns an optimal feasible solution to model (8-13) in time that is polynomial in the number of flights and the number of time intervals.* □

In Sec. 3.3.2, we describe a different proof to show that if the primal problem (8-13) and the dual problem are feasible, then Algorithm 1 will produce a primal and a dual solution with equal values, and hence these two solutions are primal optimal and dual optimal respectively.

2.7 Modifications to Algorithm 1

Observe that cancelling flights in the algorithm leads to infeasibility. To minimise the chances of this occuring, we suggest the following. If the current time period t equals the last time period during which a flight f can arrive (that is, if $t = L_f$), such flights should be given highest priority for arrival at t.

Since such a policy does not involve cost considerations, it may not lead to an optimal solution.

2.8 Example

We now provide an example for Alg. 1. Recall that r_f is the planned arrival time of f and that $\sigma_{f,\ g}$ is the service time necessary between f's arrival and g's departure. Let us set the maximum delay $\Delta_{max} = 3$.

Dual CS: Recall that if $x_{f,t}$ is set to one, then the corresponding dual constraint (either (16) or (17) or (18)) should be set to an equality.

Arrival capacities in the first three time periods are as follows (at the main airport, $z = 1$):

Period	1	2	3
Arrival capacity	4	5	6
No. of flights scheduled to arrive	5	6	4

Flight f	1	2	3	4	5
Cost c_f	10	20	30	40	50
r_f	1				

Flight f	6	7	8	9	10	11	12	13	14	15	16	$\sigma_{5,\ 16} = 1$
Cost c_f	11	21	31	41	51	61	12	22	32	42	33	$\sigma_{6,\ 16} = 1$
r_f	2						3				8	$\Delta_{max} = 3$
Flight 16 is a successor of two flights (5 and 6). It's flight time $\theta_{16} = 5$.												

Flight 16 departs from Airport 1 and lands at a different airport.

Period 1 (or Interval 1): The arrival capacity is 4. Flights 1-5 will be considered. Out of these, Flights 2-5 will be scheduled for arrival during that period, since these four have higher cost. Set $\gamma_f = tc_f = (1)(c_f) = c_f$ for the four scheduled flights as:

$\gamma_2 = 20$, $\gamma_3 = 30$, $\gamma_4 = 40$, $\gamma_5 = 50$.

Flight 1 will be rolled over for consideration in Period 2.

Dual variable $\Phi(5, 1, 16)$: Since Flt. 5 has a successor (Flt. 16), we need to consider $\Phi(5, \tau, 16)$. When $\tau = 1$, as per (12), $U = 1 + 1 + 5 = 7$. But $U < r_{16} = 8$, hence the coupling constraint (12) and $\Phi(5, \tau, 16)$ are undefined for $\tau = 1$.

However for $2 \leq \tau \leq 5$, we have $8 \leq U \leq 11$. These U values fall within the range of allowed arrival times for Flt 16. Hence constraint (12) and $\Phi(5, \tau, 16)$ *are* defined for $2 \leq \tau \leq 5$.

Period 2: One delayed flight (Flight 1) and Flights 6-11 with $r_f = 2$ will be considered. Their cost factors c_f for these seven flights are 10, 11, 21, 31, 41, 51 and 61 respectively. The arrival capacity is 5. Hence we choose five flights with the highest cost for arrival in Period 2; these are Flights 7-11, with costs 21, 31, 41, 51 and 61 respectively.

Set $\gamma_f = tc_f = 2c_f$ for the five scheduled flights as:

$\gamma_7 = 42$, $\gamma_8 = 62$, $\gamma_9 = 82$, $\gamma_{10} = 102$, $\gamma_{11} = 122$.

Flights 1 and 6 will be rolled over for consideration in Period 3.

For $\Phi(6, \tau, 16)$: $\tau = (2 + i) \Rightarrow U = (2 + i) + 1 + 5 = (8 + i)$ for $0 \leq i \leq 3$. These U values fall within the range of allowed arrival times for Flt 16. Hence constraint (12) and $\Phi(6, \tau, 16)$ *are* defined for $2 \leq \tau \leq 5$.

Period 3: Two delayed flights (Flights 1 and 6) and Flights 12-15 with $r_f = 3$ will be considered. Their cost factors c_f for these six flights are 10, 11, 12, 22, 32 and 42 respectively. The arrival capacity is 6. Hence all considered flights can be scheduled to arrive at $t = 3$. The dual constraint corresponding to $x_{f,t} = x_{6,3}$ should be set to an equality:

$-\alpha_{1,3} + \gamma_6 + \Phi(6, 3.16) + \Phi(6, 4, 16) + \Phi(6, 5.16) = 3c_6 = 33.$

Set $\gamma_f = tc_f = 3c_f$ for the six scheduled flights as:

$\gamma_1 = 30$, $\gamma_6 = 33$, $\gamma_{12} = 36$, $\gamma_{13} = 66$, $\gamma_{14} = 96$, $\gamma_{15} = 126$.

