Minimizing recovery cost of network optimization problems

Zacharie Alès, Sourour Elloumi

Minimizing recovery cost of network optimization problems

Zacharie Ales[†,‡] Sourour Elloumi[†,‡]

[†] UMA, ENSTA Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France
[‡] CEDRIC, CNAM, Paris, France

{firstname.lastname}@ensta-paris.fr

Abstract

We propose a two-stage recoverable robustness approach that minimizes the recovery cost. In many applications, once the uncertainty ξ is revealed, it can be more important to recover a solution x^ξ which is as similar as possible to the nominal solution x^{nom} than to minimize the nominal objective value of x^ξ. This for example occurs when the nominal solution is implemented on a regular basis or when the uncertainty is revealed late. We define the *proactive problem* which minimizes the weighted recovery costs over a discrete set of scenarios while ensuring optimality of the nominal objective value of x^{nom}. We model the recovery cost of a scenario by a distance between the first-stage nominal solution and the second-stage solution recovered for this scenario.

We show for two different solution distances d_{val} and d_{struct} that the proactive problem is $\mathcal{NP}$-hard for both the integer min-cost flow problem with uncertain arc demands and for the integer max-flow problem with uncertain arc capacities. For these two problems, we prove that once uncertainty is revealed, even identifying a *reactive solution* x^r with a minimal distance to a given solution x^{nom} is $\mathcal{NP}$-hard for d_{struct}, and is polynomial for d_{val}.

We highlight the benefits of the proactive approach in a case study on a railroad planning problem. First, we compare it to the anchored and the k-distance approaches. Then, we show the efficiency of the proactive solution over reactive solutions. Finally, we illustrate the recovery cost reduction when relaxing the optimality constraint on the nominal objective of the proactive solution x^{nom}.

We also consider the min-max version of the proactive problem where we minimize the maximal recovery cost over all scenarios. We show that the same complexity results hold for this version. We also exhibit a class of problems for which the set of extreme points of the convex hull of a discrete uncertainty set always contain a worst-case scenario. We show that this result does not hold for three distinct classes deduced from the first one.

Keywords: Robust optimization; Uncertainty modelling; Recoverable robustness; $\mathcal{NP}$-hardness; Proactive problem; Mixed-Integer Linear Programming.

1 Introduction

Recoverable robustness can be viewed as a particular case of two-stage robustness. It identifies a *recoverable* solution of a deterministic nominal problem P. A solution of P is said to be recoverable if it can be recovered by limited means in all likely scenarios [23]. In this article, we focus on recoverable solutions which minimize recovery costs.

In many real-world problems, all solutions are recoverable as an operational solution must be implemented regardless of the scenario that occurs. For example, a railway company with a failing network will carry as many passengers as possible. In this context, recoverable robustness amounts to solving the deterministic problem as it would return any of the optimal solutions of P. However, not all of these optimal solutions are equivalent in terms of recovery costs. Thus, we propose a two-stage recoverable robustness approach that identifies an optimal solution of P and that additionally minimizes the sum of recovery costs. In this framework, a nominal solution is determined in the first stage while the second stage associates a solution to each scenario of a discrete uncertainty set.

A first motivation for minimizing the recovery costs is that the later the uncertainty is revealed, the more limited the recovery actions are. For example, in railway scheduling, a recovery may require communication between different departments to synchronize their resources or early transfers of trains between stations. These type of actions with a high recovery cost cannot be performed on short notice which may lead to worse-than-expected recovered solutions.

Moreover, the risk of human errors increases with the number of recovery actions. This is particularly the case in industrial applications where a solution is used on a regular basis. For example, in railway

planning a single schedule can be repeated every week. Since such a solution is regularly implemented, people involved in its realization (e.g., employees, clients, ...) are used to it. Therefore, when it must be recovered, they may, by habit, disrupt the realisation of the recovered solution. For example, users may miss a train if it leaves earlier than usual or an operator may activate a railroad switch that is no longer required. These disruptions may lead to incidents and customers dissatisfaction or even the need to recover the solution again. Consequently, the closer the recovered solutions are to the nominal solution, the less likely human errors are. Thus, we model the recovery costs by distances between the nominal solution of the deterministic problem P and the recovered solutions for each scenario.

We apply our approach to two robust network optimization problems: the integer min-cost flow with uncertain demands and the integer max-flow with uncertain capacities. This article is inspired by a practical flexible scheduling problem of SNCF, the French national railway company [24, 25]. This previous work motivated our choice to focus on two flow problems as they correspond to sub-problems of network design problems encountered by railway companies. In this context, the graph under consideration can represent a railway network and the flow the path of passengers or trains. Recovery costs for such problems can be expensive, hence the need to design relevant robust approaches.

In Section 2 we formally present this approach and introduce the nominal, the reactive and the proactive problems. We discuss links with existing approaches in Section 2.4. Sections 3 and Section 4 are dedicated to two network problems namely, the min-cost flow problem and the max-flow problem within the framework of solution robustness. We characterize the complexity of the reactive and the proactive problems for two different solution distances. We present a case study on a railroad planning problem in Section 5, highlighting the benefits of the proactive approach over the reactive approach. We compare the proactive approach to two other approaches from the literature: the anchored approach and the k-distance approach. Eventually, in Section 6 we consider a version of the proactive problem in which we consider the recovery cost of the worst-case scenario rather than the sum of the recovery costs over all scenarios. We show that the complexity results found for the sum of the recovery costs also hold in this worst-case version. We also show a class of problems for which considering a discrete set of uncertain scenarios is equivalent to considering its convex hull.

2 Reactive and proactive robustness solutions

We consider an optimization problem in which the set of feasible solutions depends on uncertain parameters within a set of possible realizations U. Let ξ^{nom} be the nominal value of these parameters. In the deterministic version of this optimization problem, called the *nominal problem*, the uncertain parameters take their nominal value ξ^{nom}. The nominal optimization problem is:

$$P \begin{cases} \min & f(x) \\ \text{s.t.} & x \in X(\xi^{nom}) \end{cases}$$

where $X(\xi^{nom})$ is the mapping which provides the feasibility set associated with parameters ξ^{nom}. The objective function $f : \mathbb{R}^n \to \mathbb{R}$ is called the *nominal objective*.

