Learning dual inequalities for column generation

Guy Desaulniers
Polytechnique Montréal and GERAD, Canada
Co-authors: Louis Popovic, Quentin Cappart
In collaboration with Giro Inc.

Column Generation workshop
Montréal, Canada
May 19, 2023

POLYTECHNIQUE MONTRÉAL
TECHNOLOGICAL UNIVERSITY

GERAD
GROUPE D'ÉTUDES ET DE RECHERCHE EN ANALYSE DES DÉCISIONS

GIRO

Outline

Outline

High degeneracy in column generation

- High degeneracy occurs for many applications
  - With set partitioning/covering/packing constraints : vehicle routing, crew scheduling, cutting stock, computer vision, etc.
  - Several non-zero coefficients in each column (say, ≥ 10)
- Unstable dual variables : large oscillations from one iteration to the next
- Generation of useless columns
- Slow convergence (large number of iterations)

Some existing techniques

- Dual variable stabilization
  - Marsten et al. (1975), du Merle et al. (1999), Ben Amor et al. (2006), Oukil et al. (2007)
- Dual variable smoothing
  - Wentges (1997), Neame (2000), Pessoa et al. (2013, 2018)
- Constraint aggregation
  - El Hallaoui et al. (2005, 2008, 2010), Bouarab et al. (2017), Costa et al. (2022)
- Dual optimal inequalities
  - Valério de Carvalho (2005), Ben Amor et al. (2006), Gschwind and Irnich (2016), Yarkony et al. (2020), Haghani et al. (2022)

Goal

- Industrial project with Giro
- Application to public transit : electric bus scheduling
- Explore the use of dual inequalities to speed up column generation
- Use machine learning to predict them

Outline

Brief problem statement

Multi-depot electric vehicle scheduling problem (MDEVSP)

- Input
  - Set of timetabled trips with energy consumptions
  - Set of depots with identical electric buses (EBs)
  - Set of charging stations with identical chargers
- Find least-cost EB schedules such that
  - Each trip is covered by a bus
  - Each schedule starts and ends at the same depot
  - Each schedule is time- and energy-feasible
  - Charging station capacity is not exceeded

Mathematical model

$$\min \quad \sum_{d \in D} \sum_{s \in S^d} c_s x_s \tag{1}$$

$$\text{s.t. :} \quad \sum_{d \in D} \sum_{s \in S^d} a_s^t x_s = 1, \qquad \forall t \in T \tag{2}$$

$$\sum_{d \in D} \sum_{s \in S^d} b_s^{ph} x_s \leq u^h, \qquad \forall p \in P, h \in H \tag{3}$$

$$\sum_{s \in S^d} x_s \leq v^d, \qquad \forall d \in D \tag{4}$$

$$x_s \in \{0, 1\}, \qquad \forall d \in D, s \in S^d \tag{5}$$

$x_s = 1$ if EB schedule s is selected and 0 otherwise

Mathematical model

$$\min \quad \sum_{d\in D} \sum_{s\in S^d} c_s x_s \tag{1}$$

$$\text{s.t. :} \sum_{d\in D} \sum_{s\in S^d} a_s^t x_s = 1, \qquad \forall t \in T \tag{2}$$

$$\sum_{d\in D} \sum_{s\in S^d} b_s^{ph} x_s \leq u^h, \qquad \forall p \in P, h \in H \tag{3}$$

$$\sum_{s\in S^d} x_s \leq v^d, \qquad \forall d \in D \tag{4}$$

$$x_s \in \{0,1\}, \qquad \forall d \in D, s \in S^d \tag{5}$$

(1) : Minimize sum of fixed and variable costs

Mathematical model

$$\min \quad \sum_{d \in D} \sum_{s \in S^d} c_s x_s \tag{1}$$

$$\text{s.t. :} \sum_{d \in D} \sum_{s \in S^d} a_s^t x_s = 1, \qquad \forall t \in T \tag{2}$$

