A decision support system for forecasting and optimal procurement

Neven Miculnic

May 5, 2016

ABSTRACT

Optimal procurement in most industries involve forecasting of two quantities: prices of raw materials and customer's demand. The aim of this work is to integrate forecasts into production planning models, with the aim of minimizing overall procurement, holding and production costs under demand satisfaction constraints. The decision support system should allow the decision maker to integrate qualitative, unstructured information through a simple interface, for scenario selection and solution refinement

1

PROBLEM DEFINITION

1.1 INTRODUCTION

This problem is variation of lot sizing problem in stochastic setting. Variation is as follows. We have stochastic prediction for raw material supply cost $\mathbf{s}$ at future time moment t. The *factory* converts all bought raw materials at time t and stores them into a warehouse. Number of discrete time moments is denoted by n.

One time moment storage costs fixed amount h, that is the holding cost.

Likewise demand for the product is equally stochastic, denoted as $\mathbf{d}$. In case there's no products in storage to satisfy demand we pay backlogging cost denoted as b.

Our aim is to optimize our procurement policy by variating $\mathbf{x}$, that is product amount we buy at time moment t. However we are constrained by the maximum raw materials we can buy each day by $\mathbf{x}_{\text{max}}$.

1.2 FORMAL PROBLEM DEFINITION

1.2.1 *Definitions*

Every input value is known beforehand, and it's assumed for random variables their distributions are known. Following is deterministic problem variant, and in subsequent chapters randomness and uncertainty is embedded into problem. Font convention:

x, y, z	variables and constants
$\mathbf{x}, \mathbf{y}, \mathbf{z}$	vector variables and constants
X, Y, Z	random variables
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$	random vectors
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	probability distributions
$\mathbf{A}, \mathbf{B}, \mathbf{C}$	matrices, by context differentiated from random vectors

Now following is formal problem description for deterministic variant:

$\mathbf{s} = [s_1, s_2, \ldots, s_n]^\mathsf{T}$	Supply cost vector
$\mathbf{x} = [x_1, x_2, \ldots, x_n]^\mathsf{T}$	Procurement quantity vector
$\mathbf{x}_{\text{max}}$	Procurement quantity limits vector
$\mathbf{d} = [d_1, d_2, \ldots, d_n]^\mathsf{T}$	Demand random vector, Y_t is a random variable
b	backlogging cost
h	holding cost

1.2.2 *Variables*

$\mathbf{x}$ is our decision variable, as described previously.

1.2.3 *Constraints*

$$\mathbf{x} \in \mathbb{N}_0^n$$
$$\mathbf{x} \leq \mathbf{x}_{\text{max}}$$
$$\sum x = \sum d$$
$$b \geq 0$$
$$h \geq 0$$

1.3 OBJECTIVE FUNCTION

Definition 1.3.1. $c(t)$ defines total speeding we pay at time t.

$$c(t) = +s_t x_t + \begin{cases} h\left(\sum_{i=1}^{t} x_i - \sum_{i=1}^{t} d_i\right) & \sum_{i=1}^{t} x_i > \sum_{i=1}^{t} d_i \\ b\left(\sum_{i=1}^{t} d_i - \sum_{i=1}^{t} x_i\right) & \sum_{i=1}^{t} x_i < \sum_{i=1}^{t} d_i \end{cases} \tag{1}$$

Definition 1.3.2. $f(\mathbf{x})$ is objective function for this problem. Our aim is to minimize it.

$$f(\mathbf{x}) = \sum_{t} c(t) \tag{2}$$

2

DETERMINISTIC APPROACH

2.1 MODELLING APPROACHES

Problem as defined in 1 can be reduced to two well known problems: transportation and min-cost max flow problem. In both cases knowning $\mathbf{x}$ implies all needed network parameters given $b, h > 0$.

2.1.1 *Transportation problem reduction*

First we consider auxiliary variables x'_{ij} which represent amount assigned from supply at time i to demand at time t. This entails additional constraint:

$$x_i = \sum_j x'_{ij}$$

Definition 2.1.1. $\mathbf{C}$ matrix defines cost for satisfying demand with specific raw supply material purchase date. It's element c_{ij} equals:

$$c_{ij} = \begin{cases} b\,(i-j) + s_i & j < i \\ h\,(j-i) + s_i & j \geq i \end{cases}$$

That is using purchaces raw materials at i to satisfy demand at time moment j incurs cost c_{ij}.

