Activated Benders Decomposition for Paratransit Workforce Scheduling Under Cancellation Uncertainty

K. S. Cummings[a,*], A. Jacquillat[a,b], V. Vaze[c]

[a] Operations Research Center, MIT, Cambridge, MA, USA
[b] Sloan School of Management, MIT, Cambridge, MA, USA
[c] Thayer School of Engineering, Dartmouth College, Hanover, NH, USA
[*] Corresponding author

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

April 4, 2022

Keywords: workforce scheduling, paratransit, Benders decomposition, stochastic programming

1 INTRODUCTION

People with disabilities face significant obstacles to mobility, including inaccessibility, lack of assistive technology, inconvenient schedules, inadequate accommodations, and discrimination (World Health Organization, 2001). Toward provision of equitable services, transit agencies seek to offer high-quality *paratransit* options to populations with limited mobility, mandated under the U.S. Americans with Disabilities Act of 1990.

Paratransit systems entail complex, costly operations. As typical pickup-and-delivery systems, paratransit services face spatiotemporal imbalances, strict time windows, and traffic congestion. In addition, operators must drive accessible vehicles and accommodate passenger needs. Due to these complexities, paratransit contributes to 2–3% of transit ridership but 8–12% of transit costs (Meshram, 2018). In response, transit agencies are working to modernize paratransit by developing new partnerships with ride-hailing operators and launching digital platforms.

The paratransit operational regime defines a middle ground between fixed-route transit and on-demand ride-hailing. Providers have visibility into rider demand because of advance requests and a high proportion of recurring trips (e.g., commuting trips, trips to regular medical appointments). On the supply side, paratransit providers operate their own vehicle fleets with employee drivers. Yet, paratransit faces many uncertainties like trip cancellations, driver no-shows, and traffic congestion, requiring real-time routing adjustments. Paratransit therefore combines the planning challenges of fixed-route transit with the operational challenges of mobility on-demand.

Paratransit systems thus require dedicated analytics capabilities to provide low-cost, high-quality services to populations with limited mobility. Paratransit operations are a special case of the dial-a-ride (DAR) problem (Cordeau & Laporte, 2007). Bertsimas & Yan (2021) recently developed a cluster-then-route optimization approach to support vehicle routing for paratransit. In contrast, planning and design decisions in DAR systems have received more limited research.

In response, this work optimizes paratransit workforce schedules. We partner with an anonymous technology-based transportation provider that operates paratransit services in many US cities. Our stochastic program generates itineraries for drivers ahead of operations that respect trip cancellation uncertainty. Our collabor uses its digital platform to outsource trips to self-scheduled contractors on the day of operations, lending extra flexibility when optimizing employee drivers' shifts. From a technical standpoint, driver scheduling is complicated by the integration of two hard combinatorial optimization problems: workforce planning and downstream DAR operations. To tackle this challenge, we develop a new scalable approach involving an efficient formulation and a generalizable decomposition algorithm.

2 AN EFFICIENT TWO-STAGE STOCHASTIC PROGRAM

Our first contribution is to develop a formulation to optimize driver shifts in paratransit systems that are robust to trip cancellation uncertainty. Our formulation leverages the vehicle-sharing network from Vazifeh *et al.* (2018) to integrate DAR operations into a strategic workforce scheduling model. The formulation is a two-stage stochastic program with integer recourse, but the vehicle-sharing network induces a tight formulation—much tighter than canonical DAR models.

The model takes as inputs the number of available drivers W and a set of trip requests, each characterized by a pickup time and pickup/drop-off locations. We construct a vehicle-sharing network in which nodes are trip requests and directed arcs connect two requests that the same vehicle can serve. The first-stage problem determines driver itineraries, defined as ordered sets of trip requests. We generate an extensive set of candidate itineraries with the k-shortest paths algorithm from Yen (1971) over the vehicle-sharing network. Parameter C_i captures labor costs and binary trip delay penalties for each itinerary $i \in \mathcal{I}$. A set partitioning formulation selects itineraries $\boldsymbol{x} \in \{0,1\}^{\mathcal{I}}$, subject to cover constraints ensuring that all requests are served and budget constraints ensuring that at most W itineraries are selected.

