Structure-exploiting methods for large-scale stochastic optimization

Welington de Oliveira François Pacaud

January 13, 2025

Research domains

Mathematical programming, stochastic optimization, interior-point methods.

1 Subject

When a decision-maker has to face uncertain future outcomes, they usually have no other choice than resorting to stochastic optimization. In contrast to their deterministic counterparts, stochastic optimization methods model the unknown parameters explicitly using probability distributions: as a consequence, the decision variables become themselves random variables [20]. This approach is widely used in energy systems to compute the operational planning under uncertainty (in future demands or in renewable productions) [19]. If in addition the decision-maker has to manage storage (hydro-valleys, fleet of batteries), the problem inherits a sequential nature that is usually encoded through a dynamic, motivating the introduction of *multistage stochastic optimization problems* where the decision-maker makes decisions at multiple points in time.

As soon as the decision variables are encoded as random variables, the stochastic optimization problems becomes infinite dimensional. One has to resort to discretization methods to render it tractable. Unfortunately the discretization is subject to the *curse of dimensionality*, which prohibits an exact solution when the dimension of the uncertainties becomes large. As a consequence, one has to find relevant approximation scheme to compute *near-optimal* decisions, the *sampling average approximation* (SAA) being one of the most prominent method.

Once the uncertainties discretized, the problem rewrites as a mathematical program in finite dimension, solvable using classical optimization algorithm. However, the size of the problem grows linearly as we increase the number of samples in our discretization scheme. As such, the resulting problem is often large-scale, pushing the solvers to their limits. For that reason, it is recommended to exploit the structure of the problem (associated to the sampling we used before) in order to alleviate the numerical burden and speed-up the solution method.

This thesis aims at providing methodological contributions to the development of structure-exploiting solvers for large-scale stochastic optimization. To that goal, we will leverage the interior-point method (IPM) [15] — known for its versatility — to exploit the structure of the problem at the linear algebra level. The student will develop new solution algorithms based on IPM to solve previously intractable problems in large-scale stochastic optimization. In particular, we will harness the power offered by modern GPUs to achieve better scalability, using the modular framework offered by the GPU-accelerated solver MadNLP [21]. The development will lead to the development of an user-friendly package for multistage stochastic optimization, developed in Julia.

2 Scientific Project

In discrete time, every dynamical system has an inherent sequential structure. The structure can be exploited afterwards in the solution algorithm, either at the level of the optimization (using decomposition methods) or at the linear algebra level (using structured linear algebra).

2.1 Objective 1: two-stage program under ambiguity set

The two-stage stochastic program is one of the most standard stochastic optimization problem. It assumes a sequential information structure: after an initial decision $\boldsymbol{x}$ is decided, an uncertainty $\boldsymbol{w}$ occurs and the decision-maker adjusts their decision using a recourse $\boldsymbol{y}(\boldsymbol{w})$ depending on the uncertainty. The decision structure is sequential:

$$x \rightsquigarrow \boldsymbol{w} \rightsquigarrow \boldsymbol{y} \; .$$

Using probabilistic notations, the two-stage program is formally defined as

$$\min_{x,\boldsymbol{y}} \; c^\top x + \mathbb{E}\big[q(\boldsymbol{w})^\top \boldsymbol{y}\big] \quad \text{subject to} \quad \begin{cases} Ax = b \; , \\ T(\boldsymbol{w})x + W(\boldsymbol{w})\boldsymbol{y} = b(\boldsymbol{w}) \; . \end{cases} \tag{1}$$

The two-stage problem (1) has a separable structure we can exploit in the optimization: the idea is to *aggregate* the recourse problem inside a *value function* $Q(x, \boldsymbol{w})$ parameterized by the first stage variable x and the realization of the uncertainty $\boldsymbol{w}$. Under convexity assumption, the value function $Q(\cdot, \boldsymbol{w})$ can be approximated using a piecewise affine models (using a collection of cuts), the first-stage problem $\min_x c^\top x + \mathbb{E}\big[Q(x, \boldsymbol{w})\big]$ being solved using a coordination algorithm to find a global optimum (using a Benders or a bundle method [16, 7]). Alternatively, if the probability distribution of $\boldsymbol{w}$ has a finite support, the problem (1) can be solved efficiently directly using a structure exploiting interior-point method [12, 24, 11].