$\alpha_{1,3}$ continues to remain equal to zero. We also leave $\Phi(6, 3.16)$, $\Phi(6, 4, 16)$ and $\Phi(6, 5.16)$ at zero.

Procedure goBackAdjustGamma: Now this procedure will be called twice for Period 2; first for Flight 6, then for Flight 1, since Flight 6 is scheduled first in Period 3 due its higher cost. From the dual constraints, for Period 2:

$$\begin{aligned}
\gamma_1 - \alpha_{1,2} &\leq 2c_1 = 20, \\
\textstyle\sum_{\tau=2}^{5} \Phi(6, \tau, 16) + \gamma_6(= 33) - \alpha_{1,2} &\leq 2c_6 = 22, \\
\gamma_7(= 42) - \alpha_{1,2} &= 2c_7 = 42, \\
\gamma_8(= 62) - \alpha_{1,2} &= 2c_8 = 62, \\
\gamma_9(= 82) - \alpha_{1,2} &= 2c_9 = 82, \\
\gamma_{10}(= 102) - \alpha_{1,2} &= 2c_{10} = 102, \\
\gamma_{11}(= 122) - \alpha_{1,2} &= 2c_{11} = 122.
\end{aligned} \tag{24}$$

Flights 7-11 have been scheduled to land at $t = 2$ at airport $z = 1$, hence the equality above for these five flights. Recall that at this point, γ_1 is still unknown (not set). From the above,

$\alpha_{1,2} \geq \gamma_6 - 22 = 11$. Hence set $\alpha_{1,2} = 11$.

Since $\alpha_{1,2}$ has increased, γ_7, γ_8, γ_9, γ_{10} and γ_{11} should also be increased. The new values are
$\gamma_7 = 53$, $\gamma_8 = 63$, $\gamma_9 = 93$, $\gamma_{10} = 113$ and $\gamma_{11} = 133$.

There is no need to change $\Phi(6, \tau, 16)$ $(2 \leq \tau \leq 5)$ from their zero values.

Note about primal CS: For $\tau = 2$ and $U = 8$, flights 6 and 16 have not arrived by periods τ and U respectively. Thus in (12), $\sum_{t=r_f}^{\tau} x_{f,t} = \sum_{u=r_g}^{U} x_{g,u} = 0$. Hence setting $\Phi(6, 2, 16)$ $= 0$ will satisfy primal complementary slackness. Sec. 2.2.1 is applicable here.

Next, when Procedure goBackAdjustGamma(1, 2) is called after scheduling Flight 1's (the first flight's) arrival, all seven constraints in (24) are satisfied, hence nothing needs to be adjusted for Period 2.

Finally, when Procedure goBackAdjustGamma(1, 1) is called, we consider:

$$\begin{aligned}\gamma_1 - \alpha_{1,1} &\leq c_1 = 10, \\ \gamma_2 - \alpha_{1,1} &= c_2 = 20, \\ \gamma_3 - \alpha_{1,1} &= c_3 = 30, \\ \gamma_4 - \alpha_{1,1} &= c_4 = 40, \\ \Phi(5,2,16) + \Phi(5,3,16) + \Phi(5,4,16) + \gamma_5 - \alpha_{1,1} &= c_5 = 50.\end{aligned} \tag{25}$$

$\alpha_{1,1} \geq \gamma_1 - c_1 = 30 - 10 = 20$. Hence set $\alpha_{1,1} = 20$.

Now γ_2, γ_3, γ_4 and γ_5 should be increased, since $\alpha_{1,1}$ has increased. The new values are $\gamma_2 = 40$, $\gamma_3 = 50$, $\gamma_4 = 60$ and $\gamma_5 = 70$.

Again, there is no need to increase $\Phi(5,2,16)$, $\Phi(5,3,16)$ and $\Phi(5,4,16)$ beyond their zero values.

Note about primal CS: Flt 5 has arrived by $t = 1$ ($\leq \tau = 2$), but as we will see later, Flt 16 has not arrived by interval $U = 8$. Hence the coupling constraint (12) is a strict inequality, and thus setting $\Phi(5,2,16) = 0$ will satisfy primal complementary slackness. Flt 5 plays the role of Flt h in Sec. 2.2.2.

Dual feasibility if delay is zero: Note that these new γ_f values will satisfy the inequalities $\gamma_f - \alpha_{1,t} \leq tc_f$ for $2 \leq f \leq 5$ and $t \geq 2$. For flights with zero delay, if $t > r_f$, the increase in tc_f in the RHS from one period to the next (the increase is just c_f) more than compensates for the increase in $(-\alpha_{1,t})$ in the LHS, thus satisfying dual feasibility.