After cancellations occur, the second-stage recourse problem outsources trips to contractors if necessary. We leverage the vehicle-sharing graph over the outsourced trips to construct physically valid itineraries for contractors. In each cancellation scenario $\omega \in \Omega$ occurring with probability p_ω, we define second-stage variables $\boldsymbol{y}_\omega$ that capture trip outsourcing. The second-stage problem minimizes outsourcing costs, subject to flow balance and consistency constraints. The full formulation's objective is represented in Equation (1), where $\mathcal{F}$ denotes the set of feasible first-stage decisions, and $\mathcal{G}_\omega(\boldsymbol{x})$ denotes the set of feasible second-stage decisions in scenario ω given $\boldsymbol{x} \in \mathcal{F}$.

$$\min_{\boldsymbol{x}\in\mathcal{F}} \quad \left(\sum_{i\in\mathcal{I}} C_i \cdot x_i + \sum_{w\in\Omega} p_\omega \cdot \min_{\boldsymbol{y}_\omega \in \mathcal{G}_\omega(\boldsymbol{x})} g(\boldsymbol{y}_\omega) \right) \tag{1}$$

For conciseness, we omit the full formulation. However, we emphasize that the tight second-stage formulation comes at the cost of exponentially many coupling constraints between first- and second-stage variables, reminiscent of linking constraints in facility location, network design, and production-routing formulations. The generalized structure is shown in Equation (2).

$$y_{\omega i} \leq x_i, \ \forall \omega \in \Omega, i \in \mathcal{I}_\omega \tag{2}$$

3 AN ACTIVATED BENDERS DECOMPOSITION ALGORITHM

The second contribution of the paper is a solution algorithm termed *activated Benders decomposition* (ABD). The algorithm iterates, until convergence, between a master problem and a restricted subproblem—involving activated constraints and activated variables. This approach is motivated by the observation that only a small subset of candidate itineraries is selected in the first stage; hence many second-stage constraints become moot and many second-stage decision variables can be fixed to zero (see Eequation (2)). In a Benders decomposition scheme, we can thus solve the *activated* subproblems at each iteration by considering only the variables and constraints corresponding to activated itineraries $\{i \in \mathcal{I} : x_i = 1\}$, as opposed to solving the full subproblems. One question remains: how to retrieve global optimality cuts for the full problem?

To tackle this challenge, we develop a primal-dual approach to reconstruct a solution to the full subproblem from the activated subproblem's solution. We prove that the proposed ABD scheme provides a valid optimality cut at each iteration and converges finitely to the optimal solution of the two-stage stochastic program with relaxed second-stage variables. By embedding ABD into an integer L-shaped decomposition inspired from Laporte & Louveaux (1993), we obtain a finitely convergent algorithm for the two-stage stochastic integer program. The proposed algorithm can also accommodate traditional acceleration strategies in Benders decomposition.

ABD is generalizable to any two-stage stochastic program with binary first-stage variables and second-stage activation constraints. Examples include production constraints in facility location, arc capacity constraints in network design, and delivery constraints in production-routing. ABD provides a computationally efficient decomposition approach that restricts the subproblem to constraints and variables relevant to the selected facilities, arcs, routes, and itineraries.

4 PRELIMINARY RESULTS

Our third contribution is to demonstrate the benefits of our modeling and algorithmic approach to optimize workforce planning in paratransit systems using real-world historical data.

From a computational standpoint, ABD yields high-quality solutions, consistently outperforming baseline algorithms. Table 1 reports the optimality gap, computational times, and expected costs obtained with ABD, Benders decomposition (BD), and off-the-shelf implementations of the stochastic integer program (SIP) and the model with continuous recourse (SCP). ABD converges to the SIP-optimal solution in every trial except one, which achieves a 0.01% optimality gap. Off-the-shelf implementation leads to sub-optimal solutions, especially with many scenarios. BD returns close-to-optimal solutions but fails to consistently converge.

From a practical standpoint, Figure 1 and Table 2 show that the optimized solution significantly reduces costs compared to historical shifts, which induce significant buffers as compared to trip demand. The optimized shifts are more aggressive, leading to a more balanced system in terms of demand and supply. The optimized solution utilizes more drivers through shorter shifts and does not serve all trip requests at their requested times, taking advantage of the time windows. Unlike the historical solution, the optimized solution critically leverages the availability of contractors to outsource requests as necessary and accommodates all trips with no delays, driving costs down and level of service up. In ongoing work, we are expanding our model to incorporate additional complexities, including ride-sharing, driver no-shows, and real-time itinerary adjustments.