In the last decade, *distributionally robust optimization* (DRO) [23, 18] has become a prominent research area. DRO is a generalization of the stochastic and robust optimization frameworks, where the decision-maker optimizes directly over the space of probability distributions to look at the worst-case distribution inside a given *ambiguity set* $\mathcal{P}$. Using the DRO setting, the two-stage program (1) rewrites as

$$\min_{x,\boldsymbol{y}} \; \max_{\mathbb{P}\in\mathcal{P}} \; c^\top x + \mathbb{E}_{\mathbb{P}}\big[Q(x, \boldsymbol{w})\big] \; . \tag{2}$$

The first part of this PhD project will be devoted to the study of efficient solution methods for (2). Depending on the nature of the ambiguity set $\mathcal{P}$, the problem (2) can be reformulated as a second-order cone program (SOCP) or a semi-definite program (SDP), with a specific structure we can exploit in the interior-point algorithm. A more efficient algorithm would unlock the potential to solve larger instances, improving the relevance of the DRO methodology. Using a structure-exploiting interior-point would be a novelty considering the recent literature in the DRO community, where only first-order (ADMM, Frank-Wolfe) and generic second-order interior-point methods (Mosek) have been studied so far [13].

2.2 Objective 2: multistage stochastic program

The multistage stochastic program generalizes the previous two-stage program (1): this time, the decision-maker can take decision at different time-steps, all coupled together [20]: At a given time t, the decision-maker observes the last uncertainties $\boldsymbol{w}_0, \cdots, \boldsymbol{w}_{t-1}$ before taking a new decision $\boldsymbol{u}_t$. The information structure writes informally:

$$\boldsymbol{w}_0 \rightsquigarrow \boldsymbol{u}_0 \rightsquigarrow \boldsymbol{w}_1 \rightsquigarrow \boldsymbol{u}_1 \rightsquigarrow \cdots \rightsquigarrow \boldsymbol{u}_{T-1} \; .$$

The knowledge accumulated so far is usually aggregated in a *state* variable $\boldsymbol{x}_t$ that depends on all the previous uncertainties $\boldsymbol{w}_0, \cdots, \boldsymbol{w}_t$ (also known as the history). The multistage problem is usually formulated as a stochastic optimal control problem:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{u}} \; & \mathbb{E}\Big[\sum_{t=0}^{T-1} \ell_t(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t) + K(\boldsymbol{x}_T)\Big] \\ \text{subject to} \quad & \boldsymbol{x}_{t+1} = f_t(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t) \quad \forall t = 0, \cdots, T-1 \\ & g_t(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t) \leq 0 \text{ p.s.} \quad \forall t = 0, \cdots, T-1 . \end{aligned} \tag{3}$$

We aim at minimizing the operating cost $\ell_t(\cdot)$ over a given horizon T, the time-steps being coupled together through the dynamics $f_t(\cdot)$. Beside, the problem considers operational constraints encoded by the functions $g_t(\cdot)$, to be satisfied almost surely. The solution of (3) is usually given as a sequence of *policy* functions $\{\pi_t\}_{t=0,\cdots,T-1}$ that explicit the dependence of the control w.r.t. the past history: $\boldsymbol{u}_t = \pi_t(\boldsymbol{w}_0, \cdots, \boldsymbol{w}_t)$.

The problem (3) is generic, but is very difficult to solve in practice. There exists two main solution methods [20]. (i) If the probability are discrete, the problem (3) can be formulated on a scenario tree. However, the dimension blows up exponentially as we increase the number of time steps. (ii) Alternatively, (3) can be solved using Dynamic Programming, provided the uncertainties are stagewise independent. This time, the complexity becomes exponential in the number of states $\boldsymbol{x}_t$. If the problem remains convex, the SDDP algorithm [3, 17] is often the method of choice to solve (3), but the algorithm is slow to convergence and also subject to the curse of dimensionality w.r.t. the state's dimension.

The second part of this study will investigate the solution of Problem (3) using a variational approach, to target a solution with a stochastic version of the interior-point algorithm. The variational approach has been applied to problems formulated on scenario-tree [22, 25, 8], but as discussed in the paragraph before its complexity increases exponentially with the number of time-steps. Here, we will investigate two alternative approaches. First, we will investigate a stochastic Newton algorithm to solve (3). Assuming the uncertainties are Gaussian and linearizing locally the problem along a given (stochastic) iterate, we obtain a linear-quadratic Gaussian (LQG) problems whose solution is given using Riccati recursions [10]. This approach fits well the recent developments porting on stochastic Sequential Quadratic Programming (SQP) methods [14, 6], but with limitations (the Gaussian assumption being the strongest one). Alternatively, one can restrict the search space by parameterizing the policies $\pi_0, \cdots, \pi_{T-1}$ with a set of parameters with fixed dimension $\theta \in \mathbb{R}^p$ (e.g., affine or piecewise affine policies [9, 5]). As we restrict the search space, the problem (3) becomes easier to solve, but the sensitivities of the problem has to be evaluated. To that goal, we will leverage recent development in automatic differentiation made in the machine learning community, using the *differentiable programming* formalism [1, 2, 4].

