Column Generation and its applications

Murat Firat, dept. IE&IS, TU/e
BPI Cluster meeting
March 14, 2018

TU/e Technische Universiteit Eindhoven University of Technology

Outline

- Some real-life decision problems
  - Standard formulations
- Basics of Column Generation
  - Master formulations
- Case: Shift scheduling of airport workers
  - Problem description
  - Master formulation
  - Reduced cost and pricing problem
- Column Generation overview
- Towards an integer solution
- Using data science in the CG approach

Outline

Some real-life decision problems
- Standard formulations

Basics of Column Generation
- Master formulations

Case: Shift scheduling of airport workers
- Problem description
- Master formulation
- Reduced cost and pricing problem

Column Generation overview

Towards an integer solution

Using data science in the CG approach

Different industry sectors, different problems

- Telecommunication: Stable workforce assignments.

Different industry sectors, different problems

- Telecommunication: Stable workforce assignments.
- Telecommunication: Designing FTTH network.

Different industry sectors, different problems

- Telecommunication: Stable workforce assignments.
- Telecommunication: Designing FTTH network.
- Airport ground operations: Shift scheduling of workers.

Different industry sectors, different problems

- Telecommunication: Stable workforce assignments.
- Telecommunication: Designing FTTH network.
- Airport ground operations: Shift scheduling of workers.
- Logistics: Planning routes of vehicles.

Different industry sectors, different problems

- Telecommunication: Stable workforce assignments.
- Telecommunication: Designing FTTH network.
- Airport ground operations: Shift scheduling of workers.
- Logistics: Planning routes of vehicles.
- Machine learning: Constructing max-accuracy decision trees.

Instance sizes

- Stable workforce assignments: 100 workers, 25 jobs, 15 skills.

Instance sizes

- Stable workforce assignments: 100 workers, 25 jobs, 15 skills.
- Designing FTTH network: Local areas of 5-10K citizens,

Instance sizes

- Stable workforce assignments: 100 workers, 25 jobs, 15 skills.
- Designing FTTH network: Local areas of 5-10K citizens,
- Shift scheduling of airport workers: 200 workers, 4 weeks, 5 skills.

Instance sizes

- Stable workforce assignments: 100 workers, 25 jobs, 15 skills.
- Designing FTTH network: Local areas of 5-10K citizens,
- Shift scheduling of airport workers: 200 workers, 4 weeks, 5 skills.
- Planning routes of vehicles: 100-1K customers, 50 vehicles.

Instance sizes

- Stable workforce assignments: 100 workers, 25 jobs, 15 skills.
- Designing FTTH network: Local areas of 5-10K citizens,
- Shift scheduling of airport workers: 200 workers, 4 weeks, 5 skills.
- Planning routes of vehicles: 100-1K customers, 50 vehicles.
- Decision trees: 5K data instances, 50 attributes.

Common properties of the problems

- Strict feasibility requirements.

Common properties of the problems

- Strict feasibility requirements.
- Worst cases: NP-Hard.

Common properties of the problems

- Strict feasibility requirements.
- Worst cases: NP-Hard.
- Exponentially many feasible solutions

Common properties of the problems

- Strict feasibility requirements.
- Worst cases: NP-Hard.
- Exponentially many feasible solutions
- Many local optimal points.

Standard formulations

One big formulation, low-quality bound.

- Stable workforce assignments: Worker-job decisions.

Standard formulations

One big formulation, low-quality bound.

- Stable workforce assignments: Worker-job decisions.
- Shift scheduling of airport workers: Worker-shift decisions.

Standard formulations

One big formulation, low-quality bound.

- Stable workforce assignments: Worker-job decisions.
- Shift scheduling of airport workers: Worker-shift decisions.
- Planning routes of vehicles: Vehicle-customer decisions.

Standard formulations

One big formulation, low-quality bound.

- Stable workforce assignments: Worker-job decisions.
- Shift scheduling of airport workers: Worker-shift decisions.
- Planning routes of vehicles: Vehicle-customer decisions.
- Decision trees: Data row-tree node decisions.

