Faculty of Mathematics and Physics
Charles University

Marek Plesník

Katedra pravděpodobnosti a matematické statistiky

Optimal Lines for Railway Systems

Econometrics Seminar 1

13 April 2026

Table of Contents

1. Motivation
2. Network Modeling and Formulation
3. The Branch & Bound Algorithm
4. Model Reduction and Cutting Planes
5. Results and Bounds
6. Conclusion

Motivation

- Shift toward periodic timetables
  - Widely adopted across public transit networks
  - Vehicles operate at fixed, repeating intervals
- Tactical line planning
  - Must design a feasible line system before exact timetables and schedules are set
  - Sufficient capacity to serve a known, fixed passenger demand
- Balancing efficiency and service quality
  - Railway planning must balance operational costs with passenger convenience
  - Since exact timetable information is not yet available, the number of direct, transfer-free journeys is a natural proxy for service quality
- A difficult optimization problem
  - Large-scale problem and NP-hard

Literature

Main article:

- Bussieck, M. R., Kreuzer, P., Zimmermann, U. T. (1996). Optimal lines for railway systems. *European Journal of Operational Research*, 96(1), 54–63.

Selected references:

- Dienst, H. (1978). Linienplanung im spurgeführten Personenverkehr mit Hilfe eines heuristischen Verfahrens. *Ph.D. Thesis, TU Braunschweig.*
- Magnanti, T. L., Wong, R. T. (1984). Network design and transportation planning: Models and algorithms. *Transportation Science*, 18(1), 1–55.
- Wu, J., Shan, X., Zhao, S. (2025). Optimization of Passenger Train Line Planning Adjustments Based on Minimizing Systematic Costs. *Inventions*, 10(4), 64.

Network Modeling I

Suppose $G = (V, E)$ is an undirected graph where

- V ... set of vertices describing the stations
- E ... set of edges defining direct connections (links) between two stations

Edge evaluations:

- $T : E \to \mathbb{Z}^+$ ... travel time on a single link
- $D : E \to \mathbb{Z}^+$ ... travel distance on a single link

Possible lines:

- Modeled as simple paths in G
- $\mathcal{CY} \subseteq V$ ... classification yards (e.g., equipped to compose trains)
- $\mathcal{L}_0$ ... set of all possible lines (paths with start and endpoint in $\mathcal{CY}$)
- $f : \mathcal{L}_0 \to \mathbb{Z}^+$, $f_l := f(l)$ ... frequencies of possible lines in a fixed time interval

Traffic volume:

- $tr : \{\{a,b\} : a,b \in V,\ a \neq b\} \to \mathbb{Z}^+$ ... volume of traffic between stations
- $\mathcal{T} := \{\{a,b\} \mid a,b \in V, a \neq b,\ tr(\{a,b\}) > 0\}$

Notation:

We write $tr(t)$ for $t = \{a,b\} \in \mathcal{T}$

Assumption.

Travellers between a and b, for some $a,b \in V$, use a shortest path between a and b in G with respect to some edge evaluation.

In our context:

- Realistic assumption for passenger behavior in long-distance networks
- Assuming all shortest paths P_t are uniquely determined, the traffic load on each link is well-defined

Network Modeling III

Let C denote the maximal fixed train capacity

Traffic load & number of direct travellers:

- $tl : E \to \mathbb{Z}^+, \quad tl(e) = \sum_{\{t \in \mathcal{T},\, e \in P_t\}} tr(t)$ ... aggregated passenger volume on link e
- $d_{t,l} \in \mathbb{Z}^+$ ... the number of direct travellers between $t \in \mathcal{T}$ using line $l \in \mathcal{L}$

Line-frequency requirement:

$lfr : E \to \mathbb{Z}^+, \quad lfr(e) = \left\lceil \frac{tl(e)}{C} \right\rceil$ ... minimum number of trains required on link e

Reduced candidate line set:

$\mathcal{L} := \{l \in \mathcal{L}_0 : l$ is the shortest path between some $a, b \in \mathcal{CY}\}$

Definition.

A line system is **feasible** if the assigned frequencies satisfy the *lfr* for every edge in the network.

Line Optimization Problem (LOP)

The goal is to maximize the total number of direct travellers across the network

$$
\begin{aligned}
\max \quad & \sum_{l \in \mathcal{L}} \sum_{t \in \mathcal{T}, P_t \subseteq l} d_{t,l} && \\
\text{s.t.} \quad & \sum_{l \in \mathcal{L}, P_t \subseteq l} d_{t,l} \leq tr(t) && \forall t \in \mathcal{T} \\
& \sum_{t \in \mathcal{T}, e \in P_t \subseteq l} d_{t,l} \leq C \cdot f_l && \forall e \in E,\ l \in \mathcal{L} \\
& \sum_{l \in \mathcal{L}, e \in l} f_l = lfr(e) && \forall e \in E \\
& d_{t,l}, f_l \in \mathbb{Z}^+ && \forall t \in \mathcal{T},\ l \in \mathcal{L}
\end{aligned}
$$

Remarks:

Optional weights $w_{t,l}$ (e.g., travel distance or time), fractional travellers $d_{t,l} \geq 0$, ...

