Contents

this file is dedicated to some thoughts on stochastic properties。

We want to show the same algorithm works for both dynamic and static models.

1 Stochastic

1.1 Dynamic / Multistage model

We now consider the optimization model under uncertainty. Suppose demand $\boldsymbol{d}$ is random with respect to some unknown distribution $f \in \mathcal{F}$. Similarly, we wish to solve stochastic model that minimizes expected summation of shortages and surpluses deviated from the demand unfold within a finite horizon.

Fruitful research has been done in the field of stochastic programming. Traditionally, stochastic programming approaches solve the expected objective that might be too optimistic. (see ...) Robust optimization, in contrast, optimizes a worst-case objective subject to the ambiguity set (see ...). Recently, the distributionally robust methods (see ...) provide a paradigm to minimize the worst-case risk ...

The multistage or dynamic models are known to be intractable, where at each stage decision is made after the realization of uncertain events. SAA... Furthermore, linear decision rules (LDR), see @, and Bertsimas et al. (2019) provides detailed analysis on LDR named after *adaptive distributionally robust optimization*.

We notice the stochastic version inherits the property that it could be decomposed into a set of independent subproblems by relaxing linking constraints. The idea regarding Lagrangian relaxation to dynamic optimization models is not new. Hawkins (2003) develops the theory of Lagrangian relaxation on so-called *weakly coupled Markov decision process* with applications to queueing networks, supply-chain management problems, and multiarmed bandits, et cetera. It provides analysis of both infinite and finite horizon versions of the problem. Adelman and Mersereau (2008) later contributes to the bound and optimality gap for both Lagrangian and linear programming relaxations.

We use boldface notation to denote random variables and corresponding decision variables.
Let Ξ_d be the support for random variable $\boldsymbol{d}$. Let $\boldsymbol{y} = [\boldsymbol{m}, \boldsymbol{s}]$ be the variable under uncertainty. Since $\boldsymbol{y}$ sufficiently represents the state at period t, we can write the multistage optimization model using dynamic programming equations.

Define V_t is the optimal value with t periods to go. Consider the following multistage stochastic optimization problem:

$$z = \min_{\boldsymbol{\delta}_t^-, \boldsymbol{\delta}_t^+, \boldsymbol{u}_{it}, \boldsymbol{x}_{it}} \mathbb{E}_f \left[\sum_t^{|T|} h \cdot \boldsymbol{\delta}_t^- + b \cdot \boldsymbol{\delta}_t^+ \right]$$

While the decisions are taken under conditions:

$$\sum_i \boldsymbol{u}_{it} - \boldsymbol{\delta}_t^- + \boldsymbol{\delta}_t^+ = \boldsymbol{d}_t, \boldsymbol{s}_{it}, \boldsymbol{u}_{it}, \boldsymbol{x}_{it} \in \Omega_i$$

Now limit the scope to the finite horizon. We are interested in the expected value with known initial state $\boldsymbol{y}_0$:

$$z_T(y_0) = \mathbb{E}_f()$$

Similar to the deterministic problem, the Bellman iteration can be written as:

$$V_t(\boldsymbol{y}, \boldsymbol{d}_t) = \min_{\boldsymbol{\delta}_t^-, \boldsymbol{\delta}_t^+, \boldsymbol{u}_{it}, \boldsymbol{x}_{it}} h \cdot \boldsymbol{\delta}_t^- + b \cdot \boldsymbol{\delta}_t^+ + \mathbb{E}_f \left[V_{t-1}(\boldsymbol{y}', \boldsymbol{d}_{t-1}') \middle| \boldsymbol{y}, \boldsymbol{d}_t \right]$$

We now investigate the Lagrangian relaxation. The analysis is similar to existing results in Adelman and Mersereau (2008), Hawkins (2003). The difference lies in the fact that we do not enforce λ_t to be identical across the stages $t = 1, ..., |T|$, which is necessary for infinite dimensional problems.