$$\sum_{d \in D} \sum_{s \in S^d} b_s^{ph} x_s \leq u^h, \qquad \forall p \in P, h \in H \tag{3}$$

$$\sum_{s \in S^d} x_s \leq v^d, \qquad \forall d \in D \tag{4}$$

$$x_s \in \{0, 1\}, \qquad \forall d \in D, s \in S^d \tag{5}$$

(2) : Assign an EB to each trip

Mathematical model

$$\min \quad \sum_{d\in D}\sum_{s\in S^d} c_s x_s \tag{1}$$

$$\text{s.t. :} \quad \sum_{d\in D}\sum_{s\in S^d} a_s^t x_s = 1, \qquad \forall t \in T \tag{2}$$

$$\sum_{d\in D}\sum_{s\in S^d} b_s^{ph} x_s \leq u^h, \qquad \forall p \in P, h \in H \tag{3}$$

$$\sum_{s\in S^d} x_s \leq v^d, \qquad \forall d \in D \tag{4}$$

$$x_s \in \{0,1\}, \qquad \forall d \in D, s \in S^d \tag{5}$$

(3) : Meet charging station capacities in each period

Mathematical model

$$\min \quad \sum_{d\in D}\sum_{s\in S^d} c_s x_s \tag{1}$$

$$\text{s.t. :} \quad \sum_{d\in D}\sum_{s\in S^d} a_s^t x_s = 1, \qquad \forall t \in T \tag{2}$$

$$\sum_{d\in D}\sum_{s\in S^d} b_s^{ph} x_s \leq u^h, \qquad \forall p \in P, h \in H \tag{3}$$

$$\sum_{s\in S^d} x_s \leq v^d, \qquad \forall d \in D \tag{4}$$

$$x_s \in \{0,1\}, \qquad \forall d \in D, s \in S^d \tag{5}$$

(4) : Respect available number of EBs per depot

Mathematical model

$$\min \quad \sum_{d \in D} \sum_{s \in S^d} c_s x_s \tag{1}$$

$$\text{s.t. :} \sum_{d \in D} \sum_{s \in S^d} a_s^t x_s = 1, \qquad \forall t \in T \tag{2}$$

$$\sum_{d \in D} \sum_{s \in S^d} b_s^{ph} x_s \leq u^h, \qquad \forall p \in P, h \in H \tag{3}$$

$$\sum_{s \in S^d} x_s \leq v^d, \qquad \forall d \in D \tag{4}$$

$$x_s \in \{0, 1\}, \qquad \forall d \in D, s \in S^d \tag{5}$$

π_t : Dual of (2) for trip t

Column generation

- Large MDESVP instances solved by a column generation heuristic
  - One pricing problem per depot : shortest path problem with resource constraints
  - Heuristic branching (diving)

- Constraint perturbation is applied to reduce degeneracy

$$\begin{aligned} \sum_{d \in D} \sum_{s \in S^d} a_s^t x_s + \delta_t^+ - \delta_t^- = 1, & \qquad \forall t \in T \\ 0 \leq \delta_t^+ \leq \epsilon_t^+, & \qquad \forall t \in T \\ 0 \leq \delta_t^- \leq \epsilon_t^-, & \qquad \forall t \in T, \end{aligned}$$

where ϵ_t^+ and ϵ_t^- are small constants chosen randomly.

Column generation

- Large MDESVP instances solved by a column generation heuristic
  - One pricing problem per depot : shortest path problem with resource constraints
  - Heuristic branching (diving)

- Constraint perturbation is applied to reduce degeneracy

$$\sum_{d \in D} \sum_{s \in S^d} a_s^t x_s + \delta_t^+ - \delta_t^- = 1, \qquad \forall t \in T$$

$$0 \leq \delta_t^+ \leq \epsilon_t^+, \qquad \forall t \in T$$

$$0 \leq \delta_t^- \leq \epsilon_t^-, \qquad \forall t \in T,$$

where ϵ_t^+ and ϵ_t^- are small constants chosen randomly.