For any period t, if no flight is rolled over to the following period $(t+1)$, $\alpha_{1,t}$ will remain zero.

Successor: Since flight 6 is delayed by one time period, flight 16 should be delayed by one period; hence flight 16 now departs at period 4, and arrives at period 9. Observe that flight 16's departure is *not* affected by flight 5 whose delay is zero.

Since we set $x_{16,9} = 1$, the corresponding dual constraint (see Remark 2) is set to an equality:

$\Phi(6,2,16) + \Phi(6,3.16) + \Phi(6,4,16) + \Phi(6,5.16) + \gamma_{16} = tc_{16} = (9)(33) = 297.$

Nothing prevents us from setting γ_{16} equal to 297. In fact it is beneficial to the dual objective function, as per Remark 3. Hence $\Phi(6,\tau,16)$ for $2 \leq \tau \leq 5$ can remain equal to zero.

The final delay values are displayed below:

Flight f	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Cost c_f	10	20	30	40	50	11	21	31	41	51	61	12	22	32	42	33
r_f	1					2						3				8
Delay Δ_f	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

The final value of the objective function = Total delay cost =
$= c_1\Delta_1 + c_6\Delta_6 + c_{16}\Delta_{16} = (10)(2) + (11)(1) + (33)(1) = 64$ units.

$\alpha_{1,1} = 20$, $\alpha_{1,2} = 11$, and $\alpha_{1,3} = 0$.

3 Model with ground holding and cancellation

3.1 The primal problem

Flight cancellation enhancement to the NRJ model was introduced in [10]. To model cancellation, we introduce a variable fly_f for each flight f. If the flight is completed, fly_f has a value of one, otherwise it is zero.

We need two additional constraints to handle cancellation (though the second of them is just a modification of an existing constraint in the previous model).

(i) **Cancellation deciding constraint**. The decision as to whether a flight is cancelled is based on the delay potentially suffered by the aircraft. If the delay would be so large that the total delay cost of f would exceed μ_f, the cost of cancelling the flight, then it is more cost effective to cancel the flight. This is captured by the following constraint:

$$\mu_f fly_f - c_f\Delta_f \geq 0. \tag{26}$$

So as to set the delay to zero when a flight is cancelled, we modify the definition of Δ_f in (14) as follows:

$$\Delta_f = \sum_{t\in T_f} t x_{f,t} - r_f fly_f. \tag{27}$$

Substituting for Δ_f in (26), we get

$$\mu_f fly_f - c_f \left(\sum_{t\in T_f} t x_{f,t} - r_f fly_f \right) = (\mu_f + c_f r_f) fly_f - c_f \sum_{t\in T_f} t x_{f,t} \geq 0. \tag{28}$$

If the flight is cancelled, that is when $fly_f = 0$, if follows that $x_{f,t} = 0$ for every time interval $t \in T$, according to (37) below. Hence the delay Δ_f will also be zero as per (27) and the cancellation constraint (26) is satisfied. On the other hand, if the flight is flown, that is when $fly_f = 1$, then $x_{f,t} = 1$ for exactly one interval $t \in T_f$, according to (37).

Note that the unit delay cost c_f is assumed constant for any amount of delay Δ_f. Hence the total delay cost $c_f\Delta_f$ increases linearly with Δ_f.

If the flight is flown, then from (26), $\Delta_f \leq \left\lceil \frac{\mu_f}{c_f} \right\rceil$. Let $\Delta_{\max} = \left\lceil \frac{\mu_f}{c_f} \right\rceil$. Hence $L_f = r_f + \Delta_{\max} = r_f + \left\lceil \frac{\mu_f}{c_f} \right\rceil$. Then we could define T_f, the set of intervals when f could arrive (if f is flown), as follows:

$$T_f = [r_f,\ L_f] = \left[r_f,\ r_f + \left\lceil \frac{\mu_f}{c_f} \right\rceil \right]. \tag{29}$$

(ii) **Constraint connecting fly_f with $x_{f,t}$.** Constraint (5) in the NRJ model (Section 2) is not imposed here, since we do not require all flights to be flown. However, if the flight *is* flown, then $\sum_{t=r_f}^{L_f} x_{f,t} = 1$, since the flight lands at some time interval in $[r_f, L_f]$. Conversely if the flight is cancelled, $\sum_{t=r_f}^{L_f} x_{f,t} = 0$ since the flight never arrives. Thus the modified constraint (5) is:

$$fly_f = \sum_{t=r_f}^{L_f} x_{f,t}. \tag{30}$$

The objective function for the new model incorporating ground holding and cancellations is:

$$\text{Minimize} \ \ P_o = \sum_{f \in F} c_f \left[\sum_{t \in T_f} t x_{f,t} - r_f fly_f \right] + \sum_{f \in F} (1 - fly_f)\mu_f. \tag{31}$$

