A new approach to solve the probabilistic nonlinear seat inventory control problem

Andreea Popescu   Ellis Johnson   Pinar Keskinocak

School of Industrial and Systems Engineering

Georgia Institute of Technology

765 Ferst Drive, Atlanta, GA 30332

andreeap@isye.gatech.edu

ellis.johnson@isye.gatech.edu

pinar@isye.gatech.edu

Abstract

The seat inventory control problem is to maximize expected revenue of future scheduled flights by allocating seats to different itinerary / fare classes. We solve the probabilistic version using the Lagrangian dual formulation. We show that the Lagrangian dual can be solved by using a trust region based Newton method, in which the gradient and the Hessian are computed efficiently based on exploiting information from the primal representation. Special cases and schemes to adjust the dual and the primal variables are discussed.

1 Introduction

Revenue management aims to maximize profits generated from a limited capacity of a product by selling it to the right customer for the right price at the right time [1]. One of the successful application areas of revenue management is the airline industry. Seat inventory control across different fare classes is one of the key problems of revenue management in the airline industry. Seat inventory control is concerned with allocating seats to different fare classes in order to maximize the expected revenue of future scheduled flights.

There are several criteria used to differentiate among approaches for seat inventory control. There are open/closed or limit based control policies, and single-leg, multiple legs or origin-destination (OD) control policies. A thorough review of existing approaches is given in [2].

In this paper, we are proposing a new algorithm for solving the seat inventory control with itinerary specific demand using *bid prices.* Bid prices are threshold values used to accept / deny incoming booking requests. Suppose a customer requests one seat for a fare of f on a certain itinerary. The decision maker uses given values, the bid prices, for each leg comprised in the requested itinerary and accepts the request if the sum of the bid prices along the itinerary is lower than the proposed fare. The bid prices are usually computed each time a booking request comes in, by solving a mathematical program. Disadvantages and advantages of this approach are discussed in Williamson [3].

There are basically variations on 3 types of mathematical programs used in the literature for solving the seat inventory problem. There is a deterministic linear program that uses mean values of the demand on different itineraries to capture the probabilistic nature of passenger demand, a nonlinear program with linear capacity constraints that uses direct information on demand distribution in the objective function, and dynamic programs. In this paper we consider the nonlinear program with linear capacity constraints. A comprehensive overview of existing models is given in [4].

The next section introduces the problem and its dual. Calculations of the main ingredients of the proposed algorithm are presented in section 3. The algorithm and discussions on how to handle special cases follow subsequently. The paper is concluded by numerical results and further directions.

2 The probabilistic nonlinear seat inventory control problem and its dual

An airline's flight network for a given departure date can be modeled by a graph. Each node represents a departure / arrival time and city and each arc a flight leg. The arc capacities $b_1(t)$, ..., $b_L(t)$ of legs l_1, ..., l_L at time t are equal to the number of open seats on leg l at time t. The number of open seats on leg l at time t is the difference between the capacity of leg l and the number of seats sold on this leg up to time t. For simplicity, we assume cancellations and no-shows do not occur.

An arriving customer specifies a fare class $r \in F$ and an itinerary $p \in P$. A demand class is uniquely defined by a customer type $j = (r, p)$.

The information about demand classes can be summarized in an itinerary-leg incidence matrix A. Each column represents a demand class and each row a flight leg. An entry a_{ij} is equal to 1 if there is a request for leg $i \in$ itinerary p associated with demand class $j = (r, p)$. Otherwise $a_{ij} = 0$.

For example, consider the three nodes in the flight network, A, B and C. Suppose the following itineraries are requested: A-B, A-C, B-C, and A-B-C. For each itinerary, there are two fare classes: Y and M. Altogether, we have eight possible itinerary-demand class combinations. They are summarized in the following matrix:

$$\begin{array}{c} \\ \text{A-B} \\ \text{A-C} \\ \text{B-C} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{(A-B)Y} & \text{(A-C)Y} & \text{(B-C)Y} & \text{(A-B-C)Y} & \text{(A-B)M} & \text{(A-C)M} & \text{(B-C)M} & \text{(A-B-C)M} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

Before describing the mathematical model, we introduce the notation used:

Given:

