Optimality Conditions and Exact Algorithms for Risk-Averse Bilevel Stochastic Linear Problems

Bruno M. Pacheco

December 28, 2024

Introduction

Bilevel stochastic problems can be seen as a generalization of the two-stage problems we have seen in class. In both cases, there are two decisions to be made, before and after the realization of a random variable. The difference lies in that bilevel stochastic programming does *not* assume that both decisions are made by the same agent. In turn, this difference leads to a bilevel problem because the two stages do not share the same objective.

The properties of bilevel stochastic linear problems have been studied in the foundational works by Burtscheidt et al. (2020) and Claus (2021, 2022). The authors consider the more general risk-averse scenario, for which the risk-neutral case becomes a particular instance. They have presented proofs of the existence of optima and even optimality conditions for (classes of) bilevel problems in which the random variable appears in the right-hand side of the lower level (Burtscheidt et al., 2020), in the lower level cost function (Claus, 2021) in a quadratic manner, or in both (Claus, 2022). Although those are solid results, their interpretation and applicability is not easy to grasp, as they are proposed for abstract problem classes and assume intricate properties from the components of the mathematical programming models (constraint functions, solution space, objective function, etc.).

The overarching goal of this project is to deepen the understanding of the theoretical results for bilevel stochastic linear problems. The proposed approach is to explore the implications of these results for two classic textbook examples: the newsvendor problem and the multiproduct assembly problem. By proposing a bilevel variant of those problems and studying their theoretical properties following Burtscheidt et al. (2020), I expect to make those results tangible for risk-averse bilevel stochastic linear problems. Finally, I expect that those applications lead to a clear idea of which exact algorithms can be used to solve the proposed problems, reaching a practical conclusion.

The Newsvendor Problem

As presented in the preliminary report, the newsvendor problem can be formulated as

$$\begin{aligned} \min_{x} \quad & cx + Q(x, z) \\ \text{s.t.} \quad & 0 \le x \le u, \end{aligned} \tag{1}$$

in which

$$\begin{aligned} Q(x, z) = \min_{y,w} \quad & -qy - rw \\ \text{s.t.} \quad & y \le z \\ & y + w \le x \\ & y, w \ge 0. \end{aligned} \tag{2}$$

The decision variables x, y, and w represent, respectively, the amount of newspaper bought, the amount of newspaper sold, and the amount of newspaper returned. The problem is parameterized by the acquisition cost c, the storage capacity u, the demand z, the selling price q, and the return price r.

The traditional two-stage formulation comes from assuming that the demand is uncertain, and is modeled as a random variable $z = Z(\omega)$, where ω belongs to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, it is assumed that the realization of the random variable happens *after* the first decision is made (w.r.t. x), but *before* the second decision (w.r.t. y and w). Then, given a risk measure $\mathcal{R} : \mathcal{X} \longrightarrow \mathbb{R}$, where $\mathcal{X}$ is a linear subspace of all $\mathcal{F}$-measurable random variables, the two-stage problem becomes

$$\begin{aligned} \min_{x} \quad & \mathcal{R}[cx + Q(x, Z)] \\ \text{s.t.} \quad & 0 \le x \le u. \end{aligned} \tag{3}$$

Note that if we assume that $\mathcal{R}$ is translation invariant, then $\mathcal{R}[cx+Q(x,Z)] = cx + \mathcal{R}[Q(x,Z)]$, which is true for all three the risk measures of interest, nominally, the expected value, the value-at-risk, and the conditional value-at-risk.

The Bilevel Variant

In this work, I will assume a slight variation of the original newsvendor problem in which the lower-level decision is made by a different agent, with a different objective. This may represent, for example, a scenario in which the newspaper acquisition is made by a middle-man, which has different selling and return margins than the newspaper salesperson. Instead of (1), we have, then,

$$\begin{aligned} \min_x \quad & f(x,z) = cx + \min\{-q_u y - r_u w : (y,w) \in \Psi(x,z)\} \\ \text{s.t.} \quad & 0 \le x \le u, \end{aligned} \tag{4}$$

in which $\Psi(x,z)$ represents the set of solutions to the lower-level problem, that is,

$$\begin{aligned} \Psi(x,z) = \arg\min_{y,w} \quad & -q_l y - r_l w \\ \text{s.t.} \quad & y \le z \\ & y + w \le x \\ & y, w \ge 0. \end{aligned} \tag{5}$$

Note that the costs differ, that is, the selling and returning costs for the upper level are q_u and r_u, resp., while they are q_l and r_l for the lower level.

However, the analytic solution to the second stage (Birge & Louveaux, 2011) still holds, that is, as long as we assume that the selling price is strictly higher than the return price. More precisely, if $q_l > r_l$, the optimal solution to (5) can be written

$$y^*(x,z) = \min\{x,z\} \tag{6}$$

$$w^*(x,z) = \max\{x-z,0\}, \tag{7}$$

as long as the problem is feasible, i.e., $\forall x, z \ge 0$. In other words, $\Psi(x,z)$ is a singleton with $(\min\{x,z\}, \max\{x-z,0\})$. Thus, the upper-level cost can be written

$$f(x,z) = cx - q_u \min\{x,z\} - r_u \max\{x-z,0\}. \tag{8}$$

A slight difference arises if the second stage decision-maker does not have any incentive to sell the newspapers instead of returning them (e.g., no sales commision), in which case $q_l = r_l$ and the optimal solutions are

$$\begin{aligned}\Psi(x,z) &= \left\{(y,w) \in \mathbb{R}^2_+ : y + w = x, y \le z\right\} \\ &= \left\{(y,w) \in \mathbb{R}^2_+ : y \in [0,z]\,, w = \max\{x-y,0\}\right\}\end{aligned}.$$

Nonetheless, under the optimistic assumption, the solution that minimizes the upper-level cost[1] is also (6), as it is a feasible solution that achieves the optimal cost for the lower-level.