Branch & Bound Algorithm

A standard tree-search method for solving mixed-integer problems

Main steps:

1. **Relaxation:** Solve the continuous problem to find a theoretical upper bound
2. **Branching:** If a solution is fractional, split into integer-restricted subproblems
3. **Bounding:** Discard branches worse than the best known integer solution

In our context:

- Line frequencies (f_l) must be integers
- Systematically finds the optimum without exhaustively testing every combination

Dienst et al. propose a simplified B & B method for railway line planning

Main simplification: Infinite train capacity (i.e., $C = \infty$)

- If a direct connection exists, all corresponding travellers can use it
- Capacity limits on individual lines are ignored
- In our model, this is approximated by choosing C sufficiently large ... $C = \sum_{t \in \mathcal{T}} tr(t)$

Greedy B&B idea:

1. Build the line system step by step by adding one line at a time
2. After each step, update the remaining demand and requirements
3. At each node, choose the line with the largest current number of direct travellers

The Need for a New Approach

Dienst approach:

- Works reasonably well in practice ($< 4.1\%$ gap), but is fundamentally slow
- If stopped early, it gives no guarantee on solution quality

Complexity:

- The line optimization problem is NP-hard
- Real-world railway networks lead to a combinatorial explosion
- Standard exact methods struggle with the size of the resulting models (even today)

Consequence:

The paper therefore develops a reduced model and later strengthens it with cutting planes

In the original LOP, direct travellers are tracked line by line through variables $d_{t,l}$

Main idea: Aggregate direct travellers across all usable lines

$$D_t := \sum_{l \in \mathcal{L},\, P_t \subseteq l} d_{t,l}$$ ... total number of direct travellers for pair $t \in \mathcal{T}$

Remarks:

- We no longer track which exact line carries these travellers, unlike in the original LOP
- This makes the lop formulation much smaller
- Every feasible solution of LOP yields a feasible solution of lop
- Therefore, lop is a relaxation of LOP
- Dienst's (infinite-capacity) model is weaker still, so it is a relaxation of lop

Model Reduction (lop) II

The goal is to maximize the total number of aggregated direct travellers

$$
\begin{aligned}
\max \quad & \sum_{t \in \mathcal{T}} D_t \\
\text{s.t.} \quad & D_t \leq tr(t) && \forall t \in \mathcal{T} \\
& D_t \leq C \sum_{l \in \mathcal{L},\, P_t \subseteq l} f_l && \forall t \in \mathcal{T} \\
& \sum_{l \in \mathcal{L},\, e \in l} f_l = lfr(e) && \forall e \in E \\
& D_t \geq 0, \quad f_l \in \mathbb{Z}^+ && \forall t \in \mathcal{T},\, l \in \mathcal{L}
\end{aligned}
$$

Remarks:

Since capacities are only aggregated, lop may overestimate feasible direct travellers

Cutting Planes I

A cutting planes methods strengthens an LP relaxation by adding valid inequalities

Main idea:

1. Solve the LP relaxation
2. If the solution is fractional, add inequalities satisfied by all integer solutions
3. These cuts remove fractional solutions but keep all feasible integer ones

In our context:

- Traveller variables are relaxed, but line frequencies f_l must be integer
- Cutting planes tighten the model before the B&B phase
- This reduces computation time

General cutting planes go back to Chvátal and Gomory

Cuts induced by the capacity constraint:

In the LP relaxation, optimal solutions often contain fractional frequencies

Lemma 1.

Let f be an integral line partition and define $\Delta_t := tr(t) - C\left\lfloor \frac{tr(t)}{C} \right\rfloor$. Then, for all $t \in \mathcal{T}$, the following inequality is valid for lop:

$$D_t \leq \left\lfloor \frac{tr(t)}{C} \right\rfloor (C - \Delta_t) + \Delta_t \sum_{l \in \mathcal{L},\, P_t \subseteq l} f_l.$$

Remarks:

- If $tr(t) \leq C$, this reduces to $D_t \leq tr(t) \sum_{l \in \mathcal{L},\, P_t \subseteq l} f_l$
- Thus, integrality of f_l yields a stronger capacity bound

Cuts induced by the network structure:

Fractional line partitions can also arise from the topology of the network.

Lemma 2.