The first (second) summation refers to the ground holding (cancellation) costs respectively. Rearranging in terms of the variables, we get:

$$\text{Minimize} \ \ P_o = \sum_{f \in F} \mu_f - \sum_{f \in F} fly_f \left(c_f r_f + \mu_f \right) + \sum_{f \in F} \sum_{t \in T_f} x_{f,t} \left(t c_f \right) \tag{32}$$

As in Sec. 2.3.1, we will add the constant term $\sum_{f \in F} \mu_f$ to the dual objective function.

The constraints for the primal problem are:

$$\sum_{f \in F_z} -x_{f,t} \geq -K_{z,t} \;\; \forall z \in Z, \;\; \forall t \in T \;\; [\alpha_{\mathbf{z,t}}] \tag{33}$$

$$\sum_{t=1}^{r_f - 1} x_{f,t} = 0 \;\; \forall f \in F \;\; [\beta(\mathbf{f})] \tag{34}$$

$$\sum_{t=r_f}^{\tau} x_{f,t} - \sum_{u=r_g}^{U} x_{g,u} \geq 0 \;\; \forall f \in F, \;\; \forall \tau \in T_f, \;\; \forall g \in S_f$$
$$\text{and } \forall U \in T_g \text{ such that}$$
$$U = \tau + \sigma_{f,g} + \theta_g \;\; [\mathbf{\Phi}(\mathbf{f}, \tau, \mathbf{g})] \tag{35}$$

$$(\mu_f + c_f r_f) fly_f - c_f \sum_{t \in T_f} t x_{f,t} \geq 0 \;\; [\epsilon_{\mathbf{f}}] \tag{36}$$

$$\sum_{t=r_f}^{L_f} x_{f,t} - fly_f = 0 \;\; [\gamma_{\mathbf{f}}] \tag{37}$$

$$fly_f, \; x_{f,t} \in \{0,1\} \;\; \forall f \in F, \;\; \forall t \in T \tag{38}$$

Remark 1(a) also applies to Constraint (34) and the variables that appear in that constraint. Similarly, Remark 1(b) also applies to the model above.

Remark 6 *It is known that total unimodularity (T.U.) of the constraint matrix is a sufficient condition to guarantee that the optimal solution returned by the LP relaxation is also an integer solution (see, for example, [14]). (A T.U. matrix is one for which the determinant of every square sub-matrix is either +1 or −1 or zero.) However note that the constraint matrix in (32-38) is not T.U. due to coefficients such as c_f and r_f in (36).* □

3.2 Dual of the LP relaxation to the primal problem (32-38)

The dual of the LP relaxation to the above primal model is as follows. For every $f \in F$ and every $t \in T_f$, corresponding to primal variable $x_{f,t}$, we have these three constraints:

$$\gamma_f - \alpha_{z,t} - tc_f\epsilon_f \leq tc_f \tag{39}$$

$$\sum_{g \in S_f} \; \sum_{\tau:(\tau \in [t, L_f], \; U \in T_g)} \Phi(f, \tau, g) + \gamma_f - \alpha_{z,t} - tc_f\epsilon_f \leq tc_f \tag{40}$$

$$-\sum_{h: f \in S_h} \; \sum_{\tau:(U \in [t, L_f], \tau \in T_h)} \Phi(h, \tau, f) + \gamma_f - \alpha_{z,t} - tc_f\epsilon_f \leq tc_f \tag{41}$$

$$\gamma_f \text{ is unrestricted}; \;\; \alpha_{z,t}, \;\; \epsilon_f, \;\; \Phi() \geq 0. \tag{42}$$

where U and τ are related by $U = \tau + \sigma_{f,g} + \theta_g$ in (40) and
$U = \tau + \sigma_{h,f} + \theta_f$ in (41).

Remark 7 *Constraint (39): For flights $f \in F$ with neither successors nor predecessors.*
Constraint (40): For flights with successors, but no predecessors.
Constraint (41): For flights with predecessors, but no successors. □

The constraint corresponding to the primal varible fly_f is:

$$-\gamma_f + (\mu_f + c_f r_f)\epsilon_f \leq -(\mu_f + c_f r_f) \tag{43}$$

Remark 2 applies here, but with a different constraint:

$$-\sum_{h:f\in S_h} \sum_{\tau:(U\in[t,L_f],\tau\in T_h)} \Phi(h,\tau,f) + \gamma_f - tc_f\epsilon_f \leq tc_f \tag{44}$$

3.2.1 The dual objective function

The dual objective function is as follows:

$$\text{Maximize } D_o = \sum_{f\in F} \mu_f - \sum_{z\in Z,\ t\in T} K_{z,t}\alpha_{z,t}. \tag{45}$$

For a single airport problem, for airports other than the main airport, that if $z \neq 1$, we set $\alpha_{z,t} = 0$ since capacity is not an issue at those airports. Hence we can simplify the dual objective as:

$$\text{Maximize } \ D_o = \sum_{f\in F} \mu_f - \sum_{t\in T} K_{1,t}\alpha_{1,t}. \tag{46}$$

3.2.2 Complementary slackness

The dual CS conditions are the same as in Sec. 2.3.2.