References

Bertsimas, D., & Yan, J. 2021. The edge of optimization in large-scale vehicle routing for paratransit.

Cordeau, Jean-François, & Laporte, Gilbert. 2007. The dial-a-ride problem: models and algorithms. *Annals of Operations Research*, **153**(1), 29–46.

Laporte, Gilbert, & Louveaux, François V. 1993. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, **13**(3), 133–142.

Meshram, Aditi. 2018. *Best Practices in ADA Paratransit: Cost Reduction and Service Enhancement.* University of California, Davis.

Vazifeh, Moe, Santi, Paolo, Resta, Giovanni, Strogatz, Steven, & Ratti, Carlo. 2018. Addressing the minimum fleet problem in on-demand urban mobility. *Nature*, **557**.

World Health Organization. 2001. IFC: International Classification of Functioning, Disability and Health.

Yen, Jin. 1971. Finding the K Shortest Loopless Paths in a Network. *Management Science*, **17**(11).

Table 1 – *Computational results for each method by number of scenarios. "# solutions" and "convergence" respectively report number of instances where algorithm finds a solution and converges to provable optimality out of six instances (corresponding to six days of the week). Computational times are given in seconds, costs in percent optimality.*

Method	Scenarios	# solutions	Convergence	Avg. SCP Gap	Avg. SIP Gap	Min CPU	Max CPU	Avg CPU	Avg. Cost
ABD	1	6/6	6/6	0	0	2.86	160	34.4	100
	10	6/6	6/6	0	0	3.01	143	42.3	100
	25	6/6	6/6	0	0	2.91	361	79.2	100
	50	6/6	6/6	0	0	3.05	1884	349	100
	100	6/6	5/6	0.01%	0.01%	19.7	3600	638	100
BD	1	6/6	3/6	1.66%	1.66%	69.3	3600	1982	101
	10	6/6	2/6	2.39%	2.43%	1430	3600	3209	101
	25	6/6	1/6	2.72%	2.72%	2179	3600	3363	101
	50	6/6	0/6	3.85%	3.85%	3600	3600	3600	101
	100	6/6	0/6	45.3%	45.3%	3600	3600	3600	101
SCP	1	6/6	0/6	0.27%	0.27%	3600	3600	3600	100
	10	6/6	0/6	0.05%	0.06%	3600	3600	3600	105
	25	5/6	0/6	100%	100%	3600	3600	3600	120
	50	5/6	0/6	33.2%	33.2%	3600	3600	3600	108
	100	5/6	0/6	100%	100%	3600	3600	3600	115
SIP	1	6/6	1/6	-	19.7%	20.4	3600	3003	100
	10	6/6	0/6	-	0.28%	3600	3600	3600	100
	25	5/6	0/6	-	0.16%	3600	3600	3600	100
	50	5/6	1/6	-	24.9%	994	3600	3166	104
	100	3/6	0/6	-	0.35%	3600	3600	3600	129

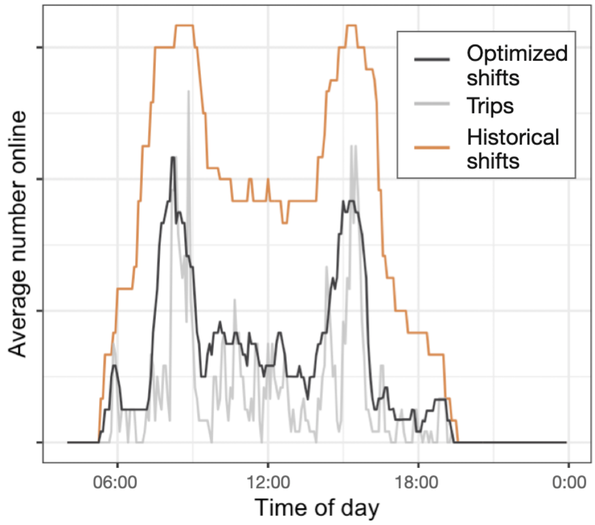

Figure 1 – *Average workforce vs. requests. Confidential y-axis scale.*

Table 2 – *Historical vs. optimized performance. Number of workers normalized for confidentiality.*

Solution	Avg. shift length (h)	Number of workers	% contractors	% trips delayed
Historical	5.39	100	0.00	19.7
Optimized	1.14	149	1.67	0.00