Let $V' \subset V$, $E' \subseteq \{\{u,v\} \in E : |\{u,v\} \cap V'| = 1\}$, $\sum_{e \in E'} lfr(e)$ be odd and define $\alpha_l := |\{e \in E : e \in l\} \cap E'|$. Then every valid integral line partition f satisfies

$$\sum_{l \in \mathcal{L},\, \alpha_l \text{ even}} \alpha_l f_l \leq \sum_{e \in E'} lfr(e) - 1 .$$

Remarks:

- An odd cut demand forces at least one line to start or end inside the cut
- This rules out fractional solutions such as multiple lines with frequency $1/2$

Implementation:

- Capacity cuts yield at most one inequality per path $t \in \mathcal{T}$
- Network cuts grow exponentially with the size of the network

To keep the model manageable, we only add a subset of the cuts from Lemma 2:

For each station $v \in V$ with $\sum_{e=\{u,v\} \in E} lfr(e)$ odd, we add

$$\sum_{l \in \mathcal{L},\ l \text{ runs through } v} 2f_l \leq \sum_{e=\{u,v\} \in E} lfr(e) - 1.$$

Effect:

Adding both cut families substantially reduces the number of fractional frequency variables in the LP relaxation

Bounds I

Why do we need bounds?

- A simple B & B approach mainly provides lower bounds during execution
- If the algorithm is stopped early due to time limits, we may have a feasible solution but no clear measure of its absolute quality

Theoretical bounds for LOP:

- Since lop is a relaxation of LOP, its optimum $\hat{D}$ is an upper bound on the true optimum D^*
- A second upper bound is the maximum possible traveller flow in the network:

$$T^* = \max \sum_{t \in \mathcal{T}} D_t$$

$$\text{s.t. } D_t \leq tr(t) \ \forall t \in \mathcal{T}, \qquad \sum_{t \in \mathcal{T},\, e \in P_t} D_t \leq C\, lfr(e) \ \forall e \in E, \qquad D_t \geq 0 \ \forall t \in \mathcal{T}.$$

Lower bound from a fixed line partition:

- Restrict LOP to active lines $\mathcal{L}^f := \{l \in \mathcal{L} \mid f_l > 0\}$ from lop solution f
- Solving this reduced LP is fast and yields optimum $D(f)$, a valid lower bound:

$$
\begin{aligned}
D(f) = \max \quad & \sum_{l \in \mathcal{L}^f} \sum_{t \in \mathcal{T}} d_{t,l} \\
\text{s.t.} \quad & \sum_{l \in \mathcal{L}^f} d_{t,l} \leq tr(t) \quad \forall t \in \mathcal{T} \\
& \sum_{t \in \mathcal{T},\, e \in l \cap P_t} d_{t,l} \leq C f_l \quad \forall e \in E, l \in \mathcal{L}^f \\
& d_{t,l} \geq 0 \quad \forall t \in \mathcal{T}, l \in \mathcal{L}^f
\end{aligned}
$$

Confidence interval: $D(f) \leq D^* \leq \min(\hat{D}, T^*)$

In practice, these intervals are small and provide a good estimate of the solution quality

Real-world railway networks used in the paper:

- Deutsche Bahn: DB-IC, DB-IR
- Nederlandse Spoorwegen: NS-IC, NS-IR, NS-CT

Network	$\|V\|$	$\|\mathcal{CY}\|$	$\|E\|$	$\|\mathcal{L}\|$	$\|\mathcal{T}\|$
NS-IC	23	23	31	253	210
NS-IR	86	86	114	3655	2147
NS-CT	385	91	428	4095	11240
DB-IC	100	100	118	4950	3136
DB-IR	307	199	398	19701	9215

Table: Instance parameters of the five railway networks based on 1995 data.

Figure: German IC/ICE, IR network (2022)

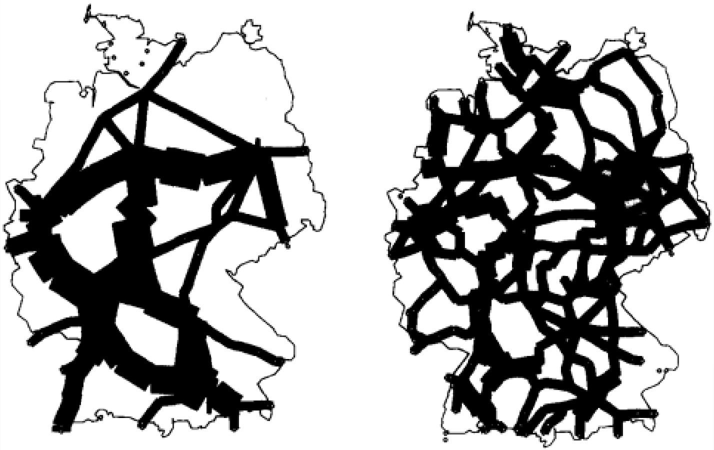

Figure: The line-frequency requirement for German IC/ICE, IR network (1996)