Outline

Master formulations

- Stable workforce assignments: Team-job decisions.

Master formulations

- Stable workforce assignments: Team-job decisions.
- Shift scheduling of airport workers: Worker-schedule decisions.

Master formulations

- Stable workforce assignments: Team-job decisions.
- Shift scheduling of airport workers: Worker-schedule decisions.
- Planning routes of vehicles: Vehicle-route decisions.

Master formulations

- Stable workforce assignments: Team-job decisions.
- Shift scheduling of airport workers: Worker-schedule decisions.
- Planning routes of vehicles: Vehicle-route decisions.
- Decision trees: Data segment-tree node decisions.

Master formulation

- Linear programming model:

Master formulation

- Linear programming model:
  - More complicated objects as decision variables.

Master formulation

- Linear programming model:
  - More complicated objects as decision variables.
- Object set

Master formulation

- Linear programming model:
  - More complicated objects as decision variables.
- Object set
  - has exponentially many items.

Master formulation

- Linear programming model:
  - More complicated objects as decision variables.
- Object set
  - has exponentially many items.
  - initially is empty or has few items.

Master formulation

- Linear programming model:
  - More complicated objects as decision variables.
- Object set
  - has exponentially many items.
  - initially is empty or has few items.
- Iteratively find promising columns and add them to object set.

Master formulation

- Linear programming model:
  - More complicated objects as decision variables.
- Object set
  - has exponentially many items.
  - initially is empty or has few items.
- Iteratively find promising columns and add them to object set.
- When no promising column exists: certificate for optimality.

Main loop of Column Generation

- Reduced cost of a column is

Main loop of Column Generation

- Reduced cost of a column is
  - violation amount of the corresponding dual constraint.

Main loop of Column Generation

- Reduced cost of a column is
  - violation amount of the corresponding dual constraint.
  - estimated cost change per unit increase in its value.

Main loop of Column Generation

- Reduced cost of a column is
  - violation amount of the corresponding dual constraint.
  - estimated cost change per unit increase in its value.
- Adding promising columns

Main loop of Column Generation

- Reduced cost of a column is
  - violation amount of the corresponding dual constraint.
  - estimated cost change per unit increase in its value.
- Adding promising columns
  - completes dual feasibility.

Main loop of Column Generation

- Reduced cost of a column is
  - violation amount of the corresponding dual constraint.
  - estimated cost change per unit increase in its value.
- Adding promising columns
  - completes dual feasibility.
  - improves primal objective.

Main loop of Column Generation

- Reduced cost of a column is
  - violation amount of the corresponding dual constraint.
  - estimated cost change per unit increase in its value.
- Adding promising columns
  - completes dual feasibility.
  - improves primal objective.
- No dual feasibility improvement $\approx$ primal optimality

Outline

Problem definition: Shift scheduling

We are given:

- multi-skilled workers with availability info.

Problem definition: Shift scheduling

We are given:

- multi-skilled workers with availability info.
- service demand within a planning horizon

Problem definition: Shift scheduling

We are given:

- multi-skilled workers with availability info.
- service demand within a planning horizon
- labor regulations about shifts:

Problem definition: Shift scheduling

We are given:

- multi-skilled workers with availability info.
- service demand within a planning horizon
- labor regulations about shifts:
  - minimum resting time between shifts,

Technische Universiteit Eindhoven University of Technology

Problem definition: Shift scheduling

We are given:

- multi-skilled workers with availability info.
- service demand within a planning horizon
- labor regulations about shifts:
  - minimum resting time between shifts,
  - maximum working time due to contracts,

Problem definition: Shift scheduling

We are given:

- multi-skilled workers with availability info.
- service demand within a planning horizon
- labor regulations about shifts:
  - minimum resting time between shifts,
  - maximum working time due to contracts,
  - night shifts: longer resting times.