For completion, let us now consider the extreme case $q_l < r_l$, arising, for example, if the sales commission is not sufficiently higher than the cost of selling the newspapers, implying that simply returning the newspaper is the more profitable choice. Thus, the optimal solution is simply $y^* = 0$ and $w^* = x$, $\forall x,z \ge 0$, which is completely independent of the actual demand. The upper-level cost becomes $f(x,z) = (c - r_u)x$, which, of course, has $x = 0$ as an optimal solution[2].

At this point, it is natural to investigate the domain of $f(x,z)$ and its continuity properties. It is clear in all scenarios that $\mathrm{dom} f = \mathbb{R}^2_+$, as that renders both lower-level and the upper-level feasible. Note that the feasible space of the first-stage feasible set here lies within the domain of f (when projected into the x dimension). Therefore, any upper-level feasible assignment for x and z renders the lower-level feasible, a property similar to relative complete recourse. In fact, if we consider that $Z(\omega) \in \mathbb{R}_+ \forall \omega \in \Omega$, which is completely within the problem's context, we can even say that the problem has complete recourse.

Another interesting property to investigate is the differentiability of f. For $q_l \ge r_l$, Eq. (8) shows that the upper-level cost function is continuously differentiable almost everywhere with respect to the Lebesgue measure. This

[1] It is implicitly assumed that $q_U > r_U$. However, if that is not the case, it is easy to see that an even simpler analytical solution exists, for which the same results apply.

[2] That is, assuming $c < r_u$, a purchase cost smaller than the return cost, as the opposite would be nonsensical in the problem's context.

is due to f having continuous derivatives at the entirety of its domain, except for points in $\left\{(x,z)\in\mathbb{R}^2_+ : x=z\right\}$, which has null Lebesgue measure.

The Bilevel Stochastic Newsvendor

As for the two-stage problem, given a random demand Z and a risk measure $\mathcal{R}$, our bilevel newsvendor is interested in solving

$$\begin{aligned} \min_{x} \quad & \mathcal{R}[F(x)] \\ \text{s.t.} \quad & 0 \le x \le u, \end{aligned} \tag{9}$$

where $F(x) = f(x, Z)$ is a random variable parameterized by the upper-level decision x.

I recall that the only case worth investigating here is $q_l \ge r_l$, as otherwise the problem is demand independent. For the case of interest, the changes from the two-stage variant are negligible, but I nonetheless proceed for completion.

Before diving in the properties of the random variable $F(x)$ and the cost function $\mathcal{R}[F(x)]$, it is necessary to lay out some definitions about Z. Let μ_Z be the Borel probability measure induced by Z. This means that, in face of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the probability that μ_Z associates to a given set $\{z_1, z_2, \ldots\}$ of demand values is equal to the probability of the subset of Ω that contains the respective realization values. In other words, if $\{\omega_1, \omega_2, \ldots\} \in \Omega$ is such that $z_i = Z(\omega_i)$, for each $i = 1, 2, \ldots$, then $\mu_Z(\{z_1, z_2, \ldots\} = \mathbb{P}\{\omega_1, \omega_2, \ldots\}$. I will write this relationship as $\mu_Z = \mathbb{P} \circ Z^{-1}$, following Burtscheidt and Claus (2020). In the following, I will always assume that μ_Z has finite moments of order $p \in [1, \infty)$, which is equivalent to saying that $Z \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ or

$$\mathbb{E}[|Z|^p] = \int_{\mathbb{R}_+} |z|^p \mu_Z(dz) < \infty.$$

It is easy to see that $F(x)$ is Lipschitz continuous, that is, there exists $L > 0$ such that $\|F(x) - F(x')\| \le L|x - x'|$ for any $x, x' \ge 0$. In fact, $F(x)$ is also convex, as for any realization $Z(\omega)$,

$$f(x, Z(\omega)) = \begin{cases} (c-r)x - (q-r)Z(\omega) & , x \ge Z(\omega) \\ (c-q)x & , x < Z(\omega) \end{cases}$$

and $(c - q) \geq (c - r)$, i.e., the slope of $f(x, Z(\omega))$ increases with x for any $\omega \in \Omega$. Furthermore, given that μ_Z has finite moments of order $p \in [1, \infty)$, we can see that $\forall x \geq 0$, $\mu_{F(x)}$, the probability measure induced by $F(x)$, also has finite moments of order p, as $|f(x, z)| \leq (c - r)x$, which leads to

$$\begin{aligned}\mathbb{E}[|F(x)|^p] = \mathbb{E}[|f(x, Z)|^p] &= \int_{\mathbb{R}_+} |f(x, z)|^p \mu_Z(dz) \\ &\leq (c - r)^p x^p \int_{\mathbb{R}_+} \mu_Z(dz) < \infty,\end{aligned}$$

