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

Bruno M. Pacheco

December 7, 2024

## Introduction

In this course, we have studied two-stage stochastic programming problems with recourse, in which an agent has to make two decisions, before and after the realization of a random variable. Because of the nature of the problem, the decisions are made towards the same objective. However, if we assume that the two decisions are *not* made by the same agent, and that there are conflicting objectives, the problem formulation changes. In fact, as formulated by Burtscheidt and Claus (2020), this difference in the assumption gives rise to a bilevel stochastic programming problem. In other words, the result is a bilevel problem in which the upper-level makes a here-and-now decision (prior to the realization of the random variable), while the lower-level makes a wait-and-see decisions (after the realization of the random variable).

The overarching goal of this project is to study bilevel stochastic linear problems. The proposed approach is to perform a thought experiment on two textbook examples: the newsvendor problem (Birge and Louveaux, 2011; Shapiro, Dentcheva, and A. P. Ruszczyński, 2009) and the multiproduct assembly problem (Shapiro, Dentcheva, and A. P. Ruszczyński, 2009). For both problems, the starting point will be to assume that the two decisions (first-stage and second-stage) are *not* made by the same agent, and, thus, the two stages have distinct objective functions. This assumption changes the problem formulation from a two-stage recourse problem, to a bilevel stochastic linear problem. Furthermore, I will consider a risk-averse scenario, for which the traditional assumption that the upper-level agent is optimizing the expected value is a particular case (as the expectation is a coherent risk

measure).

I aim to study optimality conditions of bilevel stochastic linear problems and reformulations that lead to exact algorithms, which will be greatly supported by the theoretical results from Burtscheidt, Claus, and Dempe (2020). The authors have provided general results for this class of problems, which I plan to present in greater detail while considering the specificities of the two selected problems. Note that the choice of the problems is well-thought, as the theoretical results of Burtscheidt, Claus, and Dempe (2020) only apply to problems in which the random variable appears in the right-hand side of the lower-level constraints[1].

## The News Vendor Problem

The newsvendor problem is the problem of deciding how many products (in here, newspapers) to buy before the demand is known. Following most of the notation from Birge and Louveaux (2011), the vendor decides to buy $x$ amount of newspaper. After this acquisition, the product can be sold up to a certain amount, the demand $z$. The leftover newspaper, if any, is returned at a lower price. Formally, the problem is formulated as a parametric problem

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

in which $c$ is the acquisition cost, and $u$ is the capacity of the newsvendor of carrying newspaper. The other cost component $Q(x,z)$ is the cost of selling the newspaper, which comprises the second-stage decision of the amount of newspaper to be sold $y$ and the amount of newspaper to be returned $w$, given the demand $z$. The formulation of the second-stage is

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

[1]In subsequent works (Claus, 2022; Claus, 2021), the results have been generalized to problems in which the random variable is present in the lower level's objective function as well. However, to the best of my knowledge, the same level of theoretical results have not yet been shown to problems in which the technology matrix (left-hand side) of the lower level is stochastic, which precludes the use of, for example, the farmer's problem in the proposed thought experiments.

in which $q$ is the selling price and $r$ is the return price.

To take into account the randomness of the demand, we assume it is not a parameter, but rather a random variable $z = Z(\omega)$, where $\omega$ belongs to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Therefore, the newsvendor problem becomes a traditional two-stage stochastic programming problem with recourse

$$\min_{x} \quad \mathbb{E}\left[f(x, Z(\omega))\right] = cx + \mathbb{E}\left[Q(x, Z(\omega))\right]$$
$$\text{s.t.} \quad 0 \leq x \leq u.$$

Note that the above, following Birge and Louveaux (2011), assumes a risk-neutral position of the upper-level agent. However, the problem could more generally be formulated with any risk measure $\mathcal{R} : \mathcal{X} \longrightarrow \mathbb{R}$, where $\mathcal{X}$ is a linear subspace of $L^0(\Omega, \mathcal{F}, \mathbb{P})$, the space of all measurable random variables, such that $f(x, Z(\omega)) \in \mathcal{X} \forall 0 \leq x \leq u$.