Basic notation

- Let $\mathcal{S}$ be set of all schedules st. a schedule $s \in \mathcal{S}$

Basic notation

- Let $\mathcal{S}$ be set of all schedules st. a schedule $s \in \mathcal{S}$
  - covers time interval $i \in N$ if $s^i = 1$, not otherwise.

Basic notation

- Let $\mathcal{S}$ be set of all schedules st. a schedule $s \in \mathcal{S}$
  - covers time interval $i \in N$ if $s^i = 1$, not otherwise.
  - should comply all regulations

Basic notation

- Let $\mathcal{S}$ be set of all schedules st. a schedule $s \in \mathcal{S}$
  - covers time interval $i \in N$ if $s^i = 1$, not otherwise.
  - should comply all regulations
  - should have length bet. 3-9 hrs.

Basic notation

- Let $\mathcal{S}$ be set of all schedules st. a schedule $s \in \mathcal{S}$
  - covers time interval $i \in N$ if $s^i = 1$, not otherwise.
  - should comply all regulations
  - should have length bet. 3-9 hrs.
- For worker $w \in W$

Basic notation

- Let $\mathcal{S}$ be set of all schedules st. a schedule $s \in \mathcal{S}$
  - covers time interval $i \in N$ if $s^i = 1$, not otherwise.
  - should comply all regulations
  - should have length bet. 3-9 hrs.
- For worker $w \in W$
  - Sk_w^d indicates if worker w is skilled in skill type $d \in D$.

Technische Universiteit Eindhoven University of Technology

Basic notation

- Let $\mathcal{S}$ be set of all schedules st. a schedule $s \in \mathcal{S}$
  - covers time interval $i \in N$ if $s^i = 1$, not otherwise.
  - should comply all regulations
  - should have length bet. 3-9 hrs.
- For worker $w \in W$
  - Sk_w^d indicates if worker w is skilled in skill type $d \in D$.
- Service demand $R_{i,d}$ at time i in skill type d,

Master IP formulation

$$
\begin{aligned}
\text{Min} \quad & \sum_{w \in W} \sum_{s \in \mathcal{S}} c_{sw} x_{sw} \\
\text{subject to} \quad & \\
& \sum_{w \in W} Sk_w^d s^i x_{sw} \geq R_{i,d}, \quad d \in D, i \in N \\
& \sum_{s \in \mathcal{S}} x_{sw} = 1, \quad w \in W \\
& x_{sw} \in \{0,1\}, \; s \in \mathcal{S}, w \in W
\end{aligned}
$$

LP Relaxation of Master formulation

$$
\begin{aligned}
\text{Min} \quad & \sum_{w \in W} \sum_{s \in \mathcal{S}'} c_{sw} x_{sw} \\
\text{subject to} \quad & \\
& \sum_{w \in W} Sk_w^d s^i x_{sw} \geq R_{i,d}, \quad d \in D, i \in N \\
& \sum_{s \in \mathcal{S}'} x_{sw} = 1, \quad w \in W \\
& 0 \leq x_{sw} \leq 1, \quad s \in \mathcal{S}', w \in W
\end{aligned}
$$

Note: Restricted set $\mathcal{S}' \subset \mathcal{S}$

LP Relaxation of Master formulation

$$
\begin{array}{llll}
\text{Min} & \displaystyle\sum_{w\in W}\sum_{s\in\mathcal{S}'} c_{sw}x_{sw} & & \text{Duals} \\
\text{subject to} & & & \\
& \displaystyle\sum_{w\in W} Sk_w^d s^i x_{sw} \geq R_{i,d}, & d\in D, i\in N & \pi_{i,d} \\
& \displaystyle\sum_{s\in\mathcal{S}'} x_{sw} = 1, & w\in W & \theta_w \\
& 0 \leq x_{sw} \leq 1, & s\in\mathcal{S}', w\in W &
\end{array}
$$

Note: Restricted set $\mathcal{S}' \subset \mathcal{S}$

Reduced cost of column x_{sw}

Dual constraint of column x_{sw}:

$$\sum_{d \in D} \sum_{i \in N} Sk_w^d s^i \pi_{i,d}^* + \theta_w^* \leq c_{sw} \tag{1}$$

Reduced cost of column x_{sw}:

$$\bar{c}_{sw} = c_{sw} - \sum_{d \in D} \sum_{i \in N} Sk_w^d s^i \pi_{i,d}^* - \theta_w^* \tag{2}$$

Case $\bar{c}_{sw} < 0$: (1) Estimated objective decrease (why?), (2) dual feasibility violation.