We model the uncertainty by a discrete set of scenarios that may occur $U = \{\xi_i\}_{i=1}^{|U|}$. To each scenario $\xi_i \in U$ is associated its feasibility set $X(\xi_i)$ and a weight $w_{\xi_i} \in \mathbb{R}^+$ which represents its importance or its probability of occurrence. Both feasibility sets $X(\xi^{nom})$ and $X(\xi_i)$ can be any subsets of $\mathbb{R}^n$. Each scenario in U corresponds to a possible realisation of the uncertainty (e.g., changes in the passenger demands, unavailability of resources, ...). Consequently, feasibility sets $X(\xi^{nom})$ and $X(\xi_i)$ may often in practice share similar constraints whose coefficients may vary. Note that $X(\xi_i)$ does not depend on the nominal solution x^{nom}. Thus, the reactive problem always has a solution provided that $X(\xi_i)$ is non empty.

In this article we will often consider robust flow problems which can be modelled as mixed-integer linear programs (MILP). To represent uncertain flow demands or capacities on the arcs of a graph, we consider uncertain right-hand side coefficients in Sections 3 and 6, where we prove complexity results, as well as in the case study presented in Section 5. In that context, a scenario ξ is a vector of $\mathbb{R}^m$ which directly corresponds to the value of these coefficients: $X(\xi) = \{x \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2} \mid Ax \leq \xi\}$.

We now introduce two problems that represent two possible approaches to handle the uncertainty. In the *reactive problem*, the uncertainty has not been anticipated and once it is revealed, a feasible solution must be obtained. The *proactive problem* is a robust approach in which a nominal solution and a recovery solution for each scenario are computed a priori.

2.1 Reactive robust solutions

We suppose that the implementation of a solution x^{nom} of the nominal problem P is planned and that scenario $\xi \in U$ occurs at operating time (i.e., a few days or hours before its realisation). To cause as little disruption as possible, the *reactive problem* provides a *reactive solution* $x^r \in X(\xi)$ whose distance to x^{nom} is minimal:

$$P^r(\xi, x^{nom}) \begin{cases} \min & d(x^r, x^{nom}) \\ \text{s.t.} & x^r \in X(\xi) \end{cases}$$

where $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a distance function between x^r and x^{nom} which models the recovery cost. The reactive problem P^r is considered once the uncertainty is revealed. We now introduce the proactive problem which allows us to further minimize the recovery cost by anticipating the uncertainties.

2.2 Proactive robust solutions

A reactive solution x^r may have a high nominal objective value $f(x^r)$ as f is not taken into account in the reactive problem. Furthermore, the recovery cost may also be high as it has not been anticipated. To address these issues, we introduce the two-stage *proactive problem* P^p whose variables are a *proactive solution* x^p of the nominal problem, set in the first stage, and a recovery solution x^i for each scenario $\xi_i \in U$, set in the second stage:

$$P^p(U, c^*) \begin{cases} \min\limits_{x^p \in X(\xi^{nom})} \quad \sum\limits_{\xi_i \in U} w_{\xi_i} \min\limits_{x^i \in X(\xi_i)} d(x^p, x^i) & (1) \\ \text{s.t. } f(x^p) = c^* & (2) \end{cases}$$

where c^* is the optimal value of the nominal problem.

Objective (1) minimizes the weighted sum of the recovery costs over all the scenarios while ensuring that x^p is feasible for (P) and that each x^i is feasible for scenario ξ_i. This objective is more similar to those of stochastic approaches, which consider an expectation over all scenarios, than to those of robust approaches, which tend to focus on the worst-case scenario. We study a min-max variant of the proactive problem in Section 6.

Constraint (2) ensures that the nominal objective value of x^p is equal to c^*. Consequently, x^p is an optimal solution of the nominal problem which additionally minimizes the weighted sum of the recovery costs over U. Similar to P^r, the first stage solution x^p does not constrain the second stage feasibility sets $\{X(\xi_i)\}_{\xi_i \in U}$.

Since the weights w_{ξ_i} are positive, each solution x^i is an optimal solution to the reactive problem $P^r(\xi_i, x^p)$. As a consequence, if the nominal problem has a unique optimal solution x^* then, solving $P^p(U, c^*)$ is equivalent to solving P and $P^r(\xi_i, x^*)$ for $i \in \{1, ..., |U|\}$.

The proactive problem P^p requires us to know the optimal value c^* of an optimal solution of the nominal problem. This should generally not be limiting as P is a significantly smaller problem than P^p given that P only finds one solution x while P^p simultaneously finds $|U| + 1$ solutions.

In order to further reduce the recovery costs, Constraint (2) can be relaxed to allow the nominal objective of the proactive solution $f(x^p)$ to deviate from optimality:

$$f(x^p) \leq c^*(1 + \varepsilon), \tag{3}$$

where $\varepsilon \in \mathbb{R}^+$ represents the maximal allowed percentage of increase of the nominal objective value of x^p. We observe the gains in terms of recovery costs resulting from this variation in the case study considered in Section 5.4.2.

2.3 Defining the recovery cost by two distances

A first intuitive distance corresponds to the ℓ_1 norm that we call the *distance in values*.

Definition 2.1. *The distance in values between two solutions $x^1 \in \mathbb{R}^n$ and $x^2 \in \mathbb{R}^n$ is*

$$d_{val}(x^1, x^2) \triangleq \sum_{j=1}^{n} |x^1_j - x^2_j|.$$

□

Depending on the context, d_{val} is not necessarily the most relevant distance. In rail planning, for example, increasing the number of cars in an existing train by one is much less expensive than planning a new train with only one car. This is because creating a train requires checking the availability of a locomotive and agents as well as the compatibility of the new train's schedule with that of other trains. Thus, we also introduce a *distance in structure*. Let $\mathbb{1}_b$ be equal to 1 if condition b is true and 0 otherwise.

Definition 2.2. *The distance in structure between two solutions $x^1 \in \mathbb{R}^n$ and $x^2 \in \mathbb{R}^n$ is*

$$d_{struct}(x^1, x^2) \triangleq \sum_{j=1}^{n} |\mathbb{1}_{x_j^1 > 0} - \mathbb{1}_{x_j^2 > 0}|.$$

□

Sometimes a problem has several categories of variables. For example, in a location problem, we have location variables and assignment variables. In such a context, it may be relevant to apply the distance to only one category of variables.