Finally our original cost function 1 then equals to:

$$c(t) = \sum_i x'_{it} c_{it} \tag{3}$$

Under those modeling approaches optimal x'_{ij} yields optimal $\mathbf{x}$

Transportation problem[1] is easy reduction since we have cost matrix $\mathbf{C}$ defining "transportation" costs associated with each possible assignment option.

1 Frederick S Hillier. *Introduction to Operations Research*. McGraw-Hill, 2014. ISBN: 0073523453.

For successful reduction we only need adding dummy source or destination as described in[2]

2.1.2 *Min cost max flow reduction*

We can exploit additional problem structure to achieve superior performance and modeling capabilities. In figure 2.1.2 we see network architecture.

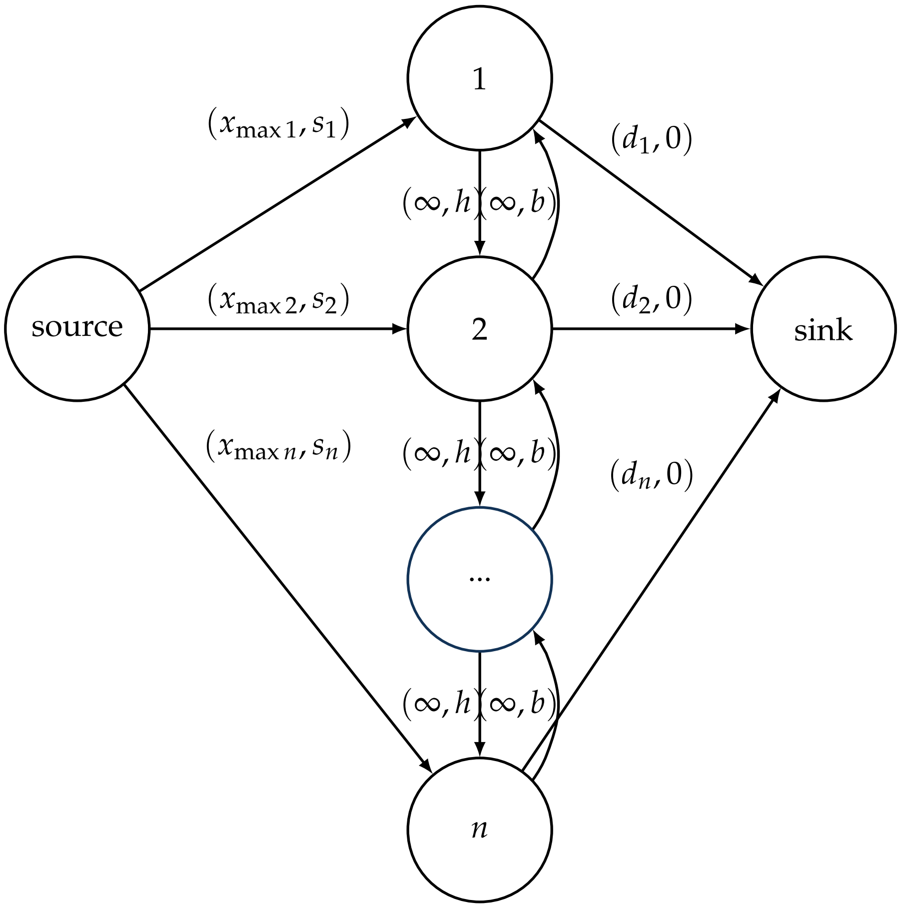

Figure 1: min cost max flot model. Arcs are labeled (capacity, cost)

2 Ibid.

2.1.3 *Feasable solution heuristic*

Since this problem has special structure, not found in original transportation problem, this heuristic enables us to construct better initial feasible solution. We can solve relaxed problem where backlogging is forbidden in $\mathcal{O}(n \log n)$ using greedy approach:

```
For each $d_i$ in order from 1 to $n$ do:
    While $d_i$ not satisfied:
        Find cheapest $x_k$ positive supply node and satisfy as much as possibl
```

By the end of this procedure we're going to satisfy all demand in optimal way in case it's possible to to do. Naive implementation results in $\mathcal{O}(n^2)$ runtime, however we can use priority queue, with heap backing implementation for $\mathcal{O}(n \log n)$ runtime.

