# A Dynamic Programming Decomposition Method for Making Overbooking Decisions over an Airline Network

Alexander Erdelyi
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
alex@orie.cornell.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
topaloglu@orie.cornell.edu

July 23, 2009

# Online Supplement

## A Online Supplement: Proof of Proposition 2

To simplify the proof, we introduce auxiliary value functions $\{\psi^i_t(\cdot) : i \in \mathcal{L},\ t \in \mathcal{T}\}$ by letting

$$\psi^i_t(x_t) = \sum_{j \in \mathcal{J}} p_{jt} \max\{F^i_j + \psi^i_{t-1}(x_t + e_j), \psi^i_{t-1}(x_t)\} + \Big[1 - \sum_{j \in \mathcal{J}} p_{jt}\Big] \psi^i_{t-1}(x_t) \tag{O.1}$$

with the boundary condition that $\psi^i_0(x_0) = -\mathbb{E}\{\phi^i(S(x_0))\}$, where

$$\phi^i(S(x_0)) = \min \quad \sum_{j \in \mathcal{J}} \Theta^i_j\, w_j \tag{O.2}$$

$$\text{subject to} \quad \sum_{j \in \mathcal{J}} a_{ij}\, [S_j(x_{j0}) - w_j] \le c_i \tag{O.3}$$

$$w_j \le S_j(x_{j0}) \qquad j \in \mathcal{J} \tag{O.4}$$

$$w_j \in \mathbb{Z}_+ \qquad j \in \mathcal{J}. \tag{O.5}$$

The following two results provide the intermediate steps to prove Proposition 2.

**Lemma 3** *For all $t \in \mathcal{T}$, we have*

$$\psi^i_t(x_t) = V^i_t(\mathcal{R}^i(x_t)) + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^i} q_j\, [-\Theta^i_j]^+\, x_{jt} + \sum_{s=1}^{t} \sum_{j \in \mathcal{J} \setminus \mathcal{J}^i} p_{js} \max\{F^i_j, F^i_j - q_j\, \Theta^i_j, 0\}.$$

**Proof of Lemma 3** We show the result by induction over the time periods. Noting the upper bounds on the decision variables $\{w_j : j \in \mathcal{J} \setminus \mathcal{J}^i\}$ in problem (O.2)-(O.5), the optimal values of these decision variables are $\{\mathbf{1}(\Theta^i_j \le 0)\, S_j(x_{j0}) : j \in \mathcal{J} \setminus \mathcal{J}^i\}$. Thus, since we have $a_{ij} = 0$ for all $j \in \mathcal{J} \setminus \mathcal{J}^i$, we have

$$\phi^i(S(x_0)) = \Gamma^i(\mathcal{R}^i(S(x_0))) - \sum_{j \in \mathcal{J} \setminus \mathcal{J}^i} [-\Theta^i_j]^+\, S_j(x_{j0}),$$

where $\Gamma^i(\mathcal{R}^i(S(x_0)))$ is the optimal objective value of problem (28)-(31). Taking expectations in the expression above and noting that $S_j(x_{j0})$ has a binomial distribution with parameters $(x_{j0}, q_j)$, we obtain $\psi^i_0(x_0) = -\mathbb{E}\{\phi^i(S(x_0))\} = -\mathbb{E}\{\Gamma^i(\mathcal{R}^i(S(x_0)))\} + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^i} q_j\, [-\Theta^i_j]^+ x_{j0} = V^i_0(\mathcal{R}^i(x_0)) + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^i} q_j\, [-\Theta^i_j]^+ x_{j0}$ and the result holds for the last time period. Assuming that the result holds for time period $t-1$, we have

$$\psi^i_{t-1}(x_t + e_j) - \psi^i_{t-1}(x_t) = \begin{cases} V^i_{t-1}(\mathcal{R}^i(x_t + e_j)) - V^i_{t-1}(\mathcal{R}^i(x_t)) & \text{if } j \in \mathcal{J}^i \\ q_j\, [-\Theta^i_j]^+ & \text{if } j \in \mathcal{J} \setminus \mathcal{J}^i. \end{cases} \tag{O.6}$$