As for the primal CS conditions, in addition to those listed in Sec. 2.3.2, we need the following in relation to the primal constraint (36):

$[(\mu_f + c_f r_f)fly_f - c_f \sum_{t\in T_f} tx_{f,t} > 0] \ \Rightarrow \ [\epsilon_f = 0]$ and

$[\epsilon_f > 0] \ \Rightarrow \ [(\mu_f + c_f r_f)fly_f - c_f \sum_{t\in T_f} tx_{f,t} = 0]$.

If ϵ_f is set to zero, then these two conditions will be satisfied.

Remark 8 *We will set μ_f (the cancellation cost) sufficiently high such that when the flight is flown, constraint (36) is satisfied with a strict inequality (and hence in order to satisfy the corresponding complementary slackness, ϵ_f should equal zero). For example, the maximum delay could be set to 4 hours, whereas μ_f could be set equal to a delay cost corresponding to 4 hours and 30 minutes of delay.* □

3.3 Primal dual algorithm for cancellations

Let G be the set of flights that are flown (*not* cancelled), and A_f be the arrival time of f.

Let us start from the primal objective function value:

$$P_o = \sum_{f\in F} \mu_f - \sum_{f\in F} fly_f \left(c_f r_f + \mu_f\right) + \sum_{f\in F} \sum_{t\in T_f} x_{f,\ t} \left(t c_f\right) \tag{47}$$

Naturally we will consider only those flights f for which $fly_f = 1$ (second term in (47)) and those $\{f,\ t\}$ for which $x_{f,\ t} = 1$ (third term in (47)). That is, $f \in G$ and $t = A_f$. Recalling that $x_{f,\ t} = 1$ for at most one period t (that is, $t = A_f$), we write:

$$P_o = \sum_{f\in F} \mu_f - \sum_{f\in G} \left(c_f r_f + \mu_f\right) + \sum_{f\in G} A_f c_f. \tag{48}$$

Applying dual CS condition to constraint (43), when $fly_f = 1$, we know that
$-\gamma_f + (\mu_f + c_f r_f)\epsilon_f = -(\mu_f + c_f r_f)$. From this, we obtain

$$(\mu_f + c_f r_f) = \frac{\gamma_f}{1+\epsilon_f}. \tag{49}$$

When $x_{f,\ t} = 1$, the third term in (48) will be equal to the right side of (39) or (40) or (41), depending on the type of flight (see Remark 7).

In constraints (40) and (41), if we set $\Phi(f,\tau,g)$ and $\Phi(h,\tau,f)$ to zero as before, and setting $t = A_f$, then $tc_f = A_f c_f = \gamma_f - \alpha_{z,A_f} - A_f c_f \epsilon_f$ for all three types of flights mentioned in Remark 7. Simplifying,

$$A_f c_f = \frac{\gamma_f - \alpha_{z,A_f}}{1+\epsilon_f}. \tag{50}$$

Applying (49) and (50) to (48),

$$P_o = \sum_{f\in F} \mu_f - \sum_{f\in G} \frac{\gamma_f}{1+\epsilon_f} + \sum_{f\in G} \frac{\gamma_f - \alpha_{z,A_f}}{1+\epsilon_f} = \sum_{f\in F} \mu_f - \sum_{f\in G} \frac{\alpha_{z,A_f}}{1+\epsilon_f}. \tag{51}$$

Suppose we set $\epsilon_f = 0$ (see Remark 8). Then

$$P_o = \sum_{f\in F} \mu_f - \sum_{f\in G} \alpha_{z,A_f} = \sum_{f\in F} \mu_f - \sum_{t\in T} \sum_{\{f\,|\,A_f=t\}} \alpha_{z,A_f} \tag{52}$$

where the last equality is obtained by rearranging the summation in terms of the time periods.