Pricing problem: Objective

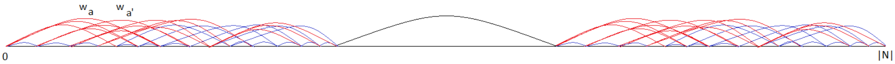
Pricing problem: Find the most promising column (schedule) with the objective

$$\min_{s\in\mathcal{S},w\in W}\left\{c_{sw}-\sum_{d\in D}\sum_{i\in N}Sk_w^d s^i \pi_{i,d}^* - \theta_w^*\right\} \tag{3}$$

$$= \max_{s\in\mathcal{S},w\in W}\left\{\sum_{i\in N}s^i\left(\sum_{d\in D}Sk_w^d \pi_{i,d}^* - c_{iw}\right)\right\} - \theta_w^* \tag{4}$$

Pricing problem: Modeling

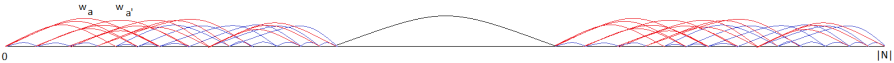
Define a graph,



- reds: shifts, blues: resting times, blacks: unavailability.

Pricing problem: Modeling

Define a graph,

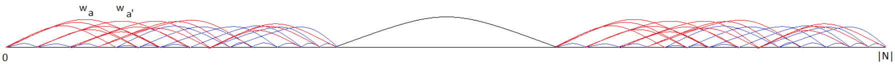


- reds: shifts, blues: resting times, blacks: unavailability.
- labor regulations: arc structure and side constraints

Technische Universiteit
Eindhoven
University of Technology

Pricing problem: Modeling

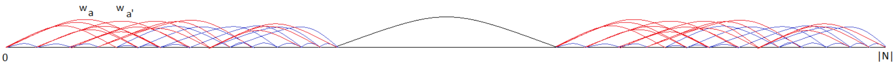
Define a graph,



- reds: shifts, blues: resting times, blacks: unavailability.
- labor regulations: arc structure and side constraints
- Solve $\max_a w_a x_a$ subject to

Pricing problem: Modeling

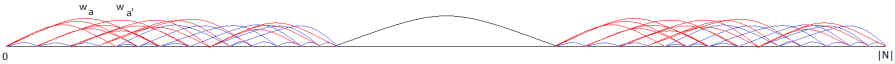
Define a graph,



- reds: shifts, blues: resting times, blacks: unavailability.
- labor regulations: arc structure and side constraints
- Solve $\max_a w_a x_a$ subject to
  - conservations, resting after shifts, passing unavailability arcs.

Pricing problem: Modeling

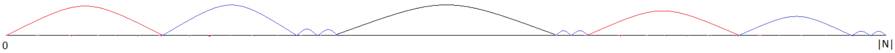
Define a graph,



- reds: shifts, blues: resting times, blacks: unavailability.
- labor regulations: arc structure and side constraints
- Solve $\max_a w_a x_a$ subject to
    - conservations, resting after shifts, passing unavailability arcs.
- Pricing: find the constrained $0 - |N|$ "Longest Path"!

Pricing problem: Modeling

A schedule s on the graph looks like:



Column Generation overview

- Initialization:

Column Generation overview

- Initialization:
  - Formulate Master ILP model, obtain the Restricted Master Problem (RMP).
  - Express the reduced costs, formulate pricing.