Proposition 1. Lagrangian relaxation provides a lower bound for any multiplier $\lambda = (\lambda_1, ..., \lambda_{|T|})$ such that $\lambda_t \in [-b, h], \ \forall t = 1, ..., |T|$.

$$V_t(\boldsymbol{y}, \boldsymbol{d}_t) \geq -\lambda_t \boldsymbol{d}_t + \sum_{i \in I} V_{it}(\boldsymbol{y}_i, \boldsymbol{d}_t)$$

Where V_{it} is the optimal equation for each i

$$V_{it}(\boldsymbol{y}, \boldsymbol{d}_t) = \boldsymbol{u}_{it} \lambda_t + \mathbb{E}_f \left[V_{i,t-1}(\boldsymbol{y}', \boldsymbol{d}_t') \middle| \boldsymbol{y}, \boldsymbol{d}_t \right]$$

PF. Relax binding constraints, since any feasible solution is the solution to the relaxed problem, we have:

$$V_t(\boldsymbol{y}, \boldsymbol{d}_t) \geq \min_{\boldsymbol{\delta}_t^-, \boldsymbol{\delta}_t^+, \boldsymbol{u}_{it}, \boldsymbol{x}_{it}} (h - \lambda_t) \cdot \boldsymbol{\delta}_t^- + (b + \lambda_t) \cdot \boldsymbol{\delta}_t^+ + \sum_i \boldsymbol{u}_{it}\lambda_t - \lambda_t \boldsymbol{d}_t + \mathbb{E}_f \left[V_{t-1}(\boldsymbol{y}', \boldsymbol{d}_t') \middle| \boldsymbol{y}, \boldsymbol{d}_t \right]$$

The RHS is unbounded unless $\lambda_t \in [-b, h]$, we have:

$$\begin{aligned} V_t(\boldsymbol{y}, \boldsymbol{d}_t) &\geq \min_{\boldsymbol{u}_{it}, \boldsymbol{x}_{it}} \sum_i \boldsymbol{u}_{it}\lambda_t - \lambda_t \boldsymbol{d}_t + \mathbb{E}_f \left[V_{t-1}(\boldsymbol{y}', \boldsymbol{d}_t') \middle| \boldsymbol{y}, \boldsymbol{d}_t \right] \\ &= -\lambda_t \boldsymbol{d}_t + \sum_{i \in I} V_{it}(\boldsymbol{y}_i, \boldsymbol{d}_t) \end{aligned}$$

The last line can be verified by induction similar to Hawkins (2003). This completes the proof. ■

Proposition 2. subgradient of λ.

1.2 Static Distributionally Robust Model

The DRO/SP model, the goal is to minimize worst-case expected unsatisfied demand and surplus (idle) flights

$$\begin{aligned} &\min \max_{f \in \mathcal{F}} \mathbb{E}_f \left[e^\top (b \cdot \boldsymbol{\delta}^+ + h \cdot \boldsymbol{\delta}^-) \right] \\ \textbf{s.t.}& \\ &\boldsymbol{U}^\top e + \boldsymbol{\delta}^+ - \boldsymbol{\delta}^- = \boldsymbol{d} \qquad && \forall \boldsymbol{d} \in \Xi_d \\ &\boldsymbol{U}_{(i,\cdot)}, \boldsymbol{X}_{(i,\cdot)}, \boldsymbol{S}_{(i,\cdot)} \in \Omega_i && \forall i \in I \end{aligned}$$

Same relaxation scheme can be used on the DRO models:

- Mean-variance, in Delage and Ye (2010).
- Likelihood, in Wang et al. (2016)

1.2.1 Moment Uncertainty

With moment uncertainty for $\boldsymbol{d} : \mathbb{E}(\boldsymbol{d}) = \mu_0, ...$, in Delage and Ye (2010). The DRO model is equivalent to the following problem:

$$
\begin{aligned}
\min_{\boldsymbol{U},\boldsymbol{Q},\boldsymbol{\beta},r,s} \quad & (\gamma_2 \boldsymbol{\Sigma}_0 - \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^\top) \bullet \boldsymbol{Q} + r + (\boldsymbol{\Sigma}_0 \bullet \boldsymbol{P}) - 2\boldsymbol{\mu}_0^\top \boldsymbol{p} + \gamma_1 s \\
\textbf{s.t.} \quad & \\
& \boldsymbol{Q} \succeq 0, \boldsymbol{\beta} \in \mathbb{R}^{|T|}, \begin{bmatrix} \boldsymbol{P} & \boldsymbol{p} \\ \boldsymbol{p}^\top & s \end{bmatrix} \succeq 0, \ \boldsymbol{\beta} = 2(\boldsymbol{p} + \boldsymbol{Q}\boldsymbol{\mu}_0) \\
& \boldsymbol{U}^\top e + \boldsymbol{\delta}^+ - \boldsymbol{\delta}^- = \boldsymbol{d} & \forall \boldsymbol{d} \in \Xi_d \\
& \boldsymbol{d}^\top \boldsymbol{Q} \boldsymbol{d} - \boldsymbol{d}^\top \boldsymbol{\beta} + r \geq e^\top (b \cdot \boldsymbol{\delta}^+ + h \cdot \boldsymbol{\delta}^-) & \forall \boldsymbol{d} \in \Xi_d \\
& \boldsymbol{X}_{(i,\cdot)}, \boldsymbol{U}_{(i,\cdot)}, \boldsymbol{S}_{(i,\cdot)} \in \Omega_i & \forall i \in I
\end{aligned}
$$