Algorithm 2 Primal dual algorithm for ground holding and cancellations for flights arriving at a single airport

---

procedure Main
  for $t = 1$ to T **do** $\text{Flt}[t] = F_{1,t}$;
  for $f = 1$ to $|F|$ **do** $\epsilon_f = 0$;
  for $f = 1$ to $|F|$, $\tau = 1$ to T and $g = 1$ to $|F|$ **do** $\Phi(f, \tau, g) = 0$;
  for $t = 1$ to T **do**
    RemCap = Cap[t]; // Remcap is the remaining capacity
    while ($\text{Flt}[t] \neq \emptyset$ & RemCap > 0) **do**
      Let f = the most expensive flight in Flt[t], with cost c_f;
      $x_{f,t} = 1$; // Schedule f to arrive at t
      $A_f = t$;
      RemCap = RemCap -1;
      $\text{Flt}[t] = \text{Flt}[t] \setminus \{f\}$;
      $\gamma_f = tc_f$;
      If ($t > r_f$) **then** goBackAdjustGamma(f, t);
    end while
    if (RemCap $= 0$ & $\text{Flt}[t] \neq \emptyset$ & $t = T$) **then**
      For each flight f in Flt[t] set $fly_f = 0$;
      // Flights in Flt[t] can never be scheduled to land, hence cancelled
    end if
    if (RemCap $= 0$ & $\exists f \in \text{Flt}[t]$ with $t = L_f$) **then**
      for every such flight f in Flt[t] (with $t = L_f$) **do**
        Set $fly_f = 0$ and remove f from Flt[t];
      end for // (f can never be scheduled to land, hence cancelled)
    end if
    if (RemCap $= 0$ & $\text{Flt}[t] \neq \emptyset$ & $t < T$) **then**
      // Notice that for every $f \in \text{Flt}[t]$, $t < L_f$
      $\text{Flt}[t+1] = \text{Flt}[t+1] \cup \text{Flt}[t]$;
    end if
  end for
  // Now schedule arrivals of flights g at airports other than the main airport:
  for every flight $f \in F_z$ where $z \neq 1$ **do**
    if any of f's predecessors have been cancelled **then** set $fly_f = 0$
    else
      Let $t_{\min}$ = earliest time that f could arrive, considering the actual arrival
        times of all of f's predecessors, as per successor constraint (35).
      if $t_{\min} > L_f$ **then** set $fly_f = 0$
        // (that is, if any predecessor arrives too late, then ancel f)
      else $x_{f,t_{\min}} = 1$; // Schedule f to arrive at $t_{\min}$
      end if
    end if
  end for
end procedure
procedure goBackAdjustGamma(ϕ, τ)

```
    // (The same as in Algorithm 1.)
end procedure
```

Alg. 2 is similar to Alg. 1, the difference being that while the former declares infeasibility if a flight f is unable to be scheduled, the latter cancels f and retains feasibility. In fact, the primal problem described in Sec. 3.1 always has a feasible solution, just by cancelling every flight (though of course, this will be far from optimal).

Complexity. The complexity of Alg. 2 is the same as that of Alg. 1, that is, $O(|F||T|) + O(|F|^2|T|) = O(|F|^2|T|)$, polynomial in $|F|$ and $|T|$.

Note that in Alg. 2, α_{z,A_f} is set as a positive quantity only if the capacity constraint has been reached during period A_f. (Also as per (33), for primal complementary slackness, $\sum_{f\in F_z} x_{f,t} = K_{z,t}$ is a necessary condition for $\alpha_{z,t}$ to be set as a positive quantity, *not* a sufficient condition.)

3.3.1 Algorithm Analysis

If we follow Alg. 2, it turns out that the primal and dual objective functions ((52) and (46), respectively) are equal.

Remark 9 $\alpha_{z,m} > 0$ *if and only if there exists a flight* f *such that we were unable to schedule* f *in period* m, *but we were able to schedule* f*'s arrival for a later period* n > m*; that is, at least one* rolledover *flight from* m *has been scheduled to land. If all such rolledover flights for a period* m *are eventually cancelled, then the procedure* goBackAdjustGamma *would never go back to* m *to adjust* $\alpha_{z,m}$. □

Furthermore, this only applies to flights scheduled to arrive at the main airport under consideration, airport 1. That is, if $z \neq 1$, $\alpha_{z,t} = 0 \; \forall t \in T$.

In particular, Remark 9 applies to those flights f that are not cancelled and $A_f > r_f$ (suffer a delay). For such flights, $\alpha_{z,t} > 0$ for $t \in [r_f, \, A_f - 1]$.

We now know that a necessary condition for $\alpha_{1,t}$ being positive is that the number of flights arriving in period t should be equal to the capacity $K_{1,t}$. That is, the capacity is fully utilised and $|\{f \mid A_f = t\}| = K_{1,t}$.