This priority queue is sorted by costs and since all costs raise by constant amount between d_i and d_{i+1} we can keep it updated in $\mathcal{O}(1)$.

2.2 VARIANTS

Problem as defined previously could seem rather simplistic and not allowing useful extensions users might want, such as current storage amount and similar. In following few subsections most useful extensions are described.

2.2.1 *Starting storage capacity*

in case we have already certain number of product in stock we can easily embed that knowledge into model by adding new supplier/node as new decision variable x_0. It's maximum, $x_{\max 0}$ is equal to starting storage capacity, and $c_{0j} = (j-1)\,h$

2.2.2 *Ending storage requirement*

For example we'd like to have some extra product in stock by the end of analysis, and it's quite easy to accommodate such requirement. Simple add to d_n ending storage requirement, thus obtaining new d_n.

2.2.3 *Time Shifted ordering*

If we order raw materials at t they might arrive at later time moment $t + \Delta_t$. This can easily be modeled via variable substitution. For different Δ_t depending on the t or multiple suppliers see subsection 2.2.4

2.2.4 *Multiple raw material suppliers*

Adding new raw material suppliers with different costs cannot be done as per original model specification, however reduction to presented problems is straightforward.

2.2.5 *Allowing future backlogging*

In model as described, time stops at time moment n, however, in realistic scenario we're looking at only short time snapshot of ongoing process. To allow such future purchases to backlog at previous times, we can extend the model with m future moments:

$$\begin{aligned} d_{n+i} &= 0 && \\ x_{\max n+i} &= y_i && \text{allowable backlogging from future supply purchases} \\ &\mathbf{C} && \text{as previously defined, simply extended} \\ &\mathbf{s} && \text{extended with future costs} \end{aligned}$$

for i in $1, 2, \ldots, m$

3

INTRODUCING RANDOMNESS

TODO... fill this in

First we are going to analise problem deeply without presuming any independence or probability distribution on random variables. Later in subsequent chapters we are going to focus more on where demand at time t has independent Gaussian distribution and mean.

3.1 COST FUNCTION

The minimizing function is:

$$\mathrm{E}\left[\mathbf{s}^{\mathsf{T}}\mathbf{x} + \sum c(t)\right]$$

Due to linerality of expectation and x being variable it's equal to:

$$\mathbf{x}\,\mathrm{E}\left[\mathbf{s}^{\mathsf{T}}\right] + \mathrm{E}\left[\sum c(t)\right]$$

Therefore only needed modeling information for supply cost is its expectation $\mathrm{E}[\mathbf{s}]$. The other part is more trickier since D_i and D_j aren't usually independent.

3.2 HANDLING DEMAND COST NON-LINEARITY

As we can see in equation 1 we have non-linearity depending whether we're satifying all demand or are we backlogging demand at time t. Therefore here are two possible solutions for minimizing objective function 2.

Simulation

We generate multiple scenarios for demand vector, d according to probability distribution. For small n and relatively small number of outcomes in each random variable we can exhaustedly model each scenario, scale it appropriately and feed to MIP solver[1]

Safety net approach

Alternatively, we can artificially add new constraints and avoiding backlogging with arbitrary probably. This model assumes backlogging cost are significantly greater than storage cost, that is backlogging penalty is severe.

Thus we chose values arbitrary realizations of random variables D_t and add additional constraints of the form:

$$X_t \geq D_i \ \forall t \tag{4}$$

which reduces cost function 1 to:

$$c(t) = h\left(X_i - D_i\right) \ X_i \geq D_i$$

and enables are faster solving approaches. Given parameters D_i chosen and their underlying distribution we derive non-increasing function $p(t)$ which represent probability of not breaking newly introduced constraints 4

Min-cost max flow with uncertaintiy

Since deterministic case can be modeled with min-cost max flow, we can investigate further in this directi

Stochastic Dynamic Programmnig

TODO.. maybe, I've got recommendation by a friend, need to read some papers.

1 There's a trick on using binary variable for discontinuity in cost function 1

BIBLIOGRAPHY

Hillier, Frederick S. *Introduction to Operations Research*. McGraw-Hill, 2014. ISBN: 0073523453.