thus implying that $F(x) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$

Let us now assume $\mathcal{R}$ is convex and monotonic. Then, for any $x, x' > 0$, and any $t \in [0, 1]$, we use the convexity of $F(x)$ and the monotonicity of $\mathcal{R}$ to see that

$$\mathcal{R}[F(tx + (1 - t)x')] \leq \mathcal{R}[tF(x) + (1 - t)F(x')],$$

which, because $\mathcal{R}$ is also convex, shows that $\mathcal{R} \circ F$ is also convex. Now, by (Shapiro et al., 2009, Proposition 6.5), we know that if $\mathcal{R}$ is convex and monotonic, it will be continuous and subdifferentiable on $L^p(\Omega, \mathcal{F}, \mathbb{P})$. This, together with the previous result that $F(x)$ is Lipschitz continuous and that its domain ($\mathbb{R}_+$) is compact, leads us to applying Shapiro et al. (2009, Theorem 6.10) to see that $\mathcal{R} \circ F$ is directionally differentiable in its domain and the derivatives have finite value. Although the proof is a bit intricate, the rough idea is that this differentiability comes from both $\mathcal{R}$ and $F(x)$ being directionally differentiable at $F(\overline{x})$ and $\overline{x}$, resp., for any $\overline{x} > 0$.

Having access to directional derivatives gives us the possibility of stating optimality conditions, as a local optima implies, and is implied by, having null derivatives in all directions. Furthermore, having directional derivatives in the entirety of the domain enables the use of gradient-based methods. In fact, since we have that $\mathcal{R} \circ F$ is convex, gradient-methods should always converge to the global optimum.

Of course, the above results are unnecessary for the risk-neutral case $\mathcal{R}_{\mathbb{E}} = \mathbb{E}$, as one can much more easily derive analytical solutions (cf. Birge and Louveaux (2011, Chapter 1.e) and Shapiro et al. (2009, Chapter 1.2.1)). On the other hand, the results are perfectly suitable for the case of the Conditional Value-at-Risk, which follows in the general case, as shown in Proposition 1.

Proposition 1 (Uryasev (2000), Proposition 2)**.** The Conditional Value-at-Risk

$$\mathrm{CVaR}_\alpha[Y] = \min_{t\in\mathbb{R}} t + \frac{1}{1-\alpha}\mathbb{E}[\max\{Y-t,0\}]$$

is convex and monotonic for any $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Proof. Note that $y \mapsto \max\{y-t, 0\}$ is convex on y for any t. Take two random variables $Y_1, Y_2 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and let t_1, t_2 be the minimizers of $\mathrm{CVaR}_\alpha[Y_1]$ and $\mathrm{CVaR}_\alpha[Y_2]$, respectively. Then, for any $\lambda \in [0,1]$, we have that, for some t,

$$\begin{aligned}
&\mathrm{CVaR}_\alpha[\lambda Y_1 + (1-\lambda)Y_2] \\
&\quad = t + \frac{1}{1-\alpha}\mathbb{E}[\lambda Y_1 + (1-\lambda)Y_2 - t]_+ \\
&\quad \leq \lambda t_1 + (1-\lambda)t_2 + \frac{1}{1-\alpha}\mathbb{E}[\lambda Y_1 + (1-\lambda)Y_2 - \lambda t_1 - (1-\lambda)t_2]_+ \\
&\quad \leq \lambda t_1 + (1-\lambda)t_2 + \frac{1}{1-\alpha}\mathbb{E}[\lambda[Y_1 - t_1]_+ + (1-\lambda)[Y_2 - t_2]_+] \\
&\quad = \lambda\left(t_1 + \frac{1}{1-\alpha}\mathbb{E}[Y_1 - t_1]_+\right) + (1-\lambda)\left(t_2 + \frac{1}{1-\alpha}\mathbb{E}[Y_2 - t_2]_+\right) \\
&\quad = \lambda \mathrm{CVaR}_\alpha[Y_1] + (1-\lambda)\mathrm{CVaR}_\alpha[Y_2],
\end{aligned}$$

where $[y]_+ = \max\{y, 0\}$.

To show monotonicity, note that $y \mapsto \max\{y-t, 0\}$ is monotonic. Take any $Y_1, Y_2 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ with $Y_1 \leq Y_2$ almost surely, i.e., $Y_1(\omega) \leq Y_2(\omega)\, \forall\omega$ such that $\mathbb{P}(\omega) > 0$. Then, following the same notation as previously,

$$\begin{aligned}
\mathrm{CVaR}_\alpha[Y_1] &= t_1 + \frac{1}{1-\alpha}\mathbb{E}[Y_1 - t_1]_+ \\
&\leq t_2 + \frac{1}{1-\alpha}\mathbb{E}[Y_1 - t_2]_+ \\
&\leq t_2 + \frac{1}{1-\alpha}\mathbb{E}[Y_2 - t_2]_+ \\
&= \mathrm{CVaR}_\alpha[Y_2].
\end{aligned}$$

