Optimization for Lot-Sizing Problems Under Uncertainty: A Data-Driven Perspective

Paula Metzker[1,2]($\boxtimes$), Simon Thevenin[1], Yossiri Adulyasak[2], and Alexandre Dolgui[1]

[1] IMT Atlantique, LS2N - UMR CNRS 6004, 44307 Nantes, France
{paula.metzker-soares,simon.thevenin,alexandre.dolgui}@imt-atlantique.fr
[2] HEC Montréal, GERAD, Montreal, Canada
yossiri.adulyasak@hec.ca

Abstract. In a manufacturing context, the lot-sizing problems (LSP) determine the quantity to produce over a planning horizon. Often, the parameters used in the LSP models are unknown when the decisions are made, and this uncertainty has a critical impact on the quality of the decisions. However, the large amount of data that can nowadays be collected from the shop floor allows inferring information on the LSP parameters and their variability. Therefore, a recent research trend is to properly account for the uncertainty in the LSP optimization models. This work presents a survey on data-driven optimization approaches for the LSPs. We also provide a comparison of some promising optimization methodologies in the context of data-driven modeling of LSPs.

Keywords: Data-driven optimization · Lot-sizing problem · Optimization under uncertainties

1 Introduction

The lot-sizing problem (LSP) [16] determines the production lots over a planning horizon that minimize overall costs and maintains a satisfactory level of service. Due to its practical importance, the LSPs attracted a wide range of research from the manufacturing and mathematical optimization communities. In fact, production and distribution systems are settled in chaotic environment where production, quality, sales, purchasing, logistics, corporate, technical, accounting and marketing department are constantly affected by unexpected events. Thus, LSPs become inadequate to meet the needs of the industry if they are not simple enough to be adapted to changes in the environment [8].

Production and distribution systems face various sources of uncertainties (demand, lead time, production yield, among others) that affect the costs and service level associated with the lot-sizes. Traditionally, these systems dampen these uncertainties by changing parameters of the planning systems, such as the safety stock, safety lead-time, and re-planning frequency. Advances in computing technologies and the massive availability of data led to the design of data-driven optimizations to directly incorporate the uncertainties within the LSP, such as stochastic programming (SP) [5], robust optimization (RO) [1], and distributionally robust optimization (DRO) [17].

While SP models often seeks to minimize the expected costs over the distribution of the uncertain LSP parameters, RO models minimize the overall costs with regard to the worst-case realization of the LSP uncertainties. Finally, DRO extends stochastic optimization by taking into account the uncertain probability distributions of the unknown parameter. Even if these methodologies aim to optimize the LSPs by mitigating uncertainties, continual modification on LSP parameters leads to constant update of the production plans [7]. To overcome this issue, a data-based perspective of optimization methodologies emerges as a rather new and promising approach to compute production plans that are flexible to changes, and whose impacts due to unexpected events is more controllable.

The data-driven models often rely on a statistical analysis of the available data [3]. Bertsimas et al. [4] propose a data-driven approach based on sample approximation algorithms to choose the decision rules that perform best from the perspective of the worst-case within a stochastic process. Jiang and Guan [10] propose a data-driven methodology to obtain robust solutions from a chance-constrained problem with inaccurate probability distributions of the uncertain parameters. Then, they proposed a data-driven approach to solve the LSP under demand uncertainty based on sample average approximation algorithm [11]. Ning et al. [12] propose some artificial intelligence techniques have been investigated for labeling the available uncertain data and to compute near-optimal solutions through a data-driven approach. In addition, they present a data-driven via the DRO methodology. A state-of-the-art in data analytics and machine learning methods for process manufacturing in the light of big data approaches is presented in [13]. Zhao et al. [19] proposes a data-driven approach based on the kernel density estimation to represent the uncertain parameters into the optimization problems based on information from the historical data.

Although data-driven optimization emerges as a rather novel methodology to deal with non-deterministic optimization problems, a data-driven perspective of the LSPs is still missing in the literature. This has inspired us to develop a survey on data-driven optimization approaches applied to the LSPs. We are not interest in an exhaustive literature review, but in a survey of the existing literature on the data-driven optimization for LSP models. The remainder of this paper is organized as follows: Sect. 2 presents the application of the different optimization approaches starting from the data up to implementable decisions. Section 3 provides the advantages and issues of these methods in terms of computation time, tractability, flexibility in handling unforeseen events, and robustness of the

solution. Section 4 gives the main research areas of the data-driven optimization to handle the LSPs. Finally, Sect. 5 summarizes the main findings of this work.