semi-infinite constraints are equivalent to (substitute $\boldsymbol{\delta}^- = \boldsymbol{u} + \boldsymbol{\delta}^+ - \boldsymbol{d}$, we have immediately)

$$
\begin{bmatrix} \boldsymbol{Q} & (he - \boldsymbol{\beta})/2 \\ (he - \boldsymbol{\beta})^\top/2 & r - (h+b)e^\top \boldsymbol{\delta}^+ - he^\top \boldsymbol{U}^\top e \end{bmatrix} \succeq 0
$$

wrap up:

$$
\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{Q},\boldsymbol{\beta},r,s} \quad & (\gamma_2 \boldsymbol{\Sigma}_0 - \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^\top) \bullet \boldsymbol{Q} + r + (\boldsymbol{\Sigma}_0 \bullet \boldsymbol{P}) - 2\boldsymbol{\mu}_0^\top \boldsymbol{p} + \gamma_1 s \\
\textbf{s.t.} \quad & \\
& \boldsymbol{Q} \succeq 0, \boldsymbol{\beta} \in \mathbb{R}^{|T|}, \begin{bmatrix} \boldsymbol{P} & \boldsymbol{p} \\ \boldsymbol{p}^\top & s \end{bmatrix} \succeq 0, \ \boldsymbol{\beta} = 2(\boldsymbol{p} + \boldsymbol{Q}\boldsymbol{\mu}_0) \\
& \begin{bmatrix} \boldsymbol{Q} & (he - \boldsymbol{\beta})/2 \\ (he - \boldsymbol{\beta})^\top/2 & r - (h+b)e^\top \boldsymbol{\delta}^+ - he^\top \boldsymbol{U}^\top e \end{bmatrix} \succeq 0 \\
& \boldsymbol{X}_{(i,\cdot)}, \boldsymbol{U}_{(i,\cdot)}, \boldsymbol{S}_{(i,\cdot)} \in \Omega_i & \forall i \in I
\end{aligned}
$$

is this too complex?

1.2.2 Finite support and likelihood robust

The problem is, let $\boldsymbol{Q} = [\boldsymbol{q}^1, ..., \boldsymbol{q}^N]$

$$\max_{\beta,\boldsymbol{\theta},\Omega_i,\forall i} \theta + \beta\gamma + \beta N - \underbrace{\beta\mathbf{N}^\top \log(\frac{\beta\mathbf{N}}{\boldsymbol{Q}e - \theta\mathbf{1}})}_{\mathcal{D}_{KL}(\beta\mathbf{N}|\boldsymbol{Q}e-\theta\mathbf{1})}$$

s.t.

$$\beta \geq 0$$

$$\boldsymbol{Q}e \geq \theta\mathbf{1}$$

$$\boldsymbol{U}^\top e + \boldsymbol{\delta}^+ - \boldsymbol{\delta}^- = \boldsymbol{d}^n \qquad \forall n = 1, ..., N$$

$$\boldsymbol{x}_{(i,\cdot)}, \boldsymbol{u}_{(i,\cdot)}, \boldsymbol{s}_{(i,\cdot)} \in \Omega_i \qquad \forall i \in I$$

Reference

Adelman D, Mersereau AJ (2008) Relaxations of weakly coupled stochastic dynamic programs. *Operations Research* 56(3):712–727.

Bertsimas D, Sim M, Zhang M (2019) Adaptive distributionally robust optimization. *Management Science*.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3).

Hawkins JT (2003) *A langrangian decomposition approach to weakly coupled dynamic optimization problems and its applications*. PhD thesis. (Massachusetts Institute of Technology).

Wang Z, Glynn PW, Ye Y (2016) Likelihood robust optimization for data-driven problems. *Computational Management Science* 13(2):241–261.