□

For the Value-at-Risk, one can use Ivanov (2014, Theorem 2) to show that it is also a continuous risk measure on the image of $F(x)$, and, thus, a solution to the problem exists. To the best of my knowledge, there is no result ensuring differentiability of the Value-at-Risk with respect to the problem variable x. However, we can see that

$$\begin{aligned}
\mathbb{P}[f(x,Z) \le \eta] &= \int_{\mathbb{R}_+} 1[f(x,z) \le \eta]\mu_Z(dz) \\
&= \int_0^x 1[f(x,z) \le \eta]\mu_Z(dz) + \int_x^\infty 1[f(x,z) \le \eta]\mu_Z(dz) \\
&= \int_0^x 1[(c-r)x - (q-r)z \le \eta]\mu_Z(dz) + \int_x^\infty 1[(c-q)x \le \eta]\mu_Z(dz) \\
&= \int_0^x 1[z \ge ((c-r)x - \eta)/(q-r)]\mu_Z(dz) + 1[(c-q)x \le \eta] \int_x^\infty \mu_Z(dz) \\
&= 1[(c-q)x \le \eta] \int_{\max\{((c-r)x-\eta)/(q-r),0\}}^x \mu_Z(dz) + 1[(c-q)x \le \eta] \int_x^\infty \mu_Z(dz) \\
&= 1[(c-q)x \le \eta] \int_{\max\{((c-r)x-\eta)/(q-r),0\}}^\infty \mu_Z(dz) \\
&= \begin{cases} 1 & , \eta \ge (c-r)x \\ \int_{((c-r)x-\eta)/(q-r)}^\infty \mu_Z(dz) & , (c-q)x \le \eta \le (c-r)x \\ 0 & , \eta \le (c-q)x \end{cases} .
\end{aligned}$$

Thus, for $\alpha \in (0,1]$, we have

$$\begin{aligned}
\mathrm{VaR}_\alpha[F(x)] &= \inf_\eta \{\eta \in \mathbb{R} : \mathbb{P}[F(x) \le \eta] \ge \alpha\} \\
&= \inf_\eta \{\eta \in \mathbb{R} : \mathbb{P}[f(x,Z) \le \eta] \ge \alpha\} \\
&= \inf_\eta \left\{\eta \in [(c-q)x, (c-r)x] : \int_{((c-r)x-\eta)/(q-r)}^\infty \mu_Z(dz) \ge \alpha \right\},
\end{aligned}$$

which, if we assume Z has a smooth cumulative distribution function (CDF), implies that $\mathrm{VaR}_\alpha \circ F$ is differentiable through a simple application of the implicit function theorem[3].

[3] This is due to $\mathrm{VaR}[F(x)]$ being formulated as a minimization (as the infimum is achieved) that is equivalent to solving $\int_{((c-r)x-\eta)/(q-r)}^\infty \mu_Z(dz) - \alpha = 0$.

The Multiproduct Assembly Problem

As in the preliminary report[4], the multiproduct assembly problem is formulated

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} + Q(\boldsymbol{x}, \boldsymbol{z}) \\ \text{s.t.} \quad & 0 \leq \boldsymbol{x} \leq \boldsymbol{u}, \end{aligned}$$

in which

$$\begin{aligned} Q(\boldsymbol{x}, \boldsymbol{z}) = \min_{\boldsymbol{y}, \boldsymbol{w}} \quad & -\boldsymbol{q}^T \boldsymbol{y} - \boldsymbol{r}^T \boldsymbol{w} \\ \text{s.t.} \quad & A^T \boldsymbol{y} + \boldsymbol{w} \leq \boldsymbol{x} \\ & \boldsymbol{y} \leq \boldsymbol{z} \\ & \boldsymbol{y}, \boldsymbol{w} \geq 0. \end{aligned}$$

In contrast to the newsvendor problem, here we have *vector* decision variables $\boldsymbol{x} = (x_1, \dots, x_m)$, $\boldsymbol{y} = (y_1, \dots, y_n)$, and $\boldsymbol{w} = (w_1, \dots, w_m)$. Each x_j and w_j represent the amount of each part j that was bought and that remained, resp., while each y_i represents the amount of product i that was assembled. The external parameters $\boldsymbol{z} = (z_1, \dots, z_n)$ represent the demand for each of the n products that can be assembled. Finally, at the first stage, $\boldsymbol{c}$ and $\boldsymbol{u}$ are the acquisition cost and the storage capacity for the parts, and at the second stage $\boldsymbol{q}$ and $\boldsymbol{r}$ are the sales price of the products and the recycling price of the parts, while in $A = [a_{ij}]$, a_{ij} is the amount of part j required to build product i.

As for the newsvendor problem, if we assume an uncertain demand modeled as the realization of a random variable $\boldsymbol{z} = Z(\omega)$, where ω belongs to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the two-stage problem is formulated very similarly to (3)

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \mathcal{R}[\boldsymbol{c}^T \boldsymbol{x} + Q(\boldsymbol{x}, Z)] \\ \text{s.t.} \quad & 0 \leq \boldsymbol{x} \leq \boldsymbol{u}. \end{aligned} \tag{10}$$

However, I highlight that, despite using the same notation, for the multiproduct assembly problem $Z(\omega) \in \mathbb{R}^n_+$, i.e., it can be seen as a vector-valued function.

[4] In the preliminary report, I used $\boldsymbol{d}$ to refer to the problem parameter, which I here decided to use $\boldsymbol{z}$ for consistency with respect to the newsvendor problem.