Therefore, we have

$$\begin{aligned} \psi^i_t(x_t) &= \sum_{j \in \mathcal{J}} p_{jt} \max\{F^i_j + \psi^i_{t-1}(x_t + e_j) - \psi^i_{t-1}(x_t), 0\} + \psi^i_{t-1}(x_t) \\ &= \sum_{j \in \mathcal{J}^i} p_{jt} \max\{F^i_j + V^i_{t-1}(\mathcal{R}^i(x_t + e_j)) - V^i_{t-1}(\mathcal{R}^i(x_t)), 0\} \\ &\qquad + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^i} p_{jt} \max\{F^i_j + q_j\, [-\Theta^i_j]^+, 0\} + V^i_{t-1}(\mathcal{R}^i(x_t)) \\ &\qquad\qquad + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^i} q_j\, [-\Theta^i_j]^+\, x_{jt} + \sum_{s=1}^{t-1} \sum_{j \in \mathcal{J} \setminus \mathcal{J}^i} p_{js} \max\{F^i_j, F^i_j - q_j\, \Theta^i_j, 0\}, \end{aligned}$$

where the first equality follows from (O.1) and the second equality follows from (O.6) and the induction assumption. Since $\max\{F^i_j + q_j\,[-\Theta^i_j]^+, 0\} = \max\{F^i_j, F^i_j - q_j\,\Theta^i_j, 0\}$, the result follows by collecting the terms on the right side of the expression above and noting the definition of $V^i_t(\mathcal{R}^i(x_t))$ in (27). □

**Lemma 4** *For all $t \in \mathcal{T}$, we have $V_t(x_t) \le \psi^i_t(x_t) - \sum_{j\in\mathcal{J}} q_j\,\Lambda^i_j\,x_{jt} + \sum_{l\in\mathcal{L}\setminus\{i\}} \lambda^*_l\,c_l$.*

**Proof of Lemma 4** We show the result by induction over the time periods. We let $\{w^*_j : j \in \mathcal{J}\}$ be the optimal solution to problem (1)-(4). We have

$$\begin{aligned}\Gamma(S(x_0)) = \sum_{j\in\mathcal{J}} \theta_j\,w^*_j &\ge \sum_{j\in\mathcal{J}} \theta_j\,w^*_j + \sum_{l\in\mathcal{L}\setminus\{i\}} \lambda^*_l \Big\{ \sum_{j\in\mathcal{J}} a_{lj}\,[S_j(x_{j0}) - w^*_j] - c_l \Big\} \\ &= \sum_{j\in\mathcal{J}} \Theta^i_j\,w^*_j + \sum_{j\in\mathcal{J}} \Lambda^i_j\,S_j(x_{j0}) - \sum_{l\in\mathcal{L}\setminus\{i\}} \lambda^*_l\,c_l \\ &\ge \phi^i(S(x_0)) + \sum_{j\in\mathcal{J}} \Lambda^i_j\,S_j(x_{j0}) - \sum_{l\in\mathcal{L}\setminus\{i\}} \lambda^*_l\,c_l, \qquad (\text{O.7})\end{aligned}$$

where the first inequality follows from the fact that the solution $\{w^*_j : j \in \mathcal{J}\}$ satisfies constraints (2) and $\lambda^*_l \ge 0$ for all $l \in \mathcal{L}\setminus\{i\}$, the second equality follows from (13) and the second inequality follows from the fact that $\{w^*_j : j \in \mathcal{J}\}$ is a feasible but not necessarily an optimal solution to problem (O.2)-(O.5). Taking expectations in the expression above, we obtain $V_0(x_0) = -\mathbb{E}\{\Gamma(S(x_0))\} \le -\mathbb{E}\{\phi^i(S(x_0))\} - \sum_{j\in\mathcal{J}} q_j\,\Lambda^i_j\,x_{j0} + \sum_{l\in\mathcal{L}\setminus\{i\}} \lambda^*_l\,c_l \;=\; \psi^i_0(x_0) - \sum_{j\in\mathcal{J}} q_j\,\Lambda^i_j\,x_{j0} + \sum_{l\in\mathcal{L}\setminus\{i\}} \lambda^*_l\,c_l$ and the result holds for the last time period. Assuming that the result holds for time period $t-1$, the induction assumption immediately implies that