- A m x n matrix with m number of legs and n demand classes;
- A_j jth column (incidence vector for demand class j);
- A^i ith row (fare classes on leg i);
- r_j fare associated with demand class j;
- b capacity vector for all legs;
- D non-negative random variable for demand with known pdf $\phi(t)$, continuous and differentiable, cdf $\Phi(t)$;

Variables and objective:

- x partitioned allocation of capacity for each of the n demand classes;
- $E[min(x_j, D_j)]$ expected sales of demand class j under partition x;

Problem formulation:

$$(PNLP) \quad \begin{array}{ll} max & \sum_{j=1}^{n} r_j \cdot E[min(x_j, D_j)] \\ \text{s.t.} & A \cdot x \leq b \\ & x \geq 0 \end{array}$$

The objective function is the expected future revenue under partition x, where the seat allocation fulfills the capacity constraints. The expected revenue under seat allocation x_j for a given demand class j is either the demand or the seat allocation for that class, whichever is lower. The problem is a relaxed problem of the real world situation, where the variables x are integral.

We denote the expected future revenue corresponding to allocation x_j for demand class j by f_j. We have:

$f_j(x_j) = r_j \cdot E[min(x_j, D_j)] = r_j \cdot \int_0^{x_j} t \cdot \phi(t)dt + r_j \cdot \int_{x_j}^{\infty} x_j \cdot \phi(t)dt.$

The PNLP is:

$$(PNLP) \quad \begin{array}{ll} max & \sum_{j=1}^{n} f_j(x_j) \\ \text{s.t.} & A \cdot x \leq b \\ & x \geq 0 \end{array}$$

Following the traditional definition, we use as bid prices the dual variables of the capacity constraints. In the following, we show how to formulate the Lagrangian dual based on the special structure of the PNLP .

It is easy to show that the PNLP has a concave separable objective function:

$f_j'(x_j) = r_j \cdot (1 - \Phi(x_j));$

$f_j''(x_j) = -r_j \cdot \phi(x_j) \leq 0 \ \forall j.$

The Lagrangian dual of the PNLP, DPNLP, is:

(DPNLP) $\max_{\pi \geq 0} \; q(\pi)$

where $q(\pi) = \inf_{x \geq 0} L(x, \pi)$ (1)

with (2) $L(x, \pi) = -\sum_{j=1}^{n} f_j(x_j) + \pi \cdot (A \cdot x - b)$ the Lagrangian function for the PNLP (we define π as a row vector).

By convexity, the minimum x_j^* in (1) exists and satisfies:

$\nabla[-\sum_{j=1}^{n} f_j(x_j^*)] + \pi \cdot A = 0$, from properties of the Lagrange multipliers. We have:

$-f_j'(x_j^*) + \pi \cdot A_j = 0 \; \forall j = 1, \ldots, n$, by separability of the objective function. We obtain:

$f_j'(x_j^*) = \pi \cdot A_j \; \forall j = 1, \ldots, n$ (3)

By investigating $f_j'(x_j) = r_j \cdot \int_{x_j}^{\infty} \phi(t)dt$, we deduce that x_j^* must satisfy:

$$\begin{aligned} f_j'(x_j^*) &= \pi \cdot A_j, \quad && \text{if } r_j = f_j'(0) > \pi \cdot A_j \\ x_j^* &= 0, && \text{if } r_j = f_j'(0) \leq \pi \cdot A_j \quad (4) \end{aligned}$$

The condition simply says that if the sum of the bid prices for the j-th itinerary / fare class exceeds the fare, set the corresponding allocation on 0, which is consistent with the bid price policy.

Plug in (4) into (1) and (2):

$q(\pi) = L(x^*, \pi) = -\sum_{j=1}^{n} f_j(x_j^*) - \pi \cdot b + \sum_{j=1}^{n} x_j^* \cdot f_j'(x_j^*) = -\pi \cdot b - \sum_{j=1}^{n} [f_j(x_j^*) - x_j^* \cdot f_j'(x_j^*)]$

We obtain the following expression for the DPNLP:

$$\begin{array}{llll} & v = \min\limits_{\pi \geq 0} & \pi \cdot b + \sum_{j=1}^{n} [f_j(x_j^*) - x_j^* \cdot f_j'(x_j^*)] & \\ (DPNLP) & x_j^* \text{ s.t.} & f_j'(x_j^*) = \pi \cdot A_j, & \text{if } r_j > \pi \cdot A_j \\ & & x_j^* = 0, & \text{if } r_j \leq \pi \cdot A_j \end{array}$$

3 The gradient and the Hessian calculations of the DPNLP

We aim to apply a trust region based Newton method to solve the DPNLP. The need of Hessian updates at each step makes Newton methods computationally expensive. In this section, we show that the Hessian updating for the DPNLP can be made efficiently using the properties of the primal variables for which the Lagrangian function of the PNLP attends the minimum.