A Bilevel Variant

Once again, the consideration that the second-stage decision is made by a different agent is used to formulate a bilevel problem. In here, this may represent, for example, the dynamics between a supplier (first stage) and a manufacturer (second stage). The bilevel variant is written

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{c}^T \boldsymbol{x} + \min \left\{ -\boldsymbol{q}_u^T \boldsymbol{y} - \boldsymbol{r}_u^T \boldsymbol{w} : (\boldsymbol{y}, \boldsymbol{w}) \in \Psi(\boldsymbol{x}, \boldsymbol{z}) \right\} \\ \text{s.t.} \quad & 0 \leq \boldsymbol{x} \leq \boldsymbol{u}, \end{aligned} \tag{11}$$

in which

$$\begin{aligned} \Psi(\boldsymbol{x}, \boldsymbol{z}) = \arg\min_{\boldsymbol{y}, \boldsymbol{w}} \quad & -\boldsymbol{q}_l \boldsymbol{y} - \boldsymbol{r}_l \boldsymbol{w} \\ \text{s.t.} \quad & A^T \boldsymbol{y} + \boldsymbol{w} \leq \boldsymbol{x} \\ & \boldsymbol{y} \leq \boldsymbol{z} \\ & \boldsymbol{y}, \boldsymbol{w} \geq 0. \end{aligned} \tag{12}$$

Because there is no analytic solution for the multiproduct assembly problem (such as the one for the newsvendor problem), the first interesting property to analyse on the bilevel variant, even before introducing the random variable, is that of the continuity of function f. With Lemma 1, we guarantee the problem is well defined and, since the feasible solution space of the upper level is compact, the problem is solvable.

Lemma 1 (Burtscheidt and Claus (2020), Lemma 17.2.1)**.** Function f is real-valued and Lipschitz continuous on $\mathbb{R}_+^{m+n}$.

Proof. It is easy to see that $\forall\, (\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}_+^{m+n}$ the lower-level problem is feasible, e.g., through the naïve assignment $(\overline{\boldsymbol{y}}, \overline{\boldsymbol{w}}) = (0, \boldsymbol{x})$. Also, the lower-level is also bounded, as $\boldsymbol{x}$ and $\boldsymbol{z}$ are, resp., upper-bounds for $\boldsymbol{w}$ and $\boldsymbol{y}$. As a consequence, the minimization problem in $f(\boldsymbol{x}, \boldsymbol{z})$ exists (i.e., is well-defined), which renders $f(\boldsymbol{x}, \boldsymbol{z})$ a real-valued function, or, equivalently, $\operatorname{dom} f = \mathbb{R}_+^{m+n}$.

Now, to demonstrate Lipschitz continuity, take any $(\boldsymbol{x}, \boldsymbol{z}), (\boldsymbol{x}', \boldsymbol{z}') \in \mathbb{R}_+^{m+n}$ and assume, without loss of generalization, that $f(\boldsymbol{x}, \boldsymbol{z}) \geq f(\boldsymbol{x}', \boldsymbol{z}')$. Then, take $(\boldsymbol{y}', \boldsymbol{w}') \in \Psi(\boldsymbol{x}', \boldsymbol{z}')$ such that $f(\boldsymbol{x}', \boldsymbol{z}') = \boldsymbol{c}^T \boldsymbol{x}' - \boldsymbol{q}_u \boldsymbol{y}' - \boldsymbol{r}_u \boldsymbol{w}'$. Parametric programming theory (Klatte & Thiere, 1995, Lemma 4.1) gives us that for every point $(\boldsymbol{y}', \boldsymbol{w}') \in \Psi(\boldsymbol{x}', \boldsymbol{z}')$, there exists $\Lambda > 0$ such that for any other point $(\boldsymbol{y}, \boldsymbol{w}) \in \Psi(\boldsymbol{x}, \boldsymbol{z})$

$$(\boldsymbol{y}', \boldsymbol{w}') = (\boldsymbol{y}, \boldsymbol{w}) + \Lambda \|(\boldsymbol{x}, \boldsymbol{z}) - (\boldsymbol{x}', \boldsymbol{z}')\| \boldsymbol{e}$$

for some a vector $\boldsymbol{e} \in \mathbb{R}^{m+n}$ with $\|\boldsymbol{e}\| \leq 1$. Thus, for any $(\boldsymbol{y}, \boldsymbol{w}) \in \Psi(\boldsymbol{x}, \boldsymbol{z})$,

$$\begin{aligned}
|f(\boldsymbol{x}, \boldsymbol{z}) - f(\boldsymbol{x}', \boldsymbol{z}')| &= f(\boldsymbol{x}, \boldsymbol{z}) - \boldsymbol{c}^T \boldsymbol{x}' + \boldsymbol{q}_u^T \boldsymbol{y}' + \boldsymbol{r}_u^T \boldsymbol{w}' \\
&\leq \boldsymbol{c}^T \boldsymbol{x} - \boldsymbol{q}_u^T \boldsymbol{y} - \boldsymbol{r}_u^T \boldsymbol{w} - \boldsymbol{c}^T \boldsymbol{x}' + \boldsymbol{q}_u^T \boldsymbol{y}' + \boldsymbol{r}_u^T \boldsymbol{w}' \\
&\leq \|\boldsymbol{c}\| \|\boldsymbol{x} - \boldsymbol{x}'\| + \|(\boldsymbol{q}_u, \boldsymbol{r}_u)\| \|(\boldsymbol{y}, \boldsymbol{w}) - (\boldsymbol{y}', \boldsymbol{w}')\| \\
&\leq \|\boldsymbol{c}\| \|\boldsymbol{x} - \boldsymbol{x}'\| + \|(\boldsymbol{q}_u, \boldsymbol{r}_u)\| \Lambda \|(\boldsymbol{x}, \boldsymbol{z}) - (\boldsymbol{x}', \boldsymbol{z}')\| \|\boldsymbol{e}\| \\
&\leq L_f \|(\boldsymbol{x}, \boldsymbol{z}) - (\boldsymbol{x}', \boldsymbol{z}')\|,
\end{aligned}$$