## The Multiproduct Assembly Problem

Suppose an assembly plant that produces $n$ products, and each product requires zero or more of $m$ distinct parts. The plant manager has to decide the amount $\boldsymbol{x} = (x_1, \ldots, x_m)$ of parts to order before knowing how many products will be demanded by the customers. After the demand of each product $\boldsymbol{d} = (d_1, \ldots, d_n)$ is known, there is still a decision on which products will be assembled and sold, given the current stock of parts. Following (mostly) the notation of Shapiro, Dentcheva, and A. P. Ruszczyński (2009), the parametric problem can be defined as

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}, \boldsymbol{d}) = \boldsymbol{c}^T \boldsymbol{x} + Q(\boldsymbol{x}, \boldsymbol{d})$$
$$\text{s.t.} \quad 0 \leq \boldsymbol{x} \leq \boldsymbol{u},$$

where $\boldsymbol{c} = (c_1, \ldots, c_m)$ is the vector of acquisition costs of each part, while $Q(x, d)$ is the second-stage problem of deciding which products to assemble and how many of each. Precisely,

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

where $y_i$ is the amount of product $i$ assembled, that is sold at a price $q_i$, $w_j$ is the amount of part $j$ that remains and is salvaged at a price $r_j$, and each entry $a_{ij}$ of matrix $A$ indicates the amount of part $j$ that is required to assemble product $i$.

Once again, modeling the demand vector as a random variable $D(\omega)$ renders the two-stage problem

$$\min_x \quad \mathcal{R}\left[f(\boldsymbol{x}, D(\omega))\right]$$
$$\text{s.t.} \quad 0 \leq \boldsymbol{x} \leq \boldsymbol{u},$$

which is now presented with a risk measure $\mathcal{R}$.

## Risk-Averse Two-Stage Linear Programming

Both the newsvendor and the multiproduct assembly problems fall into the category of two-stage stochastic linear programs with fixed recourse and fixed technology matrix (cf. Birge and Louveaux, 2011, Sec. 3.1, and Shapiro, Dentcheva, and A. P. Ruszczyński, 2009, Sec. 2.1). In other words, in the risk-neutral scenario, they are special cases of the general problem

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T\boldsymbol{x} + \mathbb{E}\left[Q(\boldsymbol{x}, Z(\omega))\right]$$
$$\text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}$$
$$\boldsymbol{x} \geq 0,$$

in which

$$Q(\boldsymbol{x}, \boldsymbol{z}) = \min_{\boldsymbol{y}} \quad \boldsymbol{q}^T\boldsymbol{y}$$
$$\text{s.t.} \quad W\boldsymbol{y} \leq T\boldsymbol{x} + \boldsymbol{z}$$
$$\boldsymbol{y} \geq 0.$$

Note that the original formulations of both Birge and Louveaux (2011) and Shapiro, Dentcheva, and A. P. Ruszczyński (2009) consider equality constraints on the second stage, which are equivalent to the formulation with inequalities above through the addition of slack variables. Furthermore, I also consider a fixed second-stage cost, as is the case of both problems of interest.

In a risk-averse setting, as was already briefly discussed in the problems' presentation, the objective function of the general problem is considered

4

to be a risk measure over the upper-level cost (Noyan, 2012; Miller and A. Ruszczyński, 2011; Ahmed, 2006). Formally, the objective becomes $\mathcal{R}\left[f_{\mathrm{SP}}(\boldsymbol{x}, Z(\omega))\right] x$, where $f_{\mathrm{SP}}(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{c}^T \boldsymbol{x} + Q(\boldsymbol{x}, \boldsymbol{z})$, and $\mathcal{R} : \mathcal{X} \longrightarrow \mathbb{R}$ is a risk measure over a linear subspace of all $\mathcal{F}$-measurable random variables in $(\Omega, \mathcal{F}, \mathbb{P})$, containing the random variables $f_{\mathrm{SP}}(\boldsymbol{x}, Z)$ for every feasible $\boldsymbol{x}$. A common risk measure (see Noyan (2012)) is the combination of the expectation and a dispersion statistic $\mathbb{D}$, resulting in a mean-risk minimization problem

$$\min_{\boldsymbol{x}} \quad \mathbb{E}\left[f_{\mathrm{SP}}(\boldsymbol{x}, Z(\omega))\right] + \lambda \mathbb{D}\left[f_{\mathrm{SP}}(\boldsymbol{x}, Z(\omega))\right]$$
$$\text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}$$
$$\boldsymbol{x} \geq 0,$$

where $\lambda$ is a weighting factor. I note, for completion, that the operator $\mathcal{R} = \mathbb{E} + \lambda \mathbb{D}$ fits the previous definition of risk measure.

