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

Bruno M. Pacheco

December 16, 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: one before and another 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 (e.g., constraint functions, solution space, objective function).

The overarching goal of this project is to deeper 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

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

in which

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

The decision variables x, y, and w represent, respectively, the amount of newspaper initially bought, the amount of newspaper sold, and the amount of newspaper returned w. The problem is parameterized by the acquisition cost c, the newspaper 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 comes from 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 (w.r.t. x), and 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

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

Note that if we assume that $\mathcal{R}$ is translation invariant (which is a common property to expect, e.g., variance, value-at-risk), then $\mathcal{R}[cx + Q(x, Z)] = cx + \mathcal{R}[Q(x, Z)]$, which is true, for example, in the risk-neutral case $\mathcal{R} = \mathbb{E}$.

A 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,

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

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

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

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.

There is, however, an analytic solution to the second stage (Birge & Louveaux, 2011) that assumes the selling price is strictly higher than the return price. More precisely, assuming $q_l > r_l$, 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., $\forall x, z \geq 0$. 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

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

The difference arises, for example, 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

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

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

Let us consider now the extreme case $q_l < r_l$, for example, the sales commission is inferior to the cost of selling, implying that simply returning the newspaper is the more profitable choice. Thus, the optimal solution is simply $y^* = 0$ and $w^* = x$, $\forall x, z \geq 0$, which is completely independent of the actual demand. The upper-level cost becomes $f(x, z) = (c - r_u)x$, which, of course, has an optimal solution at $x = 0$[1].

At this point, it would be natural to investigate the domain of $f(x, z)$ and its continuity properties. It is clear in all scenarios that $\operatorname{dom} f = \mathbb{R}^2_+$, as that renders both lower-level and the upper-level feasible, i.e., the problem has a property similar to complete recourse. For $q_l \geq r_l$, Eq. (8) shows that the upper-level cost function is continuously differentiable almost everywhere with respect to the Lebesgue measure, that is, f has continuous derivatives except at $\left\{(x, z) \in \mathbb{R}^2_+ : x = z\right\}$, 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

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

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

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 $q_l \geq r_l$, as otherwise the problem is demand independent, and for that case, the changes from the two-stage variant problem are negligible. Nevertheless, before we dive 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. In fact, $F(x)$ is also convex, as for any realization $Z(\omega)$,

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

and $(c-q) \geq (c-r)$. Furthermore, we can see that $\forall x \geq 0$, $F(x)$ has finite moments of order $p \in [1, \infty)$, as $|f(x,z)| \leq (c-r)x$, which leads to

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

Because $F(x) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, any real-valued $\mathcal{R}$ that is both convex and monotonic will be continuous and subdifferentiable on the image of $F(x)$ (Shapiro et al., 2009, Proposition 6.5). This and the convexity of $F(x)$ allow, in turn, to state that $\mathcal{R} \circ F$ is directionally differentiable at the problem domain Shapiro et al. (2009, Theorem 6.10). Furthermore, because $F(x)$ is also monotonic, then the composition will be convex whenever $\mathcal{R}$ is convex. Therefore, gradient-based methods are exact algorithms for this problem and will (likely) return globally optimal solutions.

Of course, the above results are unnecessary for the risk-neutral case, 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 is, in fact, a coherent risk measure. 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

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

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

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

which, if we assume Z has a smooth CDF, implies that $\mathrm{VaR}_\alpha \circ F$ is differentiable through a simple application of the implicit function theorem[2].

The Multiproduct Assembly Problem

Previous stuff

ADAPT FROM NEWSVENDOR: The first interesting property to analyse, even before introducing the random variable, is that of the function f.

Lemma 1 (Burtscheidt and Claus (2020), Lemma 17.2.1)**.** Function f is real-valued and Lipschitz continuous $\forall x, z \geq 0$.

Proof. It is easy to see that $\forall x, z \geq 0$, the lower-level problem is feasible, and, thus, a solution to the minimization problem in $f(x, z)$ exists, which renders $f(x, z)$ a real-valued function.

Now, to demonstrate Lipschitz continuity, take any $x, z, x', z' \geq 0$ such that $f(x, z) \geq f(x', z')$. Then, take $(y', w') \in \Psi(x', z')$, which means that $f(x', z') = cx' - q_u y' - r_u w'$, and, thus, for any $(y, w) \in \Psi(x, z)$. On top of that, by Klatte and Thiere (1995, Theorem 4.2), we have that every point $(y', w') \in \Psi(x', z')$ can be expressed as

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

for some $(y, w) \in \Psi(x, z)$, a vector $e \in \mathbb{R}^2$ with $\|e\| \leq 1$, and some constant $\Lambda > 0$. Thus, assuming that $c \geq 0$ (which is indeed expected, as it represents

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

a cost),

$$\begin{aligned}
|f(x,z) - f(x',z')| &= f(x,z) - cx' + q_u y' + r_u w' \\
&\leq cx - q_u y - r_u w - cx' + q_u y' + r_u w' \\
&\leq c|x - x'| + \|(q_u, r_u)\| \|(y,w) - (y',w')\| \\
&\leq c|x - x'| + \|(q_u, r_u)\| \Lambda \|(x,z) - (x',z')\| \|e\| \\
&\leq L_f \|(x,z) - (x',z')\|,
\end{aligned}$$

where $L_f = c + \Lambda \|(q_u, r_u)\|$. □

Furthermore, the random variable $F(x)$ is convex, as for any realization $Z(\omega)$,

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

and $(c-q) \geq (c-r)$. Thus, if $\mathcal{R}$ is convex and monotonic, then $\mathcal{R}[F(x)]$ is convex (Shapiro et al., 2009, Proposition 6.8).

More precisely, the problem is equivalent to the two-stage problem with $q = q_u$ and $r = r_u$, so the known properties hold.

This allows me to reformulate the result that presents the Lipschitz continuity of $F(x)$ for the bilevel newsvendor problem.

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

$$\|F(x) - F(x')\|_p \leq L|x - x'|, \quad \forall x, x' \geq 0,$$

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

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

Proof. First, note that, given Lemma 1, $\forall x \geq 0$

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

as μ_Z has finite moments of order p. Then, $\forall x, x' \geq 0$,

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

$\square$

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. (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

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.