where $L_f = \|\boldsymbol{c}\| + \Lambda \|(\boldsymbol{q}_u, \boldsymbol{r}_u)\|$. □

The Bilevel Stochastic Multiproduct Assembly

Knowing that the deterministic problem is solvable lets us wander into the stochastic version of the problem. Recall that the parameter $\boldsymbol{z}$ (demand) is the realization of the random variable Z, which happens prior to the lower-level decision, but after the upper-level decision. Furthermore, I consider here a risk-averse problem, such that, given a risk measure $\mathcal{R}$ as previously, the problem of interest is

$$\begin{aligned}
\min_{\boldsymbol{x}} \quad & \mathcal{R}[F(\boldsymbol{x})] \\
\text{s.t.} \quad & 0 \leq \boldsymbol{x} \leq \boldsymbol{u},
\end{aligned} \tag{13}$$

where $F(\boldsymbol{x}) = f(\boldsymbol{x}, Z)$ is a random variable parameterized by the upper-level decision $\boldsymbol{x}$. Note that $F(\boldsymbol{x})$ is well-defined as any upper-level decision $\boldsymbol{x} \in \mathbb{R}_+^m$ renders the lower-level feasible (complete recourse).

As for the newsvendor problem, let μ_Z be the Borel probability measure induced by Z. I will assume that μ_Z has finite moments of order $p \in [1, \infty)$, i.e., $Z \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. This allows me to state the Lipschitz continuity of the random variable $F(\boldsymbol{x})$.

Lemma 2 (Burtscheidt and Claus (2020), Lemma 17.2.4[a]). If μ_Z has finite moments of order $p \in [1, \infty)$, then $\exists L > 0$ such that

$$\|F(\boldsymbol{x}) - F(\boldsymbol{x}')\|_p \leq L\|\boldsymbol{x} - \boldsymbol{x}'\|, \quad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^m_+,$$

i.e., $F(\boldsymbol{x})$ is Lipschitz continuous with respect to the L^p-norm $\|F(\boldsymbol{x})\|_p = \mathbb{E}[|F(\boldsymbol{x})|^p]^{1/p}$.

[a] Except the case for probability measures with finite moments of order $p = \infty$.

Proof. First, note that, given Lemma 1, $\forall \boldsymbol{x} \in \mathbb{R}^m_+$

$$\begin{aligned}
(\|F(\boldsymbol{x})\|_p)^p &= \int_{\mathbb{R}^n_+} |f(\boldsymbol{x}, \boldsymbol{z})|^p \mu_Z(d\boldsymbol{z}) \\
&= \int_{\mathbb{R}^n_+} |f(\boldsymbol{x}, \boldsymbol{z}) - f(0,0) + f(0,0)|^p \mu_Z(d\boldsymbol{z}) \\
&\leq \int_{\mathbb{R}^n_+} |f(0,0)|^p + |f(\boldsymbol{x}, \boldsymbol{z}) - f(0,0)|^p \mu_Z(d\boldsymbol{z}) \\
&\leq |f(0,0)|^p + \int_{\mathbb{R}^n_+} L_f^p \|(\boldsymbol{x}, \boldsymbol{z})\|^p \mu_Z(d\boldsymbol{z}) \\
&\leq |f(0,0)|^p + L_f^p \int_{\mathbb{R}^n_+} \|\boldsymbol{x}\|^p + \|\boldsymbol{z}\|^p \mu_Z(d\boldsymbol{z}) \\
&= |f(0,0)|^p + L_f^p \|\boldsymbol{x}\|^p + L_f^p \int_{\mathbb{R}^n_+} \|\boldsymbol{z}\|^p \mu_Z(d\boldsymbol{z}) < \infty,
\end{aligned}$$

as μ_Z has finite moments of order p. Then, for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^m_+$,

$$\begin{aligned}
\|F(\boldsymbol{x}) - F(\boldsymbol{x}')\|_p &= \left(\int_{\mathbb{R}^n_+} |f(\boldsymbol{x}, \boldsymbol{z}) - f(\boldsymbol{x}', \boldsymbol{z})|^p \mu_Z(d\boldsymbol{z}) \right)^{1/p} \\
&\leq \left(\int_{\mathbb{R}^n_+} (L_f \|(\boldsymbol{x}, \boldsymbol{z}) - (\boldsymbol{x}', \boldsymbol{z})\|)^p \mu_Z(d\boldsymbol{z}) \right)^{1/p} = L_f \|\boldsymbol{x} - \boldsymbol{x}'\|.
\end{aligned}$$

$\square$

Continuity and Existence of Solutions

The discussion presented for the newsvendor problem is suitable for the more general multiproduct assembly problem as well. For us to show that the problem is solvable we need just to show that the risk measures are convex and monotonic, as Lemma 2 guarantees us the remaining. Nevertheless, I will provide in the following a more in-depth discussion for the three risk measures at hand.