## Bilevel Stochastic Linear Programming

In contrast to two-stage problems, bilevel stochastic linear problems (Burtscheidt and Claus, 2020) are formulated in their parametric form as

$$\min_{\boldsymbol{x}} \quad f_{\mathrm{BL}}(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{c}^T \boldsymbol{x} + \min_{\boldsymbol{y}} \left\{ \boldsymbol{q}_u^T \boldsymbol{y} : \boldsymbol{y} \in \Psi(\boldsymbol{x}, \boldsymbol{z}) \right\}$$
$$\text{s.t.} \quad \boldsymbol{x} \in X,$$

where $X$ is a nonempty polyhedron and

$$\Psi(\boldsymbol{x}, \boldsymbol{z}) = \arg\min_{\boldsymbol{y}} \quad \boldsymbol{q}_l^T \boldsymbol{y}$$
$$\text{s.t.} \quad W\boldsymbol{y} \leq T\boldsymbol{x} + \boldsymbol{z}$$

is the set of all optimal solutions of the lower level problem. Note that an optimistic formulation is considered, which is equivalent to saying that the lower-level agent favors the upper-level whenever that does not affect the lower-level objective.

Replacing the parameter vector $\boldsymbol{z}$ by the random variable $Z(\omega)$ leads to a formulation similar to that of the two-stage general problem in the risk-averse scenario:

$$\min_{\boldsymbol{x}} \quad \mathcal{R}\left[f_{\mathrm{BL}}(\boldsymbol{x}, Z(\omega))\right]$$
$$\text{s.t.} \quad \boldsymbol{x} \in X.$$

The similarities between the two classes of problems are certainly larger than their differences. Both problems model sequential decision-making, and consider the realization of the random variable between the first and the second decision. A significant difference, however, lies in the coordination of both stages. Where two-stages problem assume that the second-stage objective is precisely the quantity of interest for the first-stage agent, bilevel stochastic problems consider more general cases, in which both levels have different objectives with respect to the decision variable $\boldsymbol{y}$. In fact, the bilevel problem reduces to a two-stage problem whenever $\boldsymbol{q}_u = \boldsymbol{q}_l$.

The differences render new challenges for deriving optimality conditions of bilevel stochastic linear problems. Such is extensively demonstrated in the work of Burtscheidt, Claus, and Dempe (2020), which were the first to achieve such results for the risk-averse bilevel problem considered in this work, that is, with the random variable showing only on the right-hand side of the lower level. The authors achieved these results by proving the Lipschitzian properties of the upper-level cost function, assuming the equivalent of complete recourse for the bilevel case. Then, they show that the continuity properties of the risk measure $\mathcal{R}$ translate to $\mathcal{R}\left[f_{\mathrm{BL}}(\boldsymbol{x}, Z(\omega))\right]$, which leads to optimality conditions for problems with convex risk measures. In fact, the authors even show single-level reformulations for problems with random variables under discrete distributions and commonly used risk measures, which enables the use of traditional exact algorithms. Finally, further work extended these results to problems where the random variable and the upper-level variable appear quadratically in the lower-level objective function Claus, 2021; Claus, 2022.

## Objectives

In this work, I will assume a variation of the newsvendor and the multiproduct assembly problems that results in risk-averse bilevel stochastic linear programming formulations. With this, I aim to study such bilevel problems taking as a start position the results for two-stage stochastic problems with recourse seen in class. Furthermore, having classic textbook problems at hand will help me to provide the results from Burtscheidt, Claus, and Dempe (2020) in greater detail by guiding the assumptions made and contextualizing the proofs. I expect to unravel the known results with respect to optimality conditions and investigate their proofs thoroughly, leading to an analysis of reformulations given common risk measures and the exact algorithms that might be appropriate for them.

I believe the choice of problems is particularly suitable for the thought experiments. Not only because these problems fit the lower-level definition of Burtscheidt and Claus (2020) with respect to the influence of the random variable, but also because of both problems are very similar. In fact, the original notation from Shapiro, Dentcheva, and A. P. Ruszczyński (2009) was purposefully modified to highlight such similarities. The multiproduct assembly can be seen as a multidimensional version of the newsvendor problem, while adding tying constraints to the second-stage $y$ variables. On top of that, the known analytical solution for the second stage of the newsvendor problems reduces most of the results to trivialities, which serves as a bridge between the theory for two-stage stochastic programming and the results for bilevel stochastic linear problems. In other words, I believe the investigation of the newsvendor problem in its bilevel version will serve as an introduction for the more complex differences that exist between the traditional mutliproduct assembly product and its proposed bilevel variation.

In face of the above, my proposed objectives for this final project are:

**Formulate a variation of the newsvendor and the multiproduct assembly problems**  Adapt two classical two-stage stochastic programming problems into bilevel formulations with risk-aversion, assuming that the two stages are managed by different agents with distinct objectives.