Data III

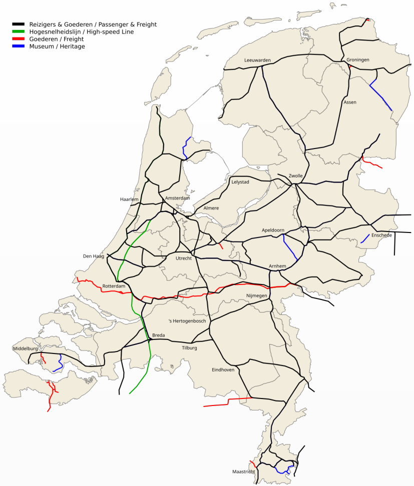

Figure: Dutch IC/ICE, IR, CT network (2017)

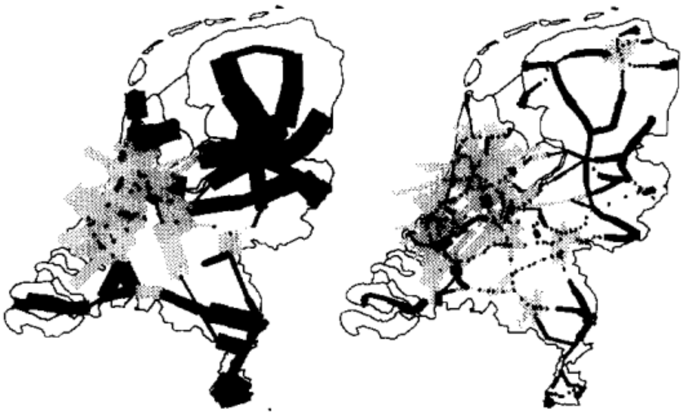

Figure: The line-frequency requirement for Dutch IC/ICE, IR, CT network (1996)

Results I

Original formulation (LOP):

- Massive formulation size (combinatorial explosion)
- LP relaxations are already expensive
- Only the smallest instance, NS-IC, could be solved to exact optimality
- Standard B & B performs poorly

Network	Variables	Constraints	LP time [s]	MIP time [s]	Status
NS-IC	2120	1017	10.55	4912.32	optimal MIP
NS-IR	116974	28487	5284.34	–	LP relaxation only
NS-CT	583598	96620	–	–	no LP solution
DB-IC	183235	43200	12777.66	–	LP relaxation only
DB-IR	900173	261308	–	–	no LP solution

Table: Size and computational status of the original LOP formulation.

Results II

Reduced formulation (lop):

- The reduced model is much smaller than the original LOP
- LP and MIP solution times decrease substantially
- However, many frequency variables are still fractional in the LP relaxation

Network	Variables	Constraints	LP time [s]	MIP time [s]	Fractional f_l
NS-IC	463	234	0.18	0.66	122
NS-IR	5802	2305	11.07	187.50	502
NS-CT	15335	11763	44.25	2128.47	390
DB-IC	8086	3276	39.54	103.23	512
DB-IR	28915	9692	94.02	1527.92	1347

Table: Computational results for the reduced formulation lop.

Reduced formulation (lop) with cutting planes:

- The strengthened model remains computationally tractable
- The number of fractional frequency variables is reduced drastically
- LP and MIP running times become very close

Network	LP time [s]	MIP time [s]	MIP* time [s]	Fractional f_l
NS-IC	0.38	0.48	0.38	0
NS-IR	32.73	37.93	30.62	8
NS-CT	143.82	161.87	358.94	0
DB-IC	50.47	56.28	55.74	0
DB-IR	312.47	408.92	420.86	3

Table: Computational results for lop tightened by cutting planes.

Bounds:

- Solving lop yields computable lower and upper bounds for the original LOP
- The resulting optimality gaps are small, never exceeding 3.1%

Network	$D(f)$	$\hat{D}$	T^*	Gap
NS-IC	8.203.412	8.203.412	9.168.554	0.0%
NS-IR	20.982.579	27.172.441	21.315.607	1.6%
NS-CT	25.079.912	37.118.270	25.863.252	3.1%
DB-IC	7.549.827	7.625.326	9.745.044	1.0%
DB-IR	6.114.448	6.114.448	8.682.953	0.0%

Table: Lower and upper bounds on the optimal value D^*.

...

Conclusion

Methodology & Results:

- MIP formulation for periodic railway line planning
- Reduced model lop with cutting planes makes real-world instances tractable
- All five tested networks solved in under 6 minutes, gaps below 3.2%

Limitations:

- Lines restricted to shortest paths and passengers assumed to follow them
- Aggregated capacities in lop may overestimate direct travellers

Extensions:

- Richer candidate lines
- Additional operational constraints and more general objectives
- Branch-cut separation
- Improved initialization heuristics

Thank You for Your Attention!