2.4 Literature review on recovery cost in robust optimization

In the line of [31], robust optimization approaches often search a nominal solution which is feasible for any value of the data in an uncertainty set [8, 11]. In this context, the recovery cost is equal to 0 as the solution is not modified once the uncertainty is revealed and the *price of robustness* [11] corresponds to the minimal increase of the nominal objective required to ensure the feasibility of a single solution in all possible scenarios.

However, in many concrete problems, a competitive nominal solution may not be feasible for all possible realisations of the uncertainty. To alleviate this problem, two-stage robust optimization approaches have been introduced where the value of recourse variables is computed only once the uncertainty is revealed [3, 7]. To ease the resolution of such problems, restrictions are often imposed on the recourse variables [7]. Their value can even be directly deduced from the uncertain data [5, 18, 28, 29], thus significantly reducing the complexity of the problem. In [6], the authors use such an approach for a robust supply chain problem in which a retailer commits to order predetermined quantities $\{w_t\}_{t=1}^T$ at each period of a fixed horizon T. The retailer can adjust the orders up and down at time t according to the demand of its clients but at the price of a penalty which is proportional to the deviation from w_t. This deviation can be viewed as a recovery cost. The objective is then to minimize a weighted sum of both the deviations and usual costs: ordering, holding, and shortage.

Liebchen et al. [23] introduced a general two-stage framework, called *recoverable robustness*, in which a solution x and a recovery algorithm A are determined such that the nominal objective is optimized and, once uncertainty is revealed, it is guaranteed that applying A to x leads to a feasible solution. Constraints can be imposed on A such as limiting its complexity or the distance between x and the recovered solution. This last constraint can be used to limit the recovery cost. However, it is usually not easy to ensure that these restrictions are satisfied for any uncertainty realization.

In [9] the authors consider a particular recoverable robustness approach for a scheduling problem in which the nominal objective is bounded and where the weighted sum of anchored jobs is maximized. A job is said to be *anchored* if it cannot be rescheduled once the uncertainty is revealed. This objective can be viewed as a way to limit the recovery cost. The authors consider uncertain processing times and show that maximizing the weight of the anchored jobs is polynomial for the box uncertainty set. It is however $\mathcal{NP}$-hard for both the budgeted and the polytope uncertainty sets.

Two other examples of recoverable robustness applied to shortest path problems are introduced in [13]. In *k-distance recoverable robustness*, the recovery actions are limited since at most k new arcs can be used once the uncertainty is revealed. In the second approach, called *rent recoverable robustness*, an edge is said to be *rented* if it is used in the first stage and it is *bought* if it is used in the second stage. For each scenario s and each arc e the cost c_e^s incurred for both renting and buying arc e is defined. The cost of the arc is lower if it is only rented (αc_e^s with $\alpha \in]0, 1[$) and higher if it is only bought ($(1+\beta)c_e^s$ with $\beta \geq 0$). This second approach is more flexible than the first one since the number of new edges used in the second stage is not constrained but it is also less generic. Indeed, the notion of renting and buying edges is relevant in problems such as railway scheduling where the network is owned by a company and exploited by others but it is not suitable for all applications. The authors study the complexity of both variants on three uncertainty sets and show that only the rent recoverable robustness on the interval set is polynomial. In Section 5, we compare the k-distance recoverable robustness and the anchored approach to our proactive approach.

Another type of recoverable robustness called *recovery-to-optimality* has been introduced in [16]. One of the main characteristics of this approach is that it constrains each recovered solution x^ξ to be optimal for the deterministic problem in which the uncertain data take their values from scenario $\xi \in U$. Similar

to our proactive approach, it minimizes the average or the worst recovery cost over all the scenarios. However, in recovery-to-optimality, a nominal solution x^{nom} may require a significantly higher recovery cost in order to be transformed into a solution x^{ξ} with an optimal nominal objective for scenario ξ. The authors apply their approach to a linear program for timetabling in public transportation. The optimality constraints of the recovered solution x^{ξ} for each scenario ξ are harder to ensure than the optimality of x^{nom} required by our approach. As a consequence, they consider a heuristic in which optimal solutions of a subset of scenarios U are first computed and then a solution x which minimizes the recovery cost to these solutions is obtained.

The notion of (a, b)-supermodel introduced in [15] is related to the recovery cost. A solution x is said to be an (a, b)-*supermodel* if whenever the value of at most a variables in x is changed, then a feasible solution of the original problem can still be obtained by changing the value of at most b other variables. The authors show that a $(1, 1)$-supermodel of SAT can be obtained by solving a larger instance of SAT. The originality of this approach is the fact that the uncertainty is not in the data but in the value of the solution itself.

The identification of efficient schedules both in terms of nominal objective and recovery cost has been considered under different names: predictability [26, 27], stability [19, 20, 21, 22], solution robustness [30, 32]. Exact approaches are rarely considered to solve such problems. One exception is [21] in which a hand-made branch-and-bound algorithm is considered to solve a single-machine scheduling problem in which only one job is expected to be disturbed. In other works, heuristic approaches are generally considered. For example, in [30] the authors present a multi-objective genetic algorithm in which both the recovery cost and the nominal objective are optimized.

In the next two sections, we apply the reactive and proactive frameworks defined in Sections 2.1 and 2.2 for the integer min-cost flow and the integer max flow problems. It is well known that the deterministic version of both of these problems can be solved with polynomial time algorithms [1]. We show that this is not the case for their proactive counterparts with uncertain arc demands for the min-cost flow problem and uncertain arc capacities for the max-flow problem. Note that polynomial problems that become $\mathcal{NP}$-hard when uncertainty is added have been studied previously for various robust approaches (see e.g., [2, 7, 10, 12, 13]).

3 Minimizing recovery cost of the min-cost flow problem with uncertain arc demands

The min-cost flow problem can be stated as:

MIN-COST FLOW PROBLEM

Input: A digraph $G = (V, A)$ with arc demands $\xi_a^{nom} \in \mathbb{N}$, capacities $u_a \in \mathbb{N}$ and unit cost $c_a \in \mathbb{R}^+$ of each arc $a \in A$ and node demands $b_v \in \mathbb{Z}$ for each node $v \in V$ ($b_v > 0$ if v is a *supply node*, $b_v < 0$ if v is a *demand node* and $b_v = 0$ if v is a *transshipment node*).