**Investigate optimality existence given commonly used risk measures**  Apply the results from Burtscheidt, Claus, and Dempe (2020) to prove the existence of optimal solutions for the expectation, the value-at-risk and the conditional value-at-risk as risk measures, re-stating the fundamental theorems while providing further details on the proofs and the implications of the structures in each problem.

**Investigate optimality conditions**  Investigate the differentiability of both problems given the three risk measures, which leads to necessary optimality conditions in terms of directional derivatives and the application of gradient-based solution algorithms.

**Derive reformulations for finite discrete distributions**  Formulate the two problems under the three risk measures as (deterministic) bilevel linear problems, which can be solved by incorporating the lower-level KKT conditions in the upper level.

Although achieving the above objectives do not require any novel result, the expected results are not straightforward guaranteed by Burtscheidt, Claus, and Dempe (2020). This is because the main theorems and lemmas have several nuances with respect to the components of the problem. An example of such is the partial randomness of the right-hand side of both problems' lower-level, which implies that the probability measure is not absolutely continuous w.r.t. the Lebesgue measure, a condition whose absence may preclude the applicability of the differentiability results (cf. Burtscheidt, Claus, and Dempe, 2020, Sec. 4). Therefore, it may be the case that some of the results expected require potentially novel results, which is out of the scope of this project.

# References

Ahmed, Shabbir (May 1, 2006). "Convexity and Decomposition of Mean-Risk Stochastic Programs". In: *Mathematical Programming* 106.3, pp. 433–446. ISSN: 1436-4646. DOI: `10.1007/s10107-005-0638-8`. URL: `https://doi.org/10.1007/s10107-005-0638-8` (visited on 10/31/2024).

Birge, John R. and François Louveaux (2011). *Introduction to Stochastic Programming*. 2nd ed. Springer Series in Operations Research. New York: Springer. 485 pp. ISBN: 978-1-4614-0237-4. URL: `https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2546867` (visited on 12/05/2024).

Burtscheidt, Johanna and Matthias Claus (2020). "Bilevel Linear Optimization Under Uncertainty". In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by Stephan Dempe and Alain Zemkoho. Cham: Springer International Publishing, pp. 485–511. ISBN: 978-3-030-52119-6. DOI: `10.1007/978-3-030-52119-6_17`. URL: `https://doi.org/10.1007/978-3-030-52119-6_17` (visited on 11/21/2024).

Burtscheidt, Johanna, Matthias Claus, and Stephan Dempe (Jan. 2020). "Risk-Averse Models in Bilevel Stochastic Linear Programming". In: *SIAM Journal on Optimization* 30.1, pp. 377–406. ISSN: 1052-6234. DOI: `10.1137/19M1242240`. URL: `https://epubs.siam.org/doi/10.1137/19M1242240` (visited on 12/03/2024).

Claus, Matthias (May 1, 2021). "On Continuity in Risk-Averse Bilevel Stochastic Linear Programming with Random Lower Level Objective Function". In: *Operations Research Letters* 49.3, pp. 412–417. ISSN: 0167-6377. DOI: `10.1016/j.orl.2021.04.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0167637721000596` (visited on 12/03/2024).

— (June 1, 2022). "Existence of Solutions for a Class of Bilevel Stochastic Linear Programs". In: *European Journal of Operational Research* 299.2, pp. 542–549. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2021.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0377221721010110 (visited on 12/03/2024).

Miller, Naomi and Andrzej Ruszczyński (2011). "Risk-Averse Two-Stage Stochastic Linear Programming: Modeling and Decomposition". In: *Operations Research* 59.1, pp. 125–132. ISSN: 0030-364X. JSTOR: 23013110. URL: https://www.jstor.org/stable/23013110 (visited on 12/07/2024).

Noyan, Nilay (Mar. 1, 2012). "Risk-Averse Two-Stage Stochastic Programming with an Application to Disaster Management". In: *Computers & Operations Research* 39.3, pp. 541–559. ISSN: 0305-0548. DOI: 10.1016/j.cor.2011.03.017. URL: https://www.sciencedirect.com/science/article/pii/S0305054811000931 (visited on 12/07/2024).

Shapiro, Alexander, Darinka Dentcheva, and Andrzej P. Ruszczyński (2009). *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM Series on Optimization. Philadelphia, Pa.: Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104). 436 pp. ISBN: 978-0-89871-875-1. URL: http://epubs.siam.org/ebooks/siam/mos-siam_series_on_optimization/mp09 (visited on 12/05/2024).