References

[1] A. Agrawal, B. Amos, S. Barratt, S. Boyd, S. Diamond, and J. Z. Kolter, *Differentiable convex optimization layers*, Advances in neural information processing systems, 32 (2019).

[2] B. Amos and J. Z. Kolter, *Optnet: Differentiable optimization as a layer in neural networks*, in International conference on machine learning, PMLR, 2017, pp. 136–145.

[3] J. R. Birge and F. V. Louveaux, *A multicut algorithm for two-stage stochastic linear programs*, European Journal of Operational Research, 34 (1988), pp. 384–392.

[4] M. Blondel and V. Roulet, *The elements of differentiable programming*, arXiv preprint arXiv:2403.14606, (2024).

[5] M. Bodur and J. R. Luedtke, *Two-stage linear decision rules for multi-stage stochastic programming*, Mathematical Programming, 191 (2022), pp. 347–380.

[6] F. E. Curtis, D. P. Robinson, and B. Zhou, *Sequential quadratic optimization for stochastic optimization with deterministic nonlinear inequality and equality constraints*, SIAM Journal on Optimization, 34 (2024), pp. 3592–3622.

[7] A. Frangioni, *Standard Bundle Methods: Untrusted Models and Duality*, in Numerical Nonsmooth Optimization, A. M. Bagirov, M. Gaudioso, N. Karmitsa, M. M. Mäkelä, and S. Taheri, eds., Springer International Publishing, Cham, 2020, pp. 61–116.

[8] G. Frison, D. Kouzoupis, M. Diehl, and J. B. Jørgensen, *A high-performance riccati based solver for tree-structured quadratic programs*, IFAC-PapersOnLine, 50 (2017), pp. 14399–14405.

[9] A. Georghiou, W. Wiesemann, and D. Kuhn, *Generalized decision rule approximations for stochastic programming via liftings*, Mathematical Programming, 152 (2015), pp. 301–338.

[10] E. Gobet and M. Grangereau, *Newton method for stochastic control problems*, SIAM Journal on Control and Optimization, 60 (2022), pp. 2996–3025.

[11] J. Gondzio and R. Sarkissian, *Parallel interior-point solver for structured linear programs*, Mathematical Programming, 96 (2003), pp. 561–584.

[12] J. Gondzio, R. Sarkissian, and J.-P. Vial, *Using an interior point method for the master problem in a decomposition approach*, European Journal of Operational Research, 101 (1997), pp. 577–587.

[13] G. A. Hanasusanto and D. Kuhn, *Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls*, Operations Research, 66 (2018), pp. 849–869.

[14] S. Na, M. Anitescu, and M. Kolar, *Inequality constrained stochastic nonlinear optimization via active-set sequential quadratic programming*, Mathematical Programming, 202 (2023), pp. 279–353.

[15] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer, 1999.

[16] W. d. Oliveira and C. Sagastizábal, *Bundle methods in the XXIst century: a bird's-eye view*, Pesquisa Operacional, 34 (2014), pp. 647–670.

[17] M. V. Pereira and L. M. Pinto, *Multi-stage stochastic optimization applied to energy planning*, Mathematical programming, 52 (1991), pp. 359–375.

[18] H. Rahimian and S. Mehrotra, *Distributionally robust optimization: A review*, arXiv preprint arXiv:1908.05659, (2019).

[19] L. A. Roald, D. Pozo, A. Papavasiliou, D. K. Molzahn, J. Kazempour, and A. Conejo, *Power systems optimization under uncertainty: A review of methods and applications*, Electric Power Systems Research, 214 (2023), p. 108725.

[20] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on stochastic programming: modeling and theory*, SIAM, 2021.

[21] S. Shin, M. Anitescu, and F. Pacaud, *Accelerating optimal power flow with gpus: Simd abstraction of nonlinear programs and condensed-space interior-point methods*, Electric Power Systems Research, 236 (2024), p. 110651.

[22] M. C. Steinbach, *Tree-sparse convex programs*, Mathematical methods of operations research, 56 (2003), pp. 347–376.

[23] W. Wiesemann, D. Kuhn, and M. Sim, *Distributionally robust convex optimization*, Operations research, 62 (2014), pp. 1358–1376.

[24] G. Zhao, *A log-barrier method with benders decomposition for solving two-stage stochastic linear programs*, Mathematical Programming, 90 (2001), pp. 507–536.

[25] ———, *A lagrangian dual method with self-concordant barriers for multi-stage stochastic convex programming*, Mathematical Programming, 102 (2005), pp. 1–24.