Lemma 3 (Burtscheidt et al. (2019), Theorem 3.4)**.** If μ_Z has finite moments of order $p = 1$, then $\mathcal{R}_{\mathbb{E}} = \mathbb{E}$ is real-valued and Lipschitz continuous.

Proof. This proof is very similar to the proof of Lemma 2. First, for any $\boldsymbol{x} \in \mathbb{R}^n_+$,

$$\begin{aligned} |\mathcal{R}_{\mathbb{E}}[F(\boldsymbol{x})]| = |\mathbb{E}[f(\boldsymbol{x}, Z)]| &\leq \int_{\mathbb{R}^n_+} |f(\boldsymbol{x}, \boldsymbol{z})| \mu_Z(d\boldsymbol{z}) \\ &\leq |f(0,0)| + L_f \|\boldsymbol{x}\| + L_f \int_{\mathbb{R}^n_+} \|\boldsymbol{z}\| \mu_Z(d\boldsymbol{z}) < \infty. \end{aligned}$$

Now, for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^m_+$,

$$\begin{aligned} |\mathcal{R}_{\mathbb{E}}[F(\boldsymbol{x})] - \mathcal{R}_{\mathbb{E}}[F(\boldsymbol{x}')]| &= \left| \int_{\mathbb{R}^n_+} \left(f(\boldsymbol{x}, \boldsymbol{z}) - f(\boldsymbol{x}', \boldsymbol{z})\right) \mu_Z(d\boldsymbol{z}) \right| \\ &\leq \int_{\mathbb{R}^n_+} |f(\boldsymbol{x}, \boldsymbol{z}) - f(\boldsymbol{x}', \boldsymbol{z})| \mu_Z(d\boldsymbol{z}) \\ &= L_f \|\boldsymbol{x} - \boldsymbol{x}'\|. \end{aligned}$$

□

The case for the conditional value-at-risk is more intricate. Generally, the results from Proposition 1 apply to the multiproduct assembly problem, guaranteeing that directional derivatives exist. If Z has *finite support*, I present here an alternative demonstration of the results through the reformulation of the CVaR presented in Birge and Louveaux (2011, Chapter 2.9)

$$\begin{aligned}
\mathrm{CVaR}_\alpha[F(\boldsymbol{x})] &= \min_{t\in\mathbb{R}} t + \frac{1}{1-\alpha}\mathbb{E}[\max\{F(\boldsymbol{x}) - t, 0\}] \\
&= \min_{t} t + \frac{1}{1-\alpha}\int_{\mathbb{R}^n_+} \max\{f(\boldsymbol{x},\boldsymbol{z}) - t, 0\}\mu_Z(d\boldsymbol{z}) \\
&= \min_{t,g(\boldsymbol{z})} t + \frac{1}{1-\alpha}\int_{\mathbb{R}^n_+} g(\boldsymbol{z})\mu_Z(d\boldsymbol{z}) \\
&\qquad \text{s.t.} \quad g(Z) + t \geq f(\boldsymbol{x}, Z), \text{ a.s.} \\
&\qquad\qquad\ \ g(Z) \geq 0, \text{ a.s.},
\end{aligned}$$

that is, the properties of $\mathcal{R}_{\mathrm{CVaR}_\alpha}$ can be shown from the study of the above parametric optimization problem. Because the problem above is clearly not degenerate[5], (Pistikopoulos et al., 2021, Theorem 2.1) shows that the objective is continuous, convex, and piecewise affine with respect to $f(\boldsymbol{x}, \overline{\boldsymbol{z}})$, $\forall \overline{\boldsymbol{z}} \in \mathrm{supp}(\mu_Z)$. As a direct consequence, Lipschitz continuity of f on $\boldsymbol{x}$ implies that $\mathcal{R}_{\mathrm{CVaR}_\alpha} \circ F$ is also Lipschitz continuous. Intuitively, as one smoothly varies $\boldsymbol{x}$, $f(\boldsymbol{x}, Z)$ also changes smoothly, which smoothly contracts or expands the polyhedron that defines the set of feasible solutions to the problem. This, in turn, leads to a smooth variation of the optimal objective value, once the problem is not degenerate. If Z has infinite support, the problem becomes an optimization problem over a functional space (i.e., $L^1(\Omega, \mathcal{F}, \mathbb{P})$). The intuition still holds, but, to the best of my knowledge, no one has shown that the above class of parametric problems is continuous with respect to the parameter, although the result seems achievable.

The value-at-risk, which is *not* convex, needs a special treatment. I recall once again that, by Ivanov (2014, Theorem 2), VaR_α is a continuous risk measure on the image of our random variable. On top of that, since $f(\boldsymbol{x}, \boldsymbol{z})$ is real-valued, then $F(\boldsymbol{x})$ is also real-valued[6], and, thus, by the definition of the value-at-risk, $\mathrm{VaR}_\alpha[F(\boldsymbol{x})]$ is also real-valued.

Therefore, for the three risk measures of interest we can guarantee that the problem is solved, as the feasible space is compact and the objective function is continuous and real-value.