Dual inequalities

Dual inequalities (DIs)

- Linear inequalities involving dual variables
- Reduce dual space
- Relax the master problem by adding new primal variables

Dual optimal inequalities (DOIs)

- DIs satisfied by all optimal dual solutions

Deep dual optimal inequalities (DDOIs)

- A set of DIs (not necessarily DOIs) satisfied by at least one optimal dual solution

Dual inequalities

Dual inequalities (DIs)

- Linear inequalities involving dual variables
- Reduce dual space
- Relax the master problem by adding new primal variables

Dual optimal inequalities (DOIs)

- DIs satisfied by all optimal dual solutions

Deep dual optimal inequalities (DDOIs)

- A set of DIs (not necessarily DOIs) satisfied by at least one optimal dual solution

Dual inequalities

Dual inequalities (DIs)

- Linear inequalities involving dual variables
- Reduce dual space
- Relax the master problem by adding new primal variables

Dual optimal inequalities (DOIs)

- DIs satisfied by all optimal dual solutions

Deep dual optimal inequalities (DDOIs)

- A set of DIs (not necessarily DOIs) satisfied by at least one optimal dual solution

Studied dual inequalities

- Focus on simple DIs of the form :

$$\pi_i \geq \pi_j, \qquad i, j \in T, i \neq j$$

- Should not be too difficult to predict
- Safe version :

$$\pi_i \geq \pi_j - \phi, \qquad i, j \in T, i \neq j$$

where $\phi > 0$ is a constant (that could depend on i, j)

- New primal variable : e_{ij}

Impact on the master problem

$$\min \quad \sum_{d \in D} \sum_{s \in S^d} c_s x_s + \phi e_{ij} \tag{6}$$

$$\text{s.t.}: \sum_{d \in D} \sum_{s \in S^d} a^i_s x_s - e_{ij} = 1 \tag{7}$$

$$\sum_{d \in D} \sum_{s \in S^d} a^j_s x_s + e_{ij} = 1 \tag{8}$$

$$e_{ij} \geq 0 \tag{9}$$

and other constraints

Variable e_{ij} allows to cover i twice (more than once) instead of covering j

Impact on the master problem

$$\min \quad \sum_{d \in D} \sum_{s \in S^d} c_s x_s + \phi e_{ij} \tag{6}$$

$$\text{s.t.}: \sum_{d \in D} \sum_{s \in S^d} a_s^i x_s - e_{ij} = 1 \tag{7}$$

$$\sum_{d \in D} \sum_{s \in S^d} a_s^j x_s + e_{ij} = 1 \tag{8}$$

$$e_{ij} \geq 0 \tag{9}$$

and other constraints

If $\pi_i^\star > \pi_j^\star$, then it should be more costly to cover i a second time rather than covering $j \Rightarrow e_{ij}^\star = 0$

Impact on the master problem

$$\min \quad \sum_{d \in D} \sum_{s \in S^d} c_s x_s + \phi e_{ij} \tag{6}$$

$$\text{s.t. :} \sum_{d \in D} \sum_{s \in S^d} a^i_s x_s - e_{ij} = 1 \tag{7}$$

$$\sum_{d \in D} \sum_{s \in S^d} a^j_s x_s + e_{ij} = 1 \tag{8}$$

$$e_{ij} \geq 0 \tag{9}$$

and other constraints

If $\pi^\star_i = \pi^\star_j$, then both options are equally costly and $\phi > 0$ breaks the tie $\Rightarrow e^\star_{ij} = 0$

Impact on the master problem

$$\min \quad \sum_{d\in D}\sum_{s\in S^d} c_s x_s + \phi e_{ij} \tag{6}$$

$$\text{s.t. :} \quad \sum_{d\in D}\sum_{s\in S^d} a_s^i x_s - e_{ij} = 1 \tag{7}$$

$$\sum_{d\in D}\sum_{s\in S^d} a_s^j x_s + e_{ij} = 1 \tag{8}$$

$$e_{ij} \geq 0 \tag{9}$$

and other constraints

If $\pi_i^\star < \pi_j^\star$ and $\phi < \pi_j^\star - \pi_i^\star$ (invalid DI), then $e_{ij}^\star > 0$

