Advanced Decomposition Methods in Stochastic Convex Optimization

Jakub Kůdela

Opponent's Report

The Dissertation consists of five parts. The first part gives an introduction to the parts of the stochastic programming theory which are used later. The second part is a reprint of journal article, introducing a decomposition algorithm for the solution of a convex two stage stochastic programs. The third part contains another journal article computing an optimal waste transportation by means of the aforementioned technique. The fourth part – the only "non-article" one – is discussing a solution method for chance constrained stochastic programming problems. The fifth part contains another application-oriented article, dealing with an optimal beam design by means of the technique discussed in part four.

In my opinion, each part (two to five) brings some scientific contribution, either in theory or in application. The text is rather well written and understandable. There are, however, some issues, which should be, in my opinion, discussed and/or clarified during the defense and possibly reflected when publishing the unpublished part.

Generally, I like the applied parts more than the theoretical ones. Although, as a theorist, I am supposed to prefer theoretical results, here I value the applied parts more because they are, simply saying, well done, following the best traditions of the candidate's institution. Both the applied articles include reasonable formulations of the models, their efficient solution, and, last but not least, clear exposition. Apart from several minor issues (see below), I have no serious objections to the applied parts (three and five).

This does not mean that the theoretical parts are not valuable; both are interesting, dealing with challenging problems, easily readable; however, there is still much to be done better here. Bellow I summarize my most serious objections. In any case, my aim is not to discourage the candidate from theoretical work. I deeply hope that he takes them as an inspiration rather than as a critique.

First of all, there is too much borrowed material in the theoretical parts bringing no or little added value (which could be e.g. better explanation or an interesting synthesis). Unlike textbooks or surveys, there is little sense in repeating known facts in research articles (hopefully correctly I suppose part four to be a ground of an article). Instead, only absolute minimum needed for understanding of the original contribution should be presented. The Bernstein approximation, for instance, needs not to be explained in detail in part four; a brief description and a reference to the original would have been sufficient. If, for instance, my article were about an engine I invented, then I certainly would have to describe it to details and demonstrate that it is better than e.g. Diesel one, but I would not write about the principles of the Diesel. The same applies to the the material the original contributions build on – the L-shaped method (part two) and the probabilistic robust design (part four). For the former, excellent monographs exist, while the latter is clearly described in primary sources. Needles to say that redundant reformulations do not help a reader much (if he wants to understand, she will have to go to the primary sources anyway) and are wasting of the energy of the the author, editors and referees (by the way, I did not check correctness of the borrowed parts).

Further, in both the theoretical parts, it not clearly distinguished, what the author's original contribution is. At the first look, it could seem that the only indisputable scientific contribution of part two is the warm start procedure. Even though it is said that the algorithm "is the first implementation of the GBD for two-stage stochastic convex programming problems of the form (2.3.1)", it is no way clear, without a detailed knowledge of the subtleties of decomposition methods, what effort had to be paid to apply the method to this kind of problems. If it is only a routine application, then it can hardly be seen as a scientific contribution; if, on the other hand, some non-trivial problems had to be solved, then this should be stressed properly to persuade the journal to publish the text and to persuade the reader that using the proposed method pays off. Nevertheless, the part has been (luckily) published as an article (yet it in a special issue of a Q4 journal), so let us concentrate to part four.

At the first look, the situation looks analogous here, yet the contribution is clearly more significant. The "discarding part" of the presented algorithm is "very similar" to a certain existing method (no differences are reported), so the only completely original contribution seems to be the pooling part of the algorithm. Even though this idea is interesting, the algorithm is implemented and numerically examined, still there is no proof that the pooling, taking non-zero computational resources, pays off. The comparison with the Bernstein method is good; however, it does not answer the question whether the new algorithm performs better than [88], which it is based on (at least I did not find any such information in the numerical results). So, as a minimum, a comparison (both theoretical and numerical) of the new algorithm and [88] has to be done for the text to have chance to be published; even then, however, I would be afraid of rejections due to insufficient contribution, if the topic were not elaborated more deeply. Let me give some hints (=possible topics of discussion during the defense)

- Assumption 4.2.5 (Feasibility) is rather restrictive. Could not something be done here? Maybe a regularization?
- Assuming the measurability of $\mathcal{V}$ is not necessary as it is, if I am not mistaken, guaranteed by the measurability of g – the problem is equivalent to the measurability of the parameter dependent integral, discussed e.g.

here.[1] Alternatively, the proof may be done without functional analysis, via discrete, approximation of the probability: the mapping from x to the sum, approximating the probability, is measurable, so the probability, being a limit of the sums, is measurable, too, by [Kallenberg 2001, Lemma 1.10. (ii)].

