Dynamic Resource Provisioning in Data Centers under Demand Uncertainty

Chaithanya Bandi * Omar El Housni † Vineet Goyal †

In this paper, we consider a dynamic resource provisioning problem that arises in data centers like Google and AWS who need to plan for resources under uncertain demand requirements. For instance, users requests arrive in an online manner with uncertain requirements of multiple resources including memory, computing power and bandwidth. The goal of the service provider (i.e. Google or AWS) is to satisfy these demand requirements with the smallest possible cost of hardware and power. It is usually significantly more costly to procure resources at short notice. Therefore, data centers aim to provision enough resources before the uncertain demands are revealed sequentially and make recourse decisions to procure additional resources in shortfall scenarios. This is referred to as the *dynamic resource provisioning* problem and raises many challenging questions.

Our Model: We consider a robust optimization approach for the dynamic resource provisioning problem. In particular, we consider a dynamic robust optimization problem under demand uncertainty and the goal is to find a feasible policy that satisfies the uncertain demand requirements and the worst-case cost is minimized. The uncertain demand in the right hand side is revealed sequentially where in each period t, an adversary selects the demand requirements from a given uncertainty set, $\mathcal{U}_t$. This problem is NP-hard even in two-stage (Feige et al. [5]). This motivates us to consider approximations for the problem. Static robust and affine adjustable solution approximations have been studied in the literature for this problem. In a static robust solution, we compute a single optimal solution that is feasible for all realizations of the uncertain right hand side. However, the performance of static solutions can be arbitrarily large for a general convex uncertainty set. Piecewise static policies also called K-adaptibility policies, are generalization of static policies where we divide the uncertainty set into several pieces and specify a static solution for each piece (El Housni and Goyal [4], Hanasusanto et al. [6]). In general, this approach with

*Kellog school of management, Northwestern University. Email: `c-bandi@kellogg.northwestern.edu`
†Industrial Engineering and Operations Research, Columbia University. Email: `{oe2148, vg2277}@columbia.edu`

a polynomial number of pieces does not improve significantly over the optimal static policy. Ben-Tal et al. [1] introduce an affine adjustable solution (also known as affine policy) to approximate adjustable robust problems. An affine policy restricts the decisions to be an affine function of the uncertain right-hand-side. An optimal affine policy can be computed efficiently for a large class of problems and has a strong empirical performance. Bertsimas et al. [3] and Iancu et al. [7] show that affine policies are optimal for a class of multistage problems where there is a single parameter uncertain in each period. Bertsimas and Goyal [2] show that affine policies are optimal if the uncertainty set is a simplex and give a worst case bound of $O(\sqrt{m})$ for general uncertainty sets where m is the dimension of the uncertainty set.

Our Contributions: Our main contributions are summarized below.

(a) We consider affine policies for dynamic robust optimization problem under right hand side uncertainty. These can be computed efficiently for a large class of uncertainty sets and these are optimal for the two-stage problem when the uncertainty set is a simplex (see [2]). However, we show that for the multi-stage problem, affine policies are not necessarily optimal even when the uncertainty sets are identical simplices. This is quite surprising and show that the structure of optimal policy for the multi-stage problem is significantly more complex than the two-stage counterpart.

(b) We consider the case of nested uncertainty sets, i.e. the uncertainty set at time period t is included in the uncertainty set at time period $t+1$. This is a reasonable assumption in multistage problems, which simply ensures that the day after tomorrow is more uncertain than tomorrow. For these class of uncertainty sets, we show that affine policies provide a good approximation for the dynamic robust optimization problem. In particular, the cost of the optimal affine policies is $O(\log T)$ times the optimal dynamic policy when the uncertainty set in each stage is a simplex. Furthermore, we show that our policy gives an $O(\sqrt{\log(T)m})$-approximation for general nested uncertainty sets where m is the dimension of the uncertainty sets. We would like to note that even for two-stage problem, the worst-case performance of affine policies can be $\Omega(\sqrt{m})$ times worse than the optimal two-stage solution [2]. The dependence of approximation bound on the number of periods as $\log T$ in this nested uncertainty case is quite analogous to the logarithmic regret bounds in dynamic learning problems.

(c) We generalize the performance bound of affine policy for two-stage problem in [2] and show an approximation bound of $O(\sqrt{mT})$ for general uncertainty sets when the assumption of nested uncertainty sets is relaxed.

(d) We apply our robust optimization approach using affine policy to the dynamic resource provisioning problem in Google data centers and compare our approach to the stochastic approach based on chance constraints optimization problems in terms of computational time and performance. We demonstrate that using our approaches we achieve near-optimal results, in a tractable manner and much faster time.

References

[1] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[2] D. Bertsimas and V. Goyal. On the Power and Limitations of Affine Policies in Two-Stage Adaptive Optimization. *Mathematical Programming*, 134(2):491–531, 2012.

[3] D. Bertsimas, D. Iancu, and P. Parrilo. Optimality of Affine Policies in Multi-stage Robust Optimization. *Mathematics of Operations Research*, 35:363–394, 2010.

[4] O. El Housni and V. Goyal. Piecewise static policies for two-stage adjustable robust linear optimization problems under uncertainty. 2015.

[5] U. Feige, K. Jain, M. Mahdian, and V. Mirrokni. Robust combinatorial optimization with exponential scenarios. *Lecture Notes in Computer Science*, 4513:439–453, 2007.

[6] G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. K-adaptability in two-stage robust binary programming, 2014.

[7] D. Iancu, M. Sharma, and M. Sviridenko. Supermodularity and affine policies in dynamic robust optimization. *Operations Research*, 61(4):941–956, 2013.