Impact on the master problem

$$\min \quad \sum_{d \in D} \sum_{s \in S^d} c_s x_s + \phi e_{ij} \tag{6}$$

$$\text{s.t. :} \quad \sum_{d \in D} \sum_{s \in S^d} a_s^i x_s - e_{ij} = 1 \tag{7}$$

$$\sum_{d \in D} \sum_{s \in S^d} a_s^j x_s + e_{ij} = 1 \tag{8}$$

$$e_{ij} \geq 0 \tag{9}$$

and other constraints

Multiple DIs can be imposed simultaneously, requiring multiple primal variables e_{ij}

Computational results with perfect information

- MDEVSP instance generator of Brasseur (2022)
- From real bus lines in Montreal
- 2 sets of 25 instances : with ∼840 (M) and ∼1030 (L) trips
- DIs deduced from optimal dual solutions ($\phi = 0$)
- Linear relaxation only

			MP	MP^{DI}	
	Trips	Buses	Time (s)	Time (s)	Acc
M	837.4	52.1	594	103	5.33
L	1030.5	62.6	1260	175	3.34

Acc = geometric mean of Time MP / Time MP^{DI}

Outline

Prediction

- Given a pair of trips i and j, we would like to predict if $\pi_i \geq \pi_j$ in an optimal linear relaxation solution of (1)–(5)
- We can solve MDEVSP instances and get optimal dual solutions $\Rightarrow$ Supervised learning
- In the training phase, features provide information about each trip and 0-1 labels are assigned to each ordered pair (1 if $\pi_i \geq \pi_j$ and 0 otherwise)

Machine learning model

- Graphical neural network (GNN) with attention, followed by a multilayer perceptron (MLP)
  - GNN takes into account the network structure between the trips
  - GNN propagates the features of the nodes and arcs through neighbor nodes and arcs
  - Given the updated features, MLP outputs a score $y_{ij} \in [0, 1]$ that can be used to make a prediction
  - $\pi_i \geq \pi_j$ if $y_{ij} \geq 1 - \beta$ or $y_{ji} \leq \beta$, a given parameter in $(0, 1]$
- Trip node features : Trip start and end times, trip duration, # trips departing from (resp. arriving at) the same terminal within 10 minutes after departure (resp. before arrival), etc.
- Arc features : Cost, consumed energy, travel time, waiting time, etc.

Machine learning model

- Graphical neural network (GNN) with attention, followed by a multilayer perceptron (MLP)
  - GNN takes into account the network structure between the trips
  - GNN propagates the features of the nodes and arcs through neighbor nodes and arcs
  - Given the updated features, MLP outputs a score $y_{ij} \in [0, 1]$ that can be used to make a prediction
  - $\pi_i \geq \pi_j$ if $y_{ij} \geq 1 - \beta$ or $y_{ji} \leq \beta$, a given parameter in $(0, 1]$
- Trip node features : Trip start and end times, trip duration, # trips departing from (resp. arriving at) the same terminal within 10 minutes after departure (resp. before arrival), etc.
- Arc features : Cost, consumed energy, travel time, waiting time, etc.

Prediction results

- 250/250 instances M/L
- 80%/10%/10% for training/validation/testing

β	Precision (%) M	Precision (%) L	DIs M	DIs L
0.15	94.7	95.0	275264	427927
0.25	92.0	92.4	314679	488256
0.35	89.2	89.6	348041	539769
0.45	86.2	86.7	379695	586923

Precision = True positives / (True positives + False positives)