It is easy to show that the gradient of the DPNLP objective function is the constraint violation of the PNLP formulation:

$\frac{\partial v}{\partial \pi_i} = b_i + \frac{\partial}{\partial \pi_i}[\sum_{j=1}^n (f_j(x_j^*) - x_j^* \cdot f_j'(x_j^*))] = b_i + \sum_{j=1}^n [f_j'(x_j^*) \cdot \frac{\partial x_j^*}{\partial \pi_i} - x_j^* \cdot \frac{\partial}{\partial \pi_i} f_j'(x_j^*) - f_j'(x_j^*) \cdot \frac{\partial x_j^*}{\partial \pi_i}] = b_i - \sum_{j=1}^n x_j^* \cdot \frac{\partial}{\partial \pi_i} f_j'(x_j^*)$ (5)

Since $f_j'(x_j^*) = \pi \cdot A_j$ if $r_j > \pi \cdot A_j$, we have $\frac{\partial}{\partial \pi_i} f_j'(x_j^*) = a_{ij}$ for all j such that $r_j > \pi \cdot A_j$.

For the remaining part of this section, we assume $r_j > \pi \cdot A_j$, $\forall j$. Handling of cases for which this condition does not hold are discussed in the next section.

We obtain:

$$\frac{\partial v}{\partial \pi_i} = b_i - \sum_{j=1}^n a_{ij} \cdot x_j^* \quad \forall\ i = 1...m, \text{ with } \ x_j^* \text{ s.t. } f_j'(x_j^*) = \pi \cdot A_j \qquad (6)$$

which is the constraint violation of the PNLP.

Based on the gradient, we proceed to calculate the Hessian of v in the DPNLP:

$\frac{\partial^2 v}{\partial \pi_i \partial \pi_k} = -\sum_{j=1}^n a_{ij} \cdot \frac{\partial x_j^*}{\partial \pi_k}$ (7).

We need to calculate $\frac{\partial x_j^*}{\partial \pi_k}$. We know: $f_j'(x_j^*) = \pi \cdot A_j$, and by defining $h_j = (f_j')^{-1}$, we deduce:

$x_j^* = h_j(\pi \cdot A_j) = (f_j')^{-1}(\pi \cdot A_j)$.

Then:

$\frac{\partial x_j^*}{\partial \pi_k} = \frac{\partial h_j(A_j^T \cdot \pi)}{\partial \pi_k} = h_j'(A_j^T \cdot \pi) \cdot \frac{\partial (A_j^T \cdot \pi)}{\partial \pi_k} = \frac{1}{f_j''(x_j^*)} \cdot a_{kj}$ (8), where the last equality comes from the following theorem from calculus:

Theorem: If f has an inverse f^{-1}, is differentiable at $f^{-1}(x)$ and $f'(f^{-1}(x)) \neq 0$, then f^{-1} is differentiable at x and the following formula holds:
$\frac{d}{dx} f^{-1}(x) = \frac{1}{f'(f^{-1}(x))}$

Plug in (8) into (7), and we obtain:

$$\frac{\partial^2 v}{\partial \pi_i \partial \pi_k} = -\sum_{j=1}^{n} a_{ij} \cdot \frac{a_{kj}}{f_j''(x_j^*)} = \sum_{j=1}^{n} \frac{a_{ij} \cdot a_{kj}}{r_j \cdot \phi(x_j^*)} \qquad \forall i, k = 1...m \qquad (9)$$

Based on (6), and (9), we design a trust region based Newton method to solve the DPNLP. The following section presents the algorithm and discusses its applicability.