Output: Find an integer flow with minimal cost.

For a node $v \in V$, let $\delta^-(v)$ and $\delta^+(v)$ be the set of predecessors and successors of v in G, respectively. An integer flow $f \in \mathbb{N}^{|A|}$ is feasible for this problem if it satisfies all the arc demands and capacities (i.e., $f_a \in [\xi_a^{nom}, u_a]$ $\forall a \in A$) and the node demands (i.e., $\sum_{u \in \delta^-(v)} f_{uv} - \sum_{u \in \delta^+(v)} f_{vu} = b_v$ $\forall v \in V$).

We assume that the uncertainties are on the arc demands ξ^{nom}. Consequently, each scenario $\xi \in U$ is a vector of arc demands in $\mathbb{N}^{|A|}$ and $X(\xi)$ represents the set of feasible integer flows when the arc demands are set to ξ. The corresponding reactive and proactive problems can be stated as follows:

REACTIVE MIN-COST FLOW PROBLEM

Input: A min-cost flow problem, one of its solutions $f^{nom} \in X(\xi^{nom})$, a distance $d : \mathbb{N}^{|A|} \times \mathbb{N}^{|A|} \to \mathbb{R}^+$ and arc demands $\xi \in \mathbb{N}^{|A|}$.

Output: Find a reactive flow $f^r \in X(\xi)$ which minimizes $d(f^{nom}, f^r)$.

PROACTIVE MIN-COST FLOW PROBLEM

Input: A min-cost flow problem with demands $\xi^{nom} \in \mathbb{N}^{|A|}$, its optimal flow value c^*, a distance $d : \mathbb{N}^{|A|} \times \mathbb{N}^{|A|} \to \mathbb{R}^+$, a discrete set of scenarios $U \subset \mathbb{N}^{|A|}$ and for each scenario $\xi \in U$, a weight $w_\xi \in \mathbb{R}^+$.

Output: Find a proactive flow $f^p \in X(\xi^{nom})$ and a flow $f^\xi \in X(\xi)$ for each scenario $\xi \in U$ which minimize $\sum_{\xi \in U} w_\xi d(f^p, f^\xi)$.

In the following subsections, we characterize the complexity of the four reactive and proactive min-cost flow problems associated with distances d_{val} and d_{struct}, as represented in Table 1. Note that all these problems are in $\mathcal{NP}$.

Problem	**Distance**	
	d_{val}	d_{struct}
Proactive	strongly $\mathcal{NP}$-hard (Section 3.1)	strongly $\mathcal{NP}$-hard (Section 3.2) even with 1 scenario
Reactive	Polynomial (Section 3.1)	strongly $\mathcal{NP}$-hard (Section 3.2)

Table 1: Complexity of the min-cost flow problems with uncertain arc demands associated with distances d_{val} and d_{struct}.

3.1 Complexity of the robust min-cost flow with distance d_{val}

We show in this section that for distance d_{val} the proactive problem is $\mathcal{NP}$-hard and that the reactive problem is polynomial.

Theorem 3.1. *The proactive min-cost flow problem with distance d_{val}, $MCF^p_{d_{val}}$, is strongly $\mathcal{NP}$-hard.*

Proof. We prove this result with a reduction from problem 3-SAT. This problem considers a boolean expression composed of n variables $X = \{x_1, ..., x_n\}$ and m clauses $\{C_1, ..., C_m\}$. Let a *literal* be either a boolean variable x_i or its negation $\overline{x}_i$. Each clause of a 3-SAT problem is a conjunction of three literals (e.g., $C_1 = (x_1 \vee \overline{x}_2 \vee x_3)$). The aim of this problem is to determine whether there exists an assignment of the variables which satisfies all the clauses.

We now present how to construct an instance I_{MCF} of the optimization problem $MCF^p_{d_{val}}$ from an instance I_{SAT} of the feasibility problem 3-SAT. Instance I_{MCF} has m scenarios. We prove that an optimal solution of I_{MCF} leads to a recovery cost of value $4mn$ if and only if I_{SAT} is a yes-instance.

Construction of I_{MCF}

As illustrated in Figure 1a, the set of nodes V of the constructed graph $G = (V, A)$ is composed of:

- one source node s with supply $b_s = n$ and one sink node t with demand $b_s = -n$;
- for each boolean variable x_i one *variable node* x_i and two *literal nodes* l_i and $\bar{l}_i$ with no demand $b_{x_i} = b_{l_i} = b_{\bar{l}_i} = 0$;
- for each clause C_p one *clause node* C_p with no demand $b_{C_p} = 0$.

The arc set A is composed of the following subsets:

- A_{sl} contains one arc from s to each literal node;
- A_{lx} contains one arc from each literal node to its corresponding variable node;
- A_{lC} contains one arc (l, C_p) for each clause C_p and each literal l in this clause;
- A_{xt} contains one arc from each variable node to t;
- A_{Ct} contains one arc from each clause node to t;
- A_{st} contains one arc from s to t.

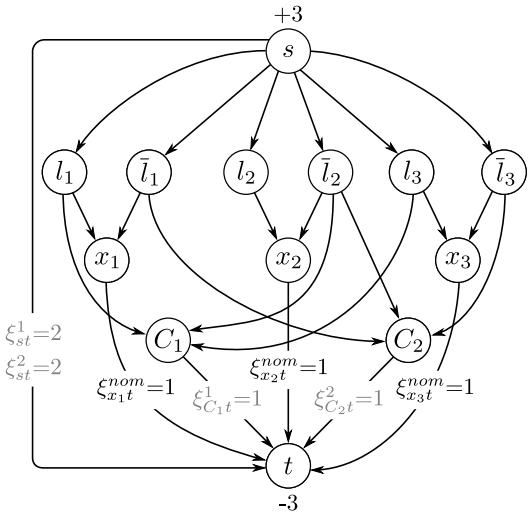
The nominal arc demand of the arcs is 0 except for the n arcs of A_{xt} where it is equal to 1. The capacity of any arc is m and its cost is 0. One can easily check that the nominal flow problem is feasible, see the illustration in Figure 1b. Any feasible flow costs 0. Consequently, c^* is equal to 0.

Note that the size of the constructed graph is polynomial in n and m. It has $3n + m + 2$ nodes and $5n + 4m + 1$ arcs.