Inequality selection

- Number of potential DIs is very large $\Rightarrow$ Select a subset
- Select inequalities forming a long sequence :

$$\pi_1 \geq \pi_2 \geq \ldots \geq \pi_\ell$$

- Network G representing the DIs
  - One source o, one sink s, and one node per trip $t \in T$
  - Arcs (o, t), (t, s), $\forall t \in T$; $(i, j) \in T \times T$ if $\pi_i \geq \pi_j$ predicted
  - Filtering of arcs applied to limit the number of cycles
- Run Bellman-Ford algorithm on G to find the longest sequence of arcs for a given number of iterations
  - If a cycle is detected, remove all arcs in the cycle
  - Otherwise, pick the DIs in the path and remove the corresponding arcs

Inequality selection

- Number of potential DIs is very large $\Rightarrow$ Select a subset
- Select inequalities forming a long sequence :

$$\pi_1 \geq \pi_2 \geq \ldots \geq \pi_\ell$$

- Network G representing the DIs
  - One source o, one sink s, and one node per trip $t \in T$
  - Arcs (o, t), (t, s), $\forall t \in T$; $(i, j) \in T \times T$ if $\pi_i \geq \pi_j$ predicted
  - Filtering of arcs applied to limit the number of cycles
- Run Bellman-Ford algorithm on G to find the longest sequence of arcs for a given number of iterations
  - If a cycle is detected, remove all arcs in the cycle
  - Otherwise, pick the DIs in the path and remove the corresponding arcs

Recovery procedure

- Selected DIs might not be DDOIs due to wrong predictions
- Might yield a weaker lower bound (some $e^{\star}_{ij} > 0$)
- Implemented recovery procedure
  - For all pairs (i, j) such that $e^{\star}_{ij} > 0$, increase ϕ_{ij} by 10 if first time or remove the corresponding DI otherwise
  - Restart column generation
  - Stop if $e^{\star}_{ij} = 0$ for all pairs (i, j) or after a fixed number of iterations

Outline

Experiments with predicted DIs

- 2 new 25-instance sets : with $\sim$870 (M) and $\sim$1080 (L) trips
- Predicted DIs ($\beta = 0.15$) : 25 sequences of DIs
  - Average of 638.9/720.7 DIs for instances M/L
  - Precision of 88.7% / 86.8%

- 3 algorithm variants
  1. $\phi = 0$, no recovery
  2. $\phi = 0$, up to two recovery iterations
  3. $\phi = 1$, up to one recovery iteration

- M/L instances split into slow (S) and fast (F) instances according to median time of algorithm without DIs
- Linear relaxation + recovery if any

Results

	Time (s)	Acc	Acc S	Acc F	LB Gap (%)	Miscov (%)
Without DIs						
M	2274					
L	7676					
With DIs, $\phi = 0$, no recovery						
M	499	1.71	2.67	1.06	0.19	6.01
L	3372	1.99	3.93	0.95	0.17	5.73
With DIs, $\phi = 0$, up to 2 rec. it.						
M	1558	0.72	1.12	0.45	0.11	3.54
L	7323	0.80	1.19	0.53	0.09	3.60
With DIs, $\phi = 1$, up to 1 rec. it.						
M	669	1.27	1.83	0.86	0.12	1.81
L	6878	1.28	2.04	0.83	0.10	1.11

Miscov : # trips not covered exactly once

Conclusion

Summary

- Prediction of simple DIs to speed up column generation
- Promising computational results for the most difficult MDEVSP instances

Future works

- Improve ML model for better prediction
- Improve DI selection
- Generate DIs dynamically (see Gschwind and Irnich, 2016)

Questions ?

Conclusion

Summary

- Prediction of simple DIs to speed up column generation
- Promising computational results for the most difficult MDEVSP instances

Future works

- Improve ML model for better prediction
- Improve DI selection
- Generate DIs dynamically (see Gschwind and Irnich, 2016)

Questions ?

Conclusion

Summary

- Prediction of simple DIs to speed up column generation
- Promising computational results for the most difficult MDEVSP instances

Future works

- Improve ML model for better prediction
- Improve DI selection
- Generate DIs dynamically (see Gschwind and Irnich, 2016)

Questions ?