- The idea of linearization seems interesting; however, it is not elaborated further in the present text. What are the benefits, when the convex solvers in fact do the same? Does it spare time/resources? Etc.
- Could not something theoretical to be said to the trade-off between the level and the optimum?
- Could not the time complexity of the new algorithm be evaluated theoretically?
- Cannot the special structure of the problem be used for more efficient computation?
- Could not paralellization help to decrease the time complexity?
- Anything else?

My final major objection is that the original material is not presented with sufficient rigor. The first keyword here is *replicability*: a sufficiently educated reader (say having PhD in the area the journal specializes in) has to be able to implement the proposed methods based on the text and other available sources. Thus, newly proposed methods, have to be described rigorously enough. Yet the presented text not tragically inexact (at least I could mostly understand it well), there are places when the reader has to guess from the context (e.g. in the algorithm description on page 43, see bellow).

Further, everything in a scientific text has to be proved. Here, for instance, the author claims that, after the pooling part, the set $\mathcal{I}$ contains (all?) the supporting scenarios. Even though the intuition says that it should be like that, it is not the task of the reader or referee to prove this fact. Here, if I am not mistaken, the statement is not true (again see below). I also miss proofs of the constraint reductions of the beam design problem (see below, too).

Seeing that considerably more space has been devoted to objections than to praise in this report, I feel obliged to stress, that, despite my objections, I see the overall quality of the work to be above standard and the text, as a whole, to demonstrate the candidate's ability of scientific work, both theoretical and applied. Thus, the Thesis can be recommended for acceptance to the defense and the candidate can be recommended to obtain the Ph.D. degree.

Finally I list some, mostly minor, issues, which I found when reading the text

[1] https://math.stackexchange.com/questions/1136665/is-the-integral-of-a-measurable-function-measurable-wrt-a-parameter

- The title of the Dissertation only partly reflects the content: in my view, only parts two and three use algorithms which can be named decomposition.
- The sentence at the turn of p3 and p4 is unintelligible.
- What is r on P7L13- (13th line from bellow)?
- P7L2- The error in probability cannot be compared with the other inaccuracies of the problem. What if the whole planet were destroyed given the event with 10^{-6} probability?
- P9L15- It is unusual to denote a constant by $O(1)$. If that means that, with $\epsilon \to 0$, the formula holds for some function $C(\epsilon)$ asymptotically equivalent to a constant, then small o should be used.
- P15, Step 2A. What does "Get $(\tilde{y}, \tilde{v})$" mean? (Of course I know it, but the algorithm should be replicable beyond any doubts)
- In part two, the solution of the problem (the main topic of the Thesis) is mentioned only briefly. In my opinion, some discussion (beyond the text of the article) might have been added. As it is not, the topic should be discussed during the defense at least.
- The last paragraph of 3.1 is non-standard: I would either refer to all the sections or to none.
- P35L6- It should be said that the set is measurable for each x (not to confuse with the joint measurability)
- P372- The author says that the ultimate goal is to find a worst-case ϵ-level solution. Later, however, no worst-case solution is sought as the objective does not depend on the chance
- 4.3.1 It is said that $\mathcal{I}$ contains the supporting scenarios. If "all s.s." is meant, then the statement is not true, as some supporting scenarios can hide below δ. Even with $\delta = 0$, this fact is not immediately clear. As this is the original part of the text, this assertion has to be proved or at least demonstrated.
- P43L3 It is not clear from the algorithm description, what to repeat k times.
- P73below (step 1) Does "randomly" mean "from $U(0,1)$"?
- P76above (definition of the problem). Some symbols (e.g v_M or w_M) are not defined here. I understand that they were defined earlier; however, these definition concerned different versions of the problem. Generally, including multiple versions of the problem and sharing the notation is not a very lucky solution. Maybe it would be better to describe the "previous" versions only informally and define the "true" one properly. Or, at least, a reference to the definitions of the missing symbols should have been added.

- P76L7 It is not clear at the first look that the $2S$ constraints may be reduced to 2
- P76L8- It is not clear at the first look that there are only two supporting scenarios

In Chocerady on August 30th, 2019, Martin Šmíd, opponent