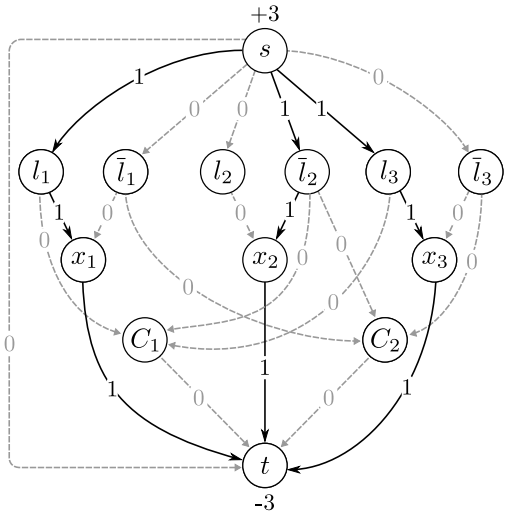
Instance I_{MCF} contains m scenarios, one scenario $\xi^p \in \mathbb{N}^{|A|}$ per clause C_p. For each scenario ξ^p the weight w_{ξ_p} is set to 1, the arc demands are all set to 0 except $\xi^p_{s,t}$ which is set to $n - 1$ and $\xi^p_{C_p,t}$ which is set to 1 (see Figure 1a).

Solving I_{SAT} from a solution of I_{MCF}

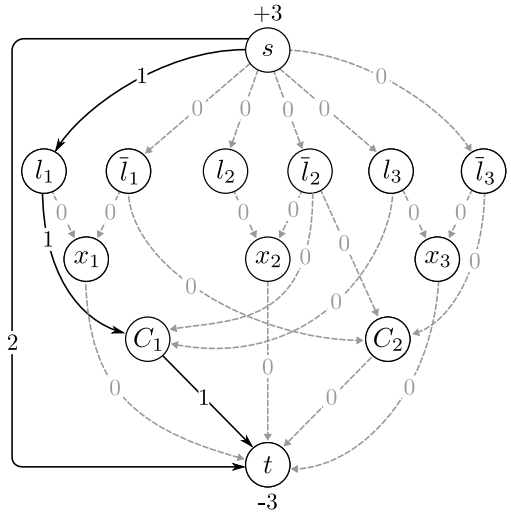
Let f be a nominal feasible flow of I_{MCF}. Since the graph does not contain any cycle, the n units supplied by s are necessarily sent on n different paths to satisfy the arc demands $\xi^{nom}_{x_i,t} = 1$ for all $i \in \{1, ..., n\}$ (see example in Figure 1b).



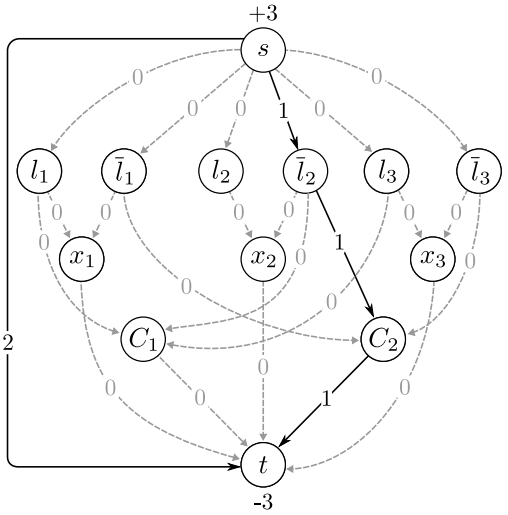
(a) Graph and scenarios of I_{MCF}. Only the positive demands are represented.



(b) Nominal flow f.



(c) Flow of the first scenario f^{ξ_1}.



(d) Flow of the second scenario f^{ξ_2}.

Figure 1: Example of the reduction to $MCF^p_{d_{val}}$ of an instance of the 3-SAT problem with three variables and two clauses $C_1 = (x_1 \vee \overline{x}_2 \vee x_3)$ and $C_2 = (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3)$. Figures 1b, 1c and 1d represent an optimal solution $(f, f^{\xi_1}, f^{\xi_2})$ to $MCF^p_{d_{val}}$. The arcs with a flow of value 0 are colored in gray.

Let f^{ξ_p} be the flow of scenario ξ_p. To satisfy the arc demands $\xi^p_{s,t} = n - 1$ and $\xi^p_{C_p,t} = 1$, flow f^{ξ_p} necessarily sends $n - 1$ units of flow on arc (s, t) and one unit of flow on a path (s, l, C_p, t) with l a literal included in clause C_p (see Figures 1c and 1d). As represented in Table 2, each scenario leads to a recovery cost of value $4n$ or $4n + 2$. A cost of $4n$ is obtained if and only if the arc of A_{sl} used in f^{ξ_p} is also used in the nominal flow.

Consequently, the recovery cost over all the scenarios is at least $4mn$. A cost of $4mn$ is obtained if and only if all the arcs of A_{sl} used by the scenarios are also used in the nominal flow (i.e., if setting to true all the variables x_i such that $f_{s,l_i} > 0$ and to false the others enables us to satisfy all the clauses). Therefore, the answer to instance I_{SAT} is true if and only if the optimal solution value of I_{MCF} is $4mn$. □

Definition 3.1. *([1], Section 1.2) A convex cost flow problem is a min-cost flow problem where the cost of an arc is a piecewise linear convex function of its flow.*

Property 3.1. *The reactive min-cost flow problem with distance d_{val}, $MCF^r_{d_{val}}$, is a convex cost flow problem.*

Proof. The cost of an arc $a \in A$ is $|x^r_a - x_a|$ which is a convex piecewise linear function of flow x^r_a. □

Arc sets	$\sum_a f_a$	$\sum_a f_a^{\xi_p}$	**Recovery cost**
A_{st}	0	$n-1$	$n-1$
A_{lx}	n	0	n
A_{xt}	n	0	n
A_{lC}	0	1	1
A_{Ct}	0	1	1
A_{sl}	n	1	$n-1$ or $n+1$
		Total	$4n$ or $4n+2$

Table 2: Recovery cost induced by scenario ξ_p for each set of arcs in the proof of Theorem 3.1. The second column corresponds to the sum of the flows for any nominal solution while the third column contains the sum of the flows for any solution of scenario ξ_p. The recovery cost associated with A_{sl} is either $n-1$ if the arc of A_{sl} used in f^{ξ_p} is also used in the nominal flow or $n+1$ otherwise.