2 Data-Driven Optimizations for the LSPs

The main steps of the data-driven optimization (DDO) are: **i.** the definition of the uncertain parameters distribution characteristics; **ii.** the analysis and processing of available data (eventually coming from various sources) to learn how to represent the uncertainties; **iii.** the formulation and modeling of the problem within the perspective of a chosen optimization method. Figure 1 gives a schematic view on this methodology, and its steps are described in the rest of this section.
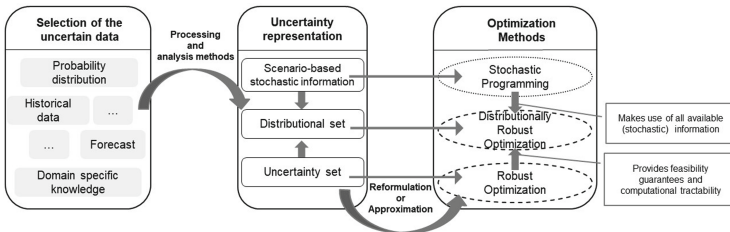

Fig. 1. DDO framework

i. Selection of uncertain data: On the one hand, ignoring uncertainties in LSPs leads to sub-optimal decisions. On the other hand, the inclusion of uncertainties increases the model complexity, and the solution might require large computation time and memory consumption. Consequently, the decision-maker must carefully analyze the historical data, forecast, probability distributions of data, experts insights, domain-specific knowledge, and any other available information to select the type of uncertainties to include in the optimization model. The decision-maker may consider the parameters whose value cannot be estimated accurately, and whose variance affects the decisions.

ii. Uncertainty representation: aims to incorporate partial information obtained from the uncertainties into the optimization methodology. For this, some data processing and analysis methods are used to manipulate uncertainties and extract as much useful and accurate information as possible. For the SP models, this step estimates the probability distributions with some statistical methods, such as analyzing historical data and the moment information, or some non-parametric statistical estimation [5]. For the RO models, the uncertainty sets are designed to preserve the computational tractability of robust models [2]. Consequently, these sets have well-defined structures such as *box and ellipsoid uncertainty set* [2]. Similarly, for the DRO, the distributional sets contain distributions with similar properties about the uncertainties, and these sets have well defined structures [18]. Among the more applied methods to build

uncertainty and distributional sets from data, we cite the statistical hypothesis testing validation [3] and machine learning techniques [12].

iii. Optimization methods: The solution approaches for the LSPs usually rely on mixed-integer linear programming, and the choice of the model must be adapted to the production context. First, multiple formulations of the LSPs exist, and the most efficient ones change depending on the context [9]. Second, the incorporation of uncertainty in these models depends on the decision framework [15]. Within a static decision framework, the lot-sizes are fixed for the entire horizon, and so they are frozen. This situation corresponds to a two-stage stochastic optimization model or a classical robust optimization model. Within a dynamic decision framework, decisions can be updated in each period t after that some uncertain parameters are revealed for periods up to t. This situation corresponds to a multi-stage stochastic optimization model or an adjustable robust optimization model. Finally, in a non-deterministic context, the models must include a mechanism to dampen uncertainty without escalating the costs. This mechanism may be a service level constraint [14], though an appropriate balance between the lost sales/backordering costs and inventory costs can also improve the quality of the solution.

Solving large-scale LSPs in an uncertain context can require intensive computations. To solve practical size instances with the SP method, the resolution approaches often rely on sampling methods, such as sample average approximation algorithm, or decomposition such as L-shaped, stochastic dual dynamic programming, or Progressive Hedging. The solution strategies for RO and DRO often cover the reformulation per constraint and dualization, and adversarial approaches, such as heuristics, branch-and-bound, or decomposition approaches.

3   Comparison of RO, SP, and DRO Methodologies

The choice of an optimization depends on decision makers' preferences, instance structures, available information, and expected trade-off in terms of solution quality and computing time. Although the problem involves three possible decision frameworks, namely static, static-dynamic and dynamic strategies, we focus our study on the static case. Based on the existing literature, we present a brief analysis of the performance of each method in terms of scalability, tractability, conservatism and flexibility of adaptation to unforeseen events.

First, RO can be used when little or no data is available, whereas SP requires large historic data to accurately estimate the probability distribution. The SP formulations often suffer from scalability issues, because the model must properly account for the uncertainty (often by relying on large scenario samples). On the contrary, RO approaches often remain tractable for practical size problems, when the robust model can be formulated as a convex problem.

RO typically optimizes the worst possible realization of the uncertain parameters, which leads to conservative solutions. On the contrary, the SP optimizes the expected costs, but it requires an sufficiently good probability distribution. Therefore SP solutions are poorly flexible to unforeseen or misrepresented events.

The DRO proposes a trade-off between these two approaches since it compensates the conservatism of the RO by taking advantage of partial distributional information obtained from the probability distributions from the SP framework. Hence, DRO emerges as a method sufficiently flexible to unforeseen events, while it remains computationally tractable, and it provides a less conservative solution.