$$\begin{aligned}\max\{f_j + V_{t-1}(x_t + e_j), V_{t-1}(x_t)\} \le \max\{f_j + \psi^i_{t-1}(x_t + e_j) - q_j\,\Lambda^i_j, \psi^i_{t-1}(x_t)\} \\ - \sum_{j\in\mathcal{J}} q_j\,\Lambda^i_j\,x_{jt} + \sum_{l\in\mathcal{L}\setminus\{i\}} \lambda^*_l\,c_l.\end{aligned}$$

Recalling that $F^i_j = f_j - q_j\,\Lambda^i_j$, one can combine the inequality above with (5) and (O.1) to obtain the result for time period $t$. □

We are now ready to finalize the proof of Proposition 2. Lemmas 3 and Lemma 4 imply that

$$\begin{aligned}V_t(x_t) \le V^i_t(\mathcal{R}^i(x_t)) + \sum_{j\in\mathcal{J}\setminus\mathcal{J}^i} q_j\,[-\Theta^i_j]^+\,x_{jt} - \sum_{j\in\mathcal{J}} q_j\,\Lambda^i_j\,x_{jt} \\ + \sum_{j\in\mathcal{J}\setminus\mathcal{J}^i} \sum_{s=1}^{t} p_{js} \max\{F^i_j, F^i_j - q_j\,\Theta^i_j, 0\} + \sum_{l\in\mathcal{L}\setminus\{i\}} \lambda^*_l\,c_l.\end{aligned}$$

The result follows by noting that the sum of the second and third terms on the right side of the expression above can be written as

$$\begin{aligned}\sum_{j\in\mathcal{J}\setminus\mathcal{J}^i} q_j\,[-\Theta^i_j]^+\,x_{jt} - \sum_{j\in\mathcal{J}} q_j\,\Lambda^i_j\,x_{jt} &= \sum_{j\in\mathcal{J}\setminus\mathcal{J}^i} q_j\,\max\{-\theta_j + \Lambda^i_j, 0\}\,x_{jt} - \sum_{j\in\mathcal{J}\setminus\mathcal{J}^i} q_j\,\Lambda^i_j\,x_{jt} - \sum_{j\in\mathcal{J}^i} q_j\,\Lambda^i_j\,x_{jt} \\ &= -\sum_{j\in\mathcal{J}\setminus\mathcal{J}^i} q_j\,\min\{\theta_j, \Lambda^i_j\}\,x_{jt} - \sum_{j\in\mathcal{J}^i} q_j\,\Lambda^i_j\,x_{jt} \\ &= -\sum_{j\in\mathcal{J}\setminus\mathcal{J}^i} \min\Big\{q_j\,\theta_j, q_j \sum_{l\in\mathcal{L}} a_{lj}\,\lambda^*_l\Big\}\,x_{jt} - \sum_{j\in\mathcal{J}^i} q_j\,\Lambda^i_j\,x_{jt},\end{aligned}$$

where the last equality follows from the fact that $\Lambda^i_j = \sum_{l \in \mathcal{L} \setminus \{i\}} a_{lj} \, \lambda^*_l$ and $a_{ij} = 0$ whenever $j \in \mathcal{J} \setminus \mathcal{J}^i$.

## B Online Supplement: Description of the Data Files

The data files that we use in our computational experiments are provided as an online supplement. The goal of this section is to describe the format of the data files. All of our data files are labeled as `rm_A_B_C.C_D.D_E.E_F.F.txt`, where `A` corresponds to the number of spokes in the airline network, `B` corresponds to the fare difference between a high fare and its corresponding low fare itinerary, `C.C` and `D.D` correspond to the parameters that we use to compute the penalty cost, `E.E` corresponds to the probability that a reservation shows up at the departure time and `F.F` corresponds to the ratio of the total expected demand to the total expected capacity. In other words, following the notation in Section 5.1, `A`, `B`, (`C.C`,`D.D`), `EE` and `F.F` respectively correspond to $N$, $\kappa$, $(\delta, \sigma)$, $q$ and $\rho$.