Theorem 3.2. *([1], Section 14.3) A convex cost flow problem with piecewise linear convex cost functions can be transformed into a min-cost flow problem.*

Corollary 3.1. *The reactive min-cost flow problem with distance d_{val}, $MCF^r_{d_{val}}$, is a polynomial problem.*

3.2 Complexity of robust min-cost flow with distance d_{struct}

We show that for distance d_{struct} both the proactive and the reactive problems are $\mathcal{NP}$-hard.

Theorem 3.3. *The proactive min-cost flow problem with distance d_{struct}, $MCF^p_{d_{struct}}$, is strongly $\mathcal{NP}$-hard.*

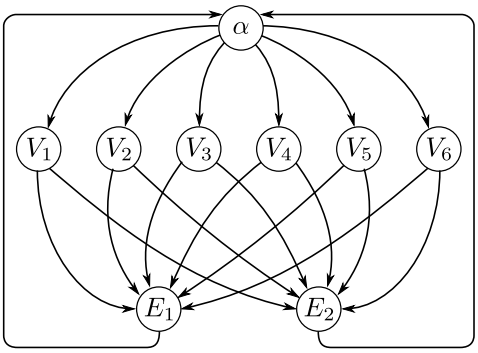
Proof. We prove the theorem with a reduction from problem 3-Partition. This problem considers a set $E = \{1, 2, ..., 3m\}$ of $3m$ elements and a positive integer B. To each element $i \in E$ is associated a size $s(i) \in]\frac{B}{4}, \frac{B}{2}[$ such that $\sum_{i \in E} s(i) = mB$. The aim of this problem is to determine whether there exists a partition $\{E_1, ..., E_m\}$ of E into m subsets such that the size of each subset E_j, $\sum_{i \in E_j} s(i)$, is equal to B. Note that since $s(i) \in]\frac{B}{4}, \frac{B}{2}[$, the cumulative sizes of 2 and 4 objects are respectively lower and greater than B. As a consequence each set necessarily contains exactly 3 elements.

We now present how to construct an instance I_{MCF} of the optimization problem $MCF^p_{d_{struct}}$ from an instance I_{3P} of the feasibility problem 3-Partition and we prove that a solution of I_{MCF} has a recovery cost of $7m$ if and only if I_{3P} is a yes-instance.

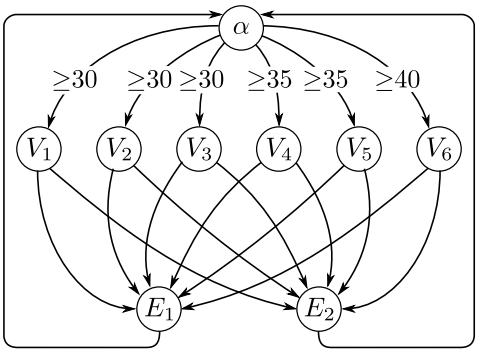
Construction of I_{MCF}

Figure 2a illustrates the graph obtained for an instance of problem 3-Partition with $m = 2$ and $B = 100$.

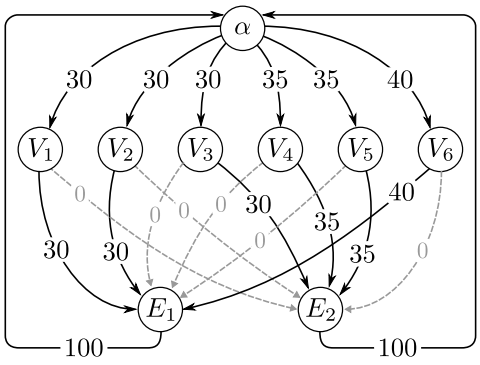
The set of nodes V of the constructed graph $G = (V, A)$ is composed of one node α, one node V_i for each element $i \in E$ and one *subset node* E_j for each subset $j \in \{1, ..., m\}$. The demand of all the nodes is equal to 0.



(a) Graph of I_{MCF}. All node demands are 0. For any arc, demand is 0, capacity is B and unitary cost is 1.



(b) Positive demands of the scenario.



(c) Optimal proactive flow inducing the 3-partition $E_1 = \{1, 2, 6\}$ and $E_2 = \{3, 4, 5\}$. The arcs with a flow of value 0 are colored in gray.

Figure 2: Example of the reduction to $MCF^p_{d_{struct}}$ of an instance of the 3-Partition problem with $m = 2$, $B = 100$ and elements of size $(30, 30, 30, 35, 35, 40)$.

The arc set A contains one arc (α, V_i) for each $i \in E$ and one arc (V_i, E_j) for each $i \in E$ and each $j \in \{1, ..., m\}$. The graph also contains an additional set of arcs, denoted by $A_{E\alpha}$, which contains one arc from each subset node to α. Note that the size of the graph is polynomial in n and m as it contains $4m + 1$ nodes and $3m^2 + 4m$ arcs.

For any arc, the nominal demand is 0, the capacity is B and the unitary cost is 1. It follows that the only optimal nominal solution is the empty flow, with the objective value $c^* = 0$.

One unique scenario ξ is considered in which the only arc demands different from 0 are ξ_{α,V_i} of arcs (α, V_i) for all $i \in E$ which are set to $s(i)$ (see Figure 2b).

Solving I_{3P} from a solution of I_{MCF}

Since $c^* = 0$, the only feasible proactive solution is the empty flow.

Let f^ξ be a feasible solution for scenario ξ. Observe that the outgoing flow from α is at least Bm. Moreover, arcs (E_j, α) have a capacity of B. Thus, the arcs (E_j, α) must all be saturated and the flow running through α is exactly Bm. Since the flow on each arc (E_j, α) is 0 in the nominal flow and B in f^ξ, these m arcs induce a recovery cost of m.

We now show that at least an additional cost of $2|E|$ is induced by the remaining arcs. Each V_i has an ingoing flow of value $s(i)$. Consequently, f^ξ must send a total flow of $s(i)$ on at least one of the m following paths $\{(\alpha, V_i, E_j)\}_{j=1}^m$. To sum up, f^ξ uses at least 2 additional arcs for each element $i \in E$, leading to a recovery cost of $m + 2|E|$ (see Figure 2c). The use of several paths to satisfy a demand $\xi_{\alpha V_i}$ would increase the recovery cost as several arcs from $\{(V_i, E_j)\}_{j=1}^m$ would be used.