Let T_F be the set of periods for which the capacity is fully utilised. Then:

$$t \in T_F \implies \sum_{f:\, A_f = t} \alpha_{1,A_f} = \sum_{f:\, A_f = t} \alpha_{1,t} = K_{1,t}\alpha_{1,t}. \tag{53}$$

For those other periods t for which $|\{f : A_f = t\}| < K_{1,t}$, that is, if $t \in T \setminus T_F$, we have $\alpha_{1,t} = 0$. Thus

$$\sum_{t\in T} \sum_{\{f:\, A_f = t\}} \alpha_{1,t} = \sum_{t\in T_F} \sum_{\{f:\, A_f = t\}} \alpha_{1,t} + \sum_{t\in (T\setminus T_F)} \sum_{\{f:\, A_f = t\}} \alpha_{1,t}$$

$$= \sum_{t \in T_F} K_{1,t}\alpha_{1,t} + 0 = \sum_{t \in T_F} K_{1,t}\alpha_{1,t} + \sum_{t \in (T \setminus T_F)} K_{1,t}\alpha_{1,t} = \sum_{t \in T} K_{1,t}\alpha_{1,t}. \quad (54)$$

Hence the primal objective function (52) can be rewritten as

$$P_o = \sum_{f \in F} \mu_f - \sum_{t \in T} \sum_{\{f:\, A_f = t\}} \alpha_{1,A_f} = \sum_{f \in F} \mu_f - \sum_{t \in T} \sum_{\{f:\, A_f = t\}} \alpha_{1,t}$$

$$= \sum_{f \in F} \mu_f - \sum_{t \in T} K_{1,t}\alpha_{1,t}, \quad (55)$$

which is precisely the dual objective function in (46). Hence we have shown that

Theorem 10 *Assuming the conditions in Remark 8, Algorithm 2 produces feasible optimal solutions for the primal problem (32-38) as well as for the dual problem (39-46) in time that is polynomial in the number of flights and the number of time intervals.* □

3.3.2 Applying the analysis of Sec. 3.3.1 to Sec. 2.1 (no cancellations)

The primal-dual analysis in Sec. 3.3.1 can be applied to Alg. 1 in Sec. 2.1, where cancellations are not allowed. It can be shown that Alg. 1 returns an optimal feasible solution if a feasible solution exists.

We set $fly_f = 1$ for every flight $f \in F$ in the primal model given by (32-38). The analysis then proceeds as in Sec. 3.3.1. We leave it to the reader to fill in the details.

3.4 How efficient are our algorithms?

In a theoretical sense, our algorithms are combinatorial and greedy, and run in polynomial time (and hence can be termed as *efficient*). Instead, one could certainly solve their Integer Programming (IP) formulations or their linear relaxations. The linear relaxation is a Linear Program (LP), so it will run in polynomial time; but will it always return integer solutions? We don't know, since there is no clear indicator such as a T.U.M. (see Remark 6). The I.P. will return integer solutions, obviously; but will it run in polynomial time? Again, we don't know; we don't have a proof yet. This depends on the formulation itself and the method to solve the IP. For problems solvable in polynomial time, in general, greedy approaches will return optimal solutions (our algorithms are indeed greedy). See for example, Chapter 12 of [16].

In a practical sense, software such as CPLEX are very good at solving IP quickly and efficiently. But so are greedy combinatorial algorithms. Many network flow problems are good examples; they can be solved in polynomial time as linear relaxations of IP, but combinatorial algorithms work faster [1].

A key advantage with our method is that we can obtain "partial solutions" up to a certain time period as the algorithm proceeds; whereas with solving the IP/LP formulation, one has to wait for the algorithm to fully halt before obtaining the schedule of *any* flight.

4 Further research

(a) The issues mentioned in Sec. 1.1 should be attended to; in particular developing algorithms for a non-linear delay cost model. One could begin by modelling the delay cost in a piecewise linear fashion. In practice, every passenger will have a unique delay cost structure with respect to time, depending on the type of passenger. Hence we need to aggregate these and determine a single delay-cost diagram for a particular flight that departs a particular airport at a particular time of the week (a Monday morning 10:30 AM flight from Sydney to Melbourne, for instance), over a planning period of say, one year.

(b) Sydney airport has a curfew period from 23:00 hours to 06:00 hours each night. During this period, flights may land, but only if they pay a penalty. Neither the NRJ model nor the models developed in this paper account for such a scenario. We plan to incorporate this in future research.

(c) Diversions to other airports due to bad weather is another common practice in the aviation industry.

(d) Capacity constraints at more than one airport (a network of airports) should be considered. Ideally both arrival and departure capacities (and the manner in which the two correlate) should be considered. To some extent, this is already being carried out in the *Metron* system implemented in Australia.

Both (b) and (c) have been modelled before [9]. However, in future research, for features that include ground holding, cancellations and those mentioned above in (a) and (b), we plan to explore the possibility of developing algorithms that either return optimal solutions or guarantee a solution that is within a fixed factor of the optimal one.