In all of our data sets, we assume that we serve $N$ spokes out of a single hub. Location 0 corresponds to the hub and locations $\{1, \ldots, N\}$ correspond to the spokes. The itineraries that connect the hub to a spoke or a spoke to the hub include one flight leg. The itineraries that connect two spokes include two flight legs, one from the origin spoke to the hub and one from the hub to the destination spoke.

Table O.1 shows the organization of the data file for a test problem with $\tau = 3$ and $N = 2$. The character "#" indicates a comment line and such lines are skipped. The entries in the five portions of the data file have the following interpretations. The first portion of the data file shows the number of time periods in the planning horizon. The second portion of the data file shows the flight legs in the airline network. The first line in this portion shows the number of flight legs. After this first line, each line corresponds to one flight leg and shows the origin location, destination location and capacity of the flight leg. The third portion of the data file shows the itineraries. The first line in this portion shows the number of itineraries. After this first line, each line corresponds to one itinerary and shows the origin location, destination location, fare level, revenue and penalty cost for the itinerary. Fare level 0 indicates a low fare itinerary and fare level 1 indicates a high fare itinerary. We emphasize that the itineraries that connect two spokes include two flight legs, one from the origin spoke to the hub and one from the hub to the destination spoke. The fourth portion of the data file shows the arrival probabilities for the requests for different itineraries. Each line in this portion corresponds to a time period in the planning horizon. Each line first shows an itinerary indicated by the triplet [origin location, destination location, fare level], followed by the probability that we observe a request for this itinerary. For example, the probability that we observe a request for the low fare itinerary from location 2 to 1 at the first time period is 0.2. Since we may not observe any itinerary arrivals at a particular time period, the probabilities in a particular line do not necessarily add up to one. The fifth portion of the data file shows the show up probabilities. Each line in this portion corresponds to one itinerary. Each line first shows an itinerary indicated by the triplet [origin location, destination location, fare level], followed by the probability that a reservation for this itinerary shows up at the departure time.

```
# beginning of data file
# portion 1
# number of time periods in decision horizon
3

# portion 2
# list of flights [in format origin location, destination location, capacity]
# first line is number of flights
4
1 0 16
2 0 21
0 1 12
0 2 20

# portion 3
# list of itineraries [in format origin location, destination location, fare level, revenue, penalty cost]
# first line is number of itineraries
7
0 1 0 24.0 48.0
0 1 1 192.0 384.0
0 2 0 34.0 68.0
1 0 0 192.0 384.0
1 2 0 53.0 106.0
2 1 0 53.0 106.0
2 1 1 212.0 442.0

# portion 4
# list of request arrival probabilities [in format itinerary, probability]
# first entry in each line indicates time period
0  [0 1 0] 0.1    [0 1 1] 0.1    [0 2 0] 0.1    [1 0 0] 0.1    [1 2 0] 0.1    [2 1 0] 0.2    [2 1 1] 0.1
1  [0 1 0] 0.1    [0 1 1] 0.1    [0 2 0] 0.1    [1 0 0] 0.1    [1 2 0] 0.1    [2 1 0] 0.1    [2 1 1] 0.1
2  [0 1 0] 0.1    [0 1 1] 0.1    [0 2 0] 0.1    [1 0 0] 0.1    [1 2 0] 0.1    [2 1 0] 0.1    [2 1 1] 0.1

# portion 5
# list of show up probabilities [in format itinerary, probability]
[0 1 0] 0.9
[0 1 1] 0.9
[0 2 0] 0.9
[1 0 0] 0.9
[1 2 0] 0.9
[2 1 0] 0.9
[2 1 1] 0.9

# end of data file
```

Table O.1: Organization of the data file for a test problem with $\tau = 3$ and $N = 2$.