[5]That is, for every possible $\boldsymbol{x}$ assignment, the optimal solution is unique and the vertex that yields it is not overdefined, i.e., the set of active constraints is also unique.

[6]One could say that $F(\boldsymbol{x})$ is real-valued *almost surely*, but I keep the understanding that the counter-domain of Z is $\mathbb{R}^m_+$.

Exact Algorithms

Given the properties seen thus far, the existence of exact algorithms for our problem can be boiled down to the study of two fundamental properties: differentiability and convexity. Differentiability of our cost function enables the use of gradient-based line search methods such as steepest ascent, which even in the absence of convexity still guarantee to converge towards a local minima. On the other hand, convexity enables the use of algorithms that converge towards global optima, such as subgradient decent or cutting-planes, which do not require differentiability.

Differently from continuity, differentiability cannot be guaranteed simply through convexity and monotonicity of the risk measures. Burtscheidt et al. (2020, Proposition 17.2.13) have provided sufficient conditions for the risk-neutral case to be differentiable, but those are *not* satisfied by the multiproduct assembly problem. This is because they require that the random variable continuously affects *every* lower-level constraint, and, thus, has a probability measure that is absolutely continuous with respect to the Lebesgue measure. As the authors highlight in Remark 17.2.15, if there is a single constraint in which the random variable has no influence, as is the case of $A^T\boldsymbol{y} + \boldsymbol{w} \leq \boldsymbol{x}$, then the extended random variable $Z' = (Z_1, Z_2)$ such that $Z_2 = Z$ and the constraint becomes $A^T\boldsymbol{y}+\boldsymbol{w} \leq \boldsymbol{x}+Z_1$, does not have a absolutely continuous probability measure $\mu_{Z'}$, as $\mathbb{P}[Z_1 = 0] = 1$. For the two other risk measures, general results are unknown.

However, convexity can be shown for special cases. Let us assume that the lower-level problem is not degenerate, that is, for every $\boldsymbol{x} \in \mathbb{R}^m_+$ and $\boldsymbol{z} \in \mathbb{R}^n_+$, the optimal solution is unique and it is uniquely determined by a combination of active constraints (the vertex is uniquely determined). This allows us to resort once again to parametric optimization theory, as (Pistikopoulos et al., 2021, Theorem 2.1) guarantees that $\Psi(\boldsymbol{x}, \boldsymbol{z})$ is a continuous and piecewise affine function. Thus, the minimization in $f(\boldsymbol{x}, \boldsymbol{z})$ is convex with respect to its parameters, which implies that f is also convex. As a consequence, $F(x)$ is convex and, thus, given a convex and monotonic risk measure, $\mathcal{R} \circ F$ is convex.

Conclusions

- For differentiability of the multiproduct assembly problem, maybe the results from Burtscheidt et al. 2019 could be applied, as they have looser assumptions than absolute continuity, but that would require a deeper investigation.

References

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming* (2nd ed). Springer. Retrieved December 5, 2024, from https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2546867
OCLC: 745001039.

Burtscheidt, J., Claus, M., & Dempe, S. (2019, January 31). *Risk-Averse Models in Bilevel Stochastic Linear Programming*. arXiv: 1901.11349 `[math]`. https://doi.org/10.48550/arXiv.1901.11349

Burtscheidt, J., & Claus, M. (2020). Bilevel Linear Optimization Under Uncertainty. In S. Dempe & A. Zemkoho (Eds.), *Bilevel Optimization: Advances and Next Challenges* (pp. 485–511). Springer International Publishing. https://doi.org/10.1007/978-3-030-52119-6_17

Burtscheidt, J., Claus, M., & Dempe, S. (2020). Risk-Averse Models in Bilevel Stochastic Linear Programming. *SIAM Journal on Optimization*, *30*(1), 377–406. https://doi.org/10.1137/19M1242240

Claus, M. (2021). On continuity in risk-averse bilevel stochastic linear programming with random lower level objective function. *Operations Research Letters*, *49*(3), 412–417. https://doi.org/10.1016/j.orl.2021.04.007

Claus, M. (2022). Existence of solutions for a class of bilevel stochastic linear programs. *European Journal of Operational Research*, *299*(2), 542–549. https://doi.org/10.1016/j.ejor.2021.12.004

Ivanov, S. V. (2014). Bilevel stochastic linear programming problems with quantile criterion. *Automation and Remote Control*, *75*(1), 107–118. https://doi.org/10.1134/S0005117914010081

Klatte, D., & Thiere, G. (1995). Error bounds for solutions of linear equations and inequalities. *Zeitschrift für Operations Research*, *41*(2), 191–214. https://doi.org/10.1007/BF01432655

Pistikopoulos, E. N., Diangelakis, N. A., & Oberdieck, R. (2021). *Multi-parametric optimization and control*. Wiley. https://doi.org/10.1002/9781119265245

Shapiro, A., Dentcheva, D., & Ruszczyński, A. P. (2009). *Lectures on stochastic programming: Modeling and theory*. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104). Retrieved December 5, 2024, from http://epubs.siam.org/ebooks/siam/mos-siam_series_on_optimization/mp09
OCLC: 693953130.

Uryasev, S. (Ed.). (2000). *Probabilistic constrained optimization: Methodology and applications*. Kluwer.