(e) It appears that the Φ's can always remain at zero level with no effect on dual feasibility. We need a mathematical proof of this.

5 Acknowledgements

I am grateful to a research grant from the Nanchang Institute of Technology which funded this research. I thank Phil Dowson at Airservices Australia for updating me on the status of ground holding in Australia in 2016. I benefitted from discussions with my colleagues Bingrong Wan, Lu Jiang and Hai-xiong Yu. A preliminary version of this paper was presented at a talk in ECCO XXIX (29th conference of the European Chapter on Combinatorial Optimization) held in Budapest, Hungary in May 2016.

References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows: Theory, Algorithms and Applications*. Prentice Hall, 1993.

[2] G. Andreatta, L. Brunetta, and G. Guastella. Multi Airport Ground Holding Problem: A Heuristic Approach based on Priority Rules. In P. Dell'Olomo L. Bianco and A.R. Odoni, editors, *Modelling and Simulation in Air Traffic Management*, pages 71–89. Springer Verlag, 1997.

[3] Airservices Australia. Air Traffic Flow Management: Ground Delay Programs. Available as `http://www.airservicesaustralia.com/projects/collaborative-decision-making-cdm/air-traffic-flow-management`, May 2017.

[4] D. Bertsimas and S.S. Patterson. The Air Traffic Flow Management with Enroute Capacities. *Operations Research*, 46(3):406–422, May-June 1998.

[5] L. Brunetta, G. Guastalla, and L. Navazio. Solving the Multi Airport Ground Holding Problem. *Annals of Operations Research*, 81:271–287, 1998.

[6] J. Clausen, A. Larsen, J. Larsen, and N.J. Rezanova. Disruption management in the airline industry: Concepts, models and methods. *Networks*, 37:809–821, 2010.

[7] Andrew Cook, Graham Tanner, and Stephen Anderson. *Evaluating the true cost to airlines of one minute of airborne or ground delay: final report*. Eurocontrol, 2004.

[8] Andrew Cook, Graham Tanner, and Adrian Lawes. The Hidden Cost of Airline Unpunctuality. *Journal of Transport Economics and Policy*, 46(2):157–173, May 2012.

[9] J.A. Filar, P. Manyem, D.M. Panton, and K. White. A Model for Adaptive Rescheduling of Flights in Emergencies (MARFE). *Journal of Industrial and Management Optimization*, 3(2):335–356, May 2007.

[10] J.A. Filar, P. Manyem, M.S. Visser, and K. White. Air Traffic Management at Sydney with Cancellations and Curfew Penalties. In P. Pardalos and V. Korotkich, editors, *Optimization and Industry: New Frontiers*, pages 113–140. Kluwer Academic Publishers, March 2003.

[11] J.A. Filar, P. Manyem, and K. White. How Airlines And Airports Recover From Schedule Perturbations: A Survey. *Annals of Operations Research*, 108:315–333, November 2001.

[12] E. P. Gilbo. Optimizing Airport Capacity Utilization in Air Traffic Flow Management Subject to Constraints at Arrival and Departure Fixes. *IEEE Transactions on Control Systems Technology*, 5(5):490–503, 1997.

[13] Eugene Gilbo and Michael West. Enhanced model for joint optimization of arrival and departure strategies and arrival/departure capacity utilization at congested airports. In *AIAA Modeling and Simulation Technologies (MST) Conference*, page 4599, 2013.

[14] Dorit Hochbaum, editor. *Approximation Algorithms for NP-Hard Problems*. PWS Publishing Company, Boston, MA, 1997.

[15] L. Navazio and G. Romanin-Jacur. The Multiple Connections Multi Airport Ground Holding Problem: Models and Algorithms. *Transportation Science*, 32(3):268–276, August 1998.

[16] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, New Jersey, USA, 1982.

[17] Everett B Peterson, Kevin Neels, Nathan Barczi, and Thea Graham. The economic cost of airline flight delay. *Journal of Transport Economics and Policy (JTEP)*, 47(1):107–121, 2013.

[18] Varun Ramanujam and Hamsa Balakrishnan. Estimation of arrival-departure capacity tradeoffs in multi-airport systems. In *Proceedings of the 48th IEEE Conference on Decision and Control 2009 held jointly with the 28th Chinese Control Conference. CDC/CCC 2009.*, pages 2534–2540. IEEE, 2009.

[19] Kevin White. *Ground Holding Optimisation and Air Schedule Recovery*. PhD thesis, University of South Australia (School Of Mathematics), July 2011.

[20] David P. Williamson and David B. Shmoys. *The Design of Approximation Algorithms*. Cambridge University Press, 2010.

[21] Duojia Yuan. *Ground Holding Optimisation and Air Schedule Recovery*. PhD thesis, RMIT University (Australia), February 2007.