Column Generation overview

- Initialization:
  - Formulate Master ILP model, obtain the Restricted Master Problem (RMP).
  - Express the reduced costs, formulate pricing.
- Warm up: Find several initial columns for a warm start of RMP.

Column Generation overview

- Initialization:
  - Formulate Master ILP model, obtain the Restricted Master Problem (RMP).
  - Express the reduced costs, formulate pricing.
- Warm up: Find several initial columns for a warm start of RMP.
- Step 1: Solve the Restricted Master Problem, pass duals to pricing.

Column Generation overview

- Initialization:
  - Formulate Master ILP model, obtain the Restricted Master Problem (RMP).
  - Express the reduced costs, formulate pricing.
- Warm up: Find several initial columns for a warm start of RMP.
- Step 1: Solve the Restricted Master Problem, pass duals to pricing.
- Step 2: Solve pricing:

Column Generation overview

- Initialization:
  - Formulate Master ILP model, obtain the Restricted Master Problem (RMP).
  - Express the reduced costs, formulate pricing.
- Warm up: Find several initial columns for a warm start of RMP.
- Step 1: Solve the Restricted Master Problem, pass duals to pricing.
- Step 2: Solve pricing:
  - If $\exists i : \bar{c}_i < 0$: update $\mathcal{S}'$, go to Step 1.
  - If $\forall i : \bar{c}_i \geq 0$: RMP is solved to optimality, go to Step 3.

Column Generation overview

- Initialization:
  - Formulate Master ILP model, obtain the Restricted Master Problem (RMP).
  - Express the reduced costs, formulate pricing.
- Warm up: Find several initial columns for a warm start of RMP.
- Step 1: Solve the Restricted Master Problem, pass duals to pricing.
- Step 2: Solve pricing:
  - If $\exists i : \bar{c}_i < 0$: update $\mathcal{S}'$, go to Step 1.
  - If $\forall i : \bar{c}_i \geq 0$: RMP is solved to optimality, go to Step 3.
- Step 3: Output the RMP solution.

Outline

Towards integer solutions

Having fractional RMP optimal solution, we have two choices:

- Use rounding heuristics:

Towards integer solutions

Having fractional RMP optimal solution, we have two choices:

- Use rounding heuristics:
  - Use meta heuristics to find a feasible solution quickly (hopefully)

Towards integer solutions

Having fractional RMP optimal solution, we have two choices:

- Use rounding heuristics:
  - Use meta heuristics to find a feasible solution quickly (hopefully)
  - Make decisions how to (smartly) round the fractional solution

Towards integer solutions

Having fractional RMP optimal solution, we have two choices:

- Use rounding heuristics:
  - Use meta heuristics to find a feasible solution quickly (hopefully)
  - Make decisions how to (smartly) round the fractional solution
- Start a smart enumeration, e.g. Branch-and-Price, either

.

Towards integer solutions

Having fractional RMP optimal solution, we have two choices:

- Use rounding heuristics:
  - Use meta heuristics to find a feasible solution quickly (hopefully)
  - Make decisions how to (smartly) round the fractional solution
- Start a smart enumeration, e.g. Branch-and-Price, either
  - to obtain "optimal" integer sol'n.
  - to output "best-found" integer sol'n with quality measure in a time limit.

.

Outline

Using data science: Advantage or waste?

Consider a data of instance history and their solutions.

- Find "similar" instances and round the current fractional solution towards corresponding solutions.

Using data science: Advantage or waste?

Consider a data of instance history and their solutions.

- Find "similar" instances and round the current fractional solution towards corresponding solutions.
- Simply find "similar" previous solutions to the current fractional solution and round accordingly.

Using data science: Advantage or waste?

Consider a data of instance history and their solutions.

- Find "similar" instances and round the current fractional solution towards corresponding solutions.
- Simply find "similar" previous solutions to the current fractional solution and round accordingly.
- Cluster instances and analyze the commonalities (patterns) in the cluster solutions.

THANKS