4 Discussion

The LSPs have been studied for decades, but it still has room for improvements and further investigation. The SP, RO and DRO methodologies have stood out either for the quality of the solution, or for the ease of calculation within an uncertain environment. Most of the studies on optimization under uncertainties rely on statistical approaches to expound available data, and to reduce the conservatism of solutions. The growth of learning methods and the increase in data availability have motivated recent works to develop some data-based approaches that deal with the uncertain information [6].

Data-driven approaches enhance the quality and the performance of the methodologies for optimization under uncertainties. Among the methodologies presented in this work, a natural application of the data-driven methodology leads to distributionally robust optimization. Here, some data processing and analysis is implemented to extract more quality information from the decision context, and propose better predictions about the information of the uncertain parameters. Therefore, a worst-case perspective can be applied over all gathered distributional data. Thus, an optimization combining the expected value from the SP and the robustness from RO would propose more realistic solutions.

DDO is an emerging field of research, whose techniques, approaches, and applications are still under development. More applications should be analyzed to report the feasibility and tractability of the proposed approaches in real applications. On the other hand, further investigation into the application and feasibility of different data-driven approaches must be carried out to deal with different versions of LSPs, considering not only different versions of the problem but also different uncertain parameters. In addition, a deeper study of data processing, data analysis, and machine learning techniques is envisaged to develop data-driven approaches, and to better understand their challenges, limits, and prospects for improving different optimization methodologies.

5 Conclusion

There is a growing interest about data-driven optimization for lot-sizing problems. These methods learn the uncertainty representation from the data, and they incorporate these uncertainties in the lot-sizing models. The DRO can be applied to tackle different types of uncertainties which would derive more benefits from the DDO approach, being more stochastic or robust according to the

decision maker's needs and decision environment. Although DRO is a promising method integrating data-driven approaches to the conception of flexible production plans, there is a lack of research on the DDO methods for the LSPs. Further studies should be envisaged to fulfill this knowledge gap.

References

1. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)
2. Bertsimas, D., Brown, D.B., Caramanis, C.: Theory and applications of robust optimization. SIAM Rev. **53**(3), 464–501 (2011)
3. Bertsimas, D., Gupta, V., Kallus, N.: Data-driven robust optimization. Math. Program. **167**(2), 235–292 (2017). https://doi.org/10.1007/s10107-017-1125-8
4. Bertsimas, D., Shtern, S., Sturt, B.: A data-driven approach for multi-stage linear optimization. Available at Optimization Online (2018)
5. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer, New York (2011). https://doi.org/10.1007/978-1-4614-0237-4
6. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**(3), 595–612 (2010)
7. Dolgui, A., Ammar, O.B., Hnaien, F., Louly, M.A., et al.: A state of the art on supply planning and inventory control under lead time uncertainty. Stud. Inform. Control **22**(3), 255–268 (2013)
8. Dolgui, A., Prodhon, C.: Supply planning under uncertainties in MRP environments: a state of the art. Ann. Rev. Control **31**(2), 269–279 (2007)
9. Gruson, M., Bazrafshan, M., Cordeau, J.F., Jans, R.: A comparison of formulations for a three-level lot sizing and replenishment problem with a distribution structure. Comput. Oper. Res. **111**, 297–310 (2019)
10. Jiang, R., Guan, Y.: Data-driven chance constrained stochastic program. Math. Program. **158**(1–2), 291–327 (2016)
11. Jiang, R., Guan, Y.: Risk-averse two-stage stochastic program with distributional ambiguity. Oper. Res. **66**(5), 1390–1405 (2018)
12. Ning, C., You, F.: Data-driven stochastic robust optimization: general computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era. Comput. Chem. Eng. **111**, 115–133 (2018)
13. Shang, C., You, F.: Data analytics and machine learning for smart process manufacturing: recent advances and perspectives in the big data era. Engineering **5**(6), 1010–1016 (2019)
14. Tempelmeier, H.: Stochastic lot sizing problems. In: Smith, J., Tan, B. (eds.) Handbook of Stochastic Models and Analysis of Manufacturing System Operations. ISOR, vol. 192, pp. 313–344. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-6777-9_10
15. Thevenin, S., Adulyasak, Y., Cordeau, J.F.: Material requirements planning under demand uncertainty using stochastic optimization. Prod. Oper. Manag. **30**(2), 475–493 (2020)
16. Wagner, H.M., Whitin, T.M.: Dynamic version of the economic lot size model. Manag. Sci. **5**(1), 89–96 (1958)
17. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Oper. Res. **62**(6), 1358–1376 (2014)

18. Zhang, Y., Shen, Z.J.M., Song, S.: Distributionally robust optimization of two-stage lot-sizing problems. Prod. Oper. Manag. **25**(12), 2116–2131 (2016)
19. Zhao, L., Ning, C., You, F.: A data-driven robust optimization approach to operational optimization of industrial steam systems under uncertainty. Comput. Aided Chem. Eng. **46**, 1399–1404 (2019)