Consequently, a recovery cost of $7m$ is obtained if and only if for each $i \in E$ exactly one path is used to satisfy the demand $\xi_{\alpha V_i}$ (i.e., if a partition with subsets of size B is obtained by assigning each element i to the unique set E_j such that $f^\xi_{V_i E_j} > 0$).

□

Theorem 3.4. *The reactive min-cost flow problem with distance d_{struct}, $MCF^r_{d_{struct}}$, is strongly $\mathcal{NP}$-hard.*

Proof. In the reduction considered in the proof of Theorem 3.3, the proactive flow is unique and known a priori as it is necessarily empty. Consequently, this flow can be given as an input of the problem rather

than being part of its solution without altering the validity of the reduction.

This new reduction leads to an instance of $MCF^r_{d_{struct}}$ since the nominal flow is fixed and only 1 scenario is considered. □

4 Minimizing recovery cost of the max-flow problem with uncertain arc capacities

The max-flow problem can be stated as:

Max-flow Problem	
Input:	A digraph $G = (V, A)$ with a source $s \in V$, a sink $t \in V$ and capacities $\xi^{nom}_a \in \mathbb{N}$ on the flow of each arc $a \in A$.
Output:	Find an integer flow with maximum value.

We consider max-flow problems with uncertainties on the capacities. Consequently, each scenario $\xi \in U$ is a vector of arc capacities in $\mathbb{N}^{|A|}$. Note that the associated reactive and proactive problems are not particular cases of the ones considered in the previous section as the uncertainty is not on the arc demands and $X(\xi)$ represents the set of feasible integer flows when the arc capacities are set to ξ. The corresponding reactive and proactive problems can be stated as follows:

Reactive Max-Flow Problem	
Input:	A max-flow problem, one of its solutions $f^{nom} \in X(\xi^{nom})$, a distance $d: \mathbb{N}^{|A|} \times \mathbb{N}^{|A|} \to \mathbb{R}^+$ and arc capacities $\xi \in \mathbb{N}^{|A|}$.
Output:	Find a reactive flow $f^r \in X(\xi)$ which minimizes $d(f^{nom}, f^r)$.

Proactive Max-Flow Problem	
Input:	A max-flow problem with capacities $\xi^{nom} \in \mathbb{N}^{|A|}$, its optimal flow value c^*, a distance $d: \mathbb{N}^{|A|} \times \mathbb{N}^{|A|} \to \mathbb{R}^+$ and for each scenario $\xi \in U$, a weight $w_\xi \in \mathbb{R}^+$.
Output:	Find a proactive flow $f^p \in X(\xi^{nom})$ and a flow $f^\xi \in X(\xi)$ for each scenario $\xi \in U$ which minimize $\sum_{\xi \in U} w_\xi d(f^p, f^\xi)$.

In the remainder of this section we determine the complexity of the four reactive and proactive max-flow problems associated with distances d_{val} and d_{struct}, as represented in Table 3.

Problem	**Distance**	
	d_{val}	d_{struct}
Proactive	strongly $\mathcal{NP}$-hard (Section 4.1)	strongly $\mathcal{NP}$-hard (Section 4.2) even with 1 scenario
Reactive	Polynomial (Section 4.1)	strongly $\mathcal{NP}$-hard (Section 4.2)

Table 3: Complexity of the max-flow problems with uncertain arc capacities associated with distances d_{val} and d_{struct}.

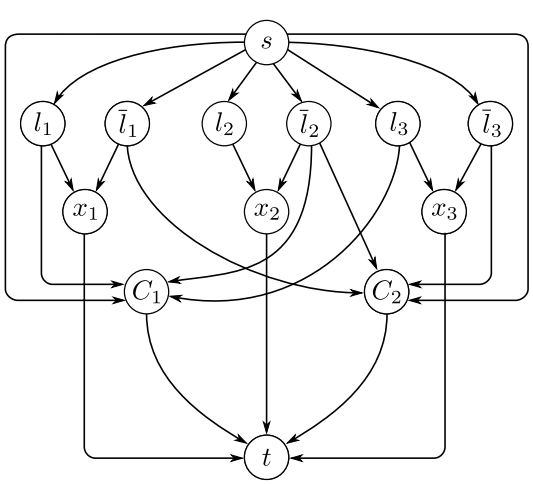
4.1 Complexity of the robust max-flow with distance d_{val}

We show in this section that for distance d_{val} the proactive problem is $\mathcal{NP}$-hard and that the reactive problem is polynomial.

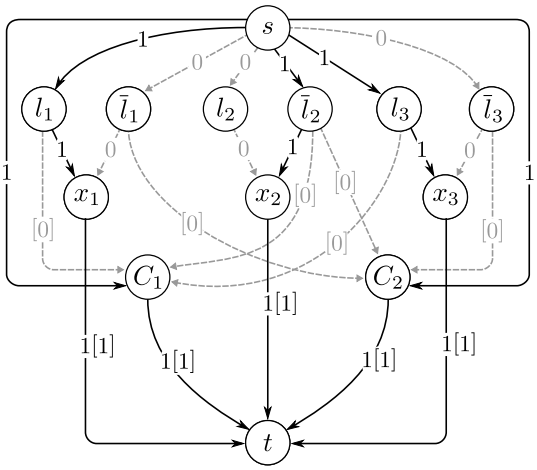
Theorem 4.1. *The proactive max-flow problem with distance d_{val}, $MF^p_{d_{val}}$, is strongly $\mathcal{NP}$-hard.*

Proof. The proof of this theorem is similar to the one of Theorem 3.1. An instance I_{MF} of the optimization problem $MF^p_{d_{val}}$ is constructed from an instance I_{SAT} of the feasibility problem 3-SAT and we prove that I_{MF} has an optimal solution with a recovery cost equal to $m(3n + 2m - 1)$ if and only if I_{SAT} is a yes-instance.