4 Solution methodology

The k-th iteration of the algorithm is:

1. For π^k calculate ${x_j^*}^k$:
   - if $\pi^k \cdot A_j = 0$ then set ${x_j^*}^k$ on the 99-th percentile of the demand distribution;
   - if $\pi^k \cdot A_j \geq r_j$ then set ${x_j^*}^k = 0$; delete column j from matrix A and continue with $A_{reduced}$;
   - otherwise set ${x_j^*}^k = \Phi^{-1}(1 - \frac{\pi^k \cdot A_j}{r_j})$.
2. delete null row i from $A_{reduced}$ and set $\pi_i^{k+1} = 0$ for all null rows i in $A_{reduced}$;
3. calculate the gradient and the Hessian for the reduced problem based on (6) and (9) respectively;
4. solve the trust region subproblem to find the descent direction

   $d^k = \arg \min_{||d|| \leq \Delta^k} v(\pi^k) + \nabla v(\pi^k)' \cdot d + \frac{1}{2} d' \cdot \nabla^2 v(\pi^k) \cdot d$

where Δ^k is some positive scalar (see e.g., [5] on how to determine the value of Δ^k and find the descent direction d^k);

5. if $v(\pi^k) + \nabla v(\pi^k)' \cdot d^k + \frac{1}{2} d^{k'} \cdot \nabla^2 v(\pi^k) \cdot d^k = v(\pi^k)$ STOP (π^k is the minimum for DPNLP);
6. otherwise set $\pi^{k+1} = \pi^k + d^k$;
7. for all i with $\pi^{k+1} < 0$, set $\pi^{k+1} = 0$;
8. set $k = k + 1$ and go to the next iteration.

The algorithm mainly uses the findings from the last section. There are some special cases that need further discussion:

1. By deleting columns in A in step 1 of the algorithm, we might create null rows in the reduced matrix, which simply means that the corresponding capacity constraints read $0 \leq b_i$ for some rows i. By setting the corresponding dual variables on 0, we ensure complementarity slackness for that capacity constraints and feasibility of the dual variables.
2. We have to ensure nonnegativity of the bid prices resulted after each iteration of the algorithm. By setting negative values of the bid prices on 0, we allow a positive or negative slack in the corresponding constraint.

   Conform (4), we have $x_j^{*^k} = \Phi^{-1}(1 - \frac{\pi^k \cdot A_j}{r_j})$; however, whenever $\pi \cdot A_j = 0$, we have an infinite value for $x_j^{*^k}$. By setting it on the 99-th percentile of the demand distribution we might violate the primal constraint, but we might enforce a positive value on the corresponding bid prices, and consequently lower allocations and eventually primal feasibility, in the next iterations.

Since the algorithm terminates based on dual optimality only, there might be necessary to use some schemes (like a Phase I linear program) to recover primal feasibility and adjust the bid prices accordingly.

5 Conclusions and further directions

The algorithm is still in its incipient phase. It was coded in Matlab and tested first on a small network with 3 legs, 6 routes, 10 demand classes, using a truncated normal demand distribution function. The algorithm successfully solved the problem when compared to the solution obtained using Matlab's nonlinear solver. When applied to a larger network, with 50 legs, 100 origin and destinations, and 150 routes, the algorithm took 14 seconds on a Intel Pentium 4, 2.40 GHz CPU, 1.00 GB RAM, and then the solution had to be adjusted for primal feasibility.

More tests need to be done. We are currently working on fine tuning the solution when flat region demand distributions are used. The issue has been addressed briefly in section 4, but more research needs to be done in order to ensure the best quality of the solution.

References

[1] K. Pak and N. Piersma, *Airline Revenue Management: An Overview of OR Techniques: 1982 - 2001*, Econometric Institute Report EI 2002-2003

[2] J. I. McGill and G. J. van Ryzin, *Revenue Management: Research Overview and Prospects*, Management Science, Vol. 33, No. 2, p.233-256, 1999

[3] E. L. Williamson, *Airline network seat control*, Ph. D. Thesis, MIT, Cambridge, MA, 1992

[4] K. T. Talluri and G. J. van Ryzin, *The theory and practice of revenue management*, Kluwer Academic Publishers, Boston/Dordrecht/London, 2004

[5] D. B. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Mass., 1995