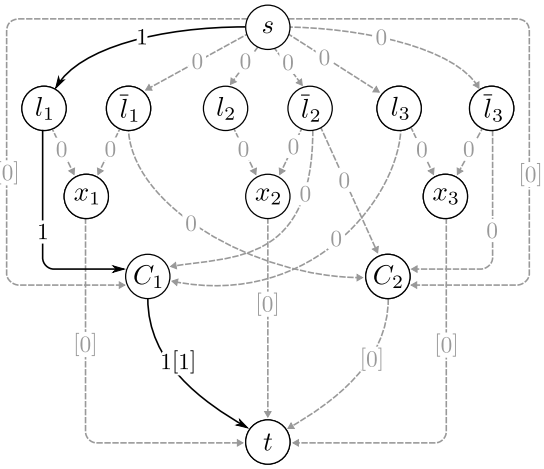
Construction of I_{MF}



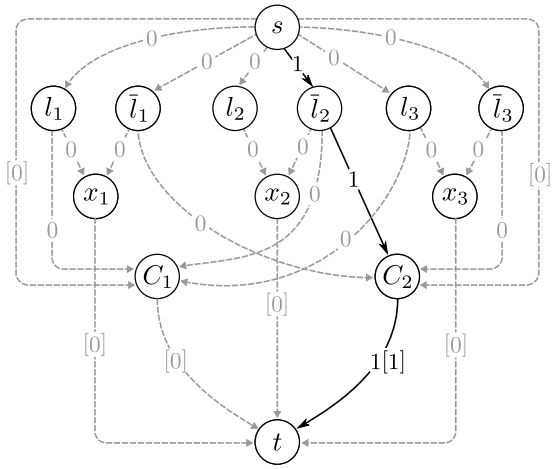
(a) Graph of I_{MF}.



(b) Nominal flow f. The nominal capacities different from $+\infty$ are represented between brackets.



(c) Flow f^{ξ^1} of the first scenario ξ^1. The scenario capacities different from $+\infty$ are represented between brackets.



(d) Flow f^{ξ^2} of the second scenario ξ^2. The scenario capacities different from $+\infty$ are represented between brackets.

Figure 3: Example of the reduction to $MF^p_{d_{val}}$ of an instance of the 3-SAT problem with three variables and two clauses $C_1 = (x_1 \vee \overline{x}_2 \vee x_3)$ and $C_2 = (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3)$ to $MF^p_{d_{val}}$. Figures 3b, 3c and 3d represent an optimal solution $(f, f^{\xi^1}, f^{\xi^2})$. The arcs with a flow of value 0 are colored in gray.

As represented in Figure 3, the graph associated with I_{MF} has the same sets of nodes and arcs as the ones in the proof of Theorem 3.1 except for the arc (s,t) which does not exist. The graph also includes an additional set of arcs A_{sC} which contains one arc between s and each clause node.

The nominal capacity of all the arcs is $+\infty$ except for the arcs in A_{xt} and A_{Ct} for which it is equal to 1 and the arcs in A_{lC} for which it is null. As proved below, c^* is equal to $n+m$.

Instance I_{MF} contains m scenarios $\{\xi^1, ..., \xi^m\}$ with weights equal to 1. The capacities of the arcs of a scenario ξ^p are all equal to $+\infty$ except for the arcs in A_{xt}, A_{sC} and $A_{Ct}\backslash\{(C_p,t)\}$ for which it is equal to 0 and $\xi^p_{C_p,t}$ which is equal to 1 (see Figure 3a).

Solving I_{SAT} from a solution of I_{MF}

Let f be a nominal feasible flow of I_{MF}. Since the $n+m$ arcs arriving in t all have a nominal capacity of 1, the nominal flow value is at most $n+m=c^*$.

Flow f necessarily sends one unit of flow for each clause C_p on path (s, C_p, t) and one unit of flow for each variable x_i on one of the following two paths (s, l_i, x_i, t) or $(s, \bar{l}_i, x_i, t)$ (see example in Figure 3b).

Let f^{ξ^p} be the flow of scenario ξ^p. Due to the scenario capacities, f^{ξ^p} can either be an empty flow (solution $f^{\xi^p,0}$) or send one unit of flow from s to t through a path (s, l, C_p, t) with l a literal of clause C_p (solution $f^{\xi^p,l}$) (see Figures 3c and 3d). As represented in Table 4, solution $f^{\xi^p,0}$ leads to a recovery cost of value $3n+2m$ and $f^{\xi^p,l}$ to a recovery cost in $[3n+2m-1, 3n+2m+1]$. A cost of $3n+2m-1$ is obtained if and only if the arc of A_{sl} used in $f^{\xi^p,l}$ is also used in the nominal flow.

Consequently, the sum of the recovery costs over all the scenarios is between $m(3n+2m-1)$ and $m(3n+2m+1)$. An optimal cost of $m(3n+2m-1)$ is obtained if and only if all the arcs of A_{sl} used

Arc sets	$\sum_a f_a$	$\sum_a f_a^{\xi^p,0}$	$\sum_a f_a^{\xi^p,l}$	Recovery cost	
				$\sum_a \lvert f_a^{\xi^p,0} - f_a \rvert$	$\sum_a \lvert f_a^{\xi^p,l} - f_a \rvert$
A_{sC}	m	0	0	m	m
A_{lC}	0	0	1	0	1
A_{Ct}	m	0	1	m	$m-1$
A_{sl}	n	0	1	n	$n \pm 1$
A_{lx}	n	0	0	n	n
A_{xt}	n	0	0	n	n
			Total	$3n+2m$	$3n+2m \pm 1$

Table 4: Recovery cost induced by scenario ξ^p for each set of arcs in the proof of Theorem 4.1. The second column corresponds to the sum of the flows for any nominal solution of value $c^* = n+m$. The third and the fourth columns contain the sum of the flows for the two possible types of solutions of scenario ξ^p: the empty flow $f^{\xi^p,0}$ and a unitary flow $f^{\xi^p,l}$. In the last column, the recovery cost associated with A_{sl} is either $n-1$ if the arc of A_{sl} used in $f^{\xi^p,l}$ is also used in the nominal flow or $n+1$ otherwise.

by the scenarios are also used in the nominal flow (i.e., if setting to true all the variables x_i such that $f_{s,l_i} > 0$ and to false the others enables us to satisfy all the clauses).

□

We show that $MF^r_{d_{val}}$ corresponds to a *convex cost flow problem* which is known to be polynomial.

Property 4.1. *The reactive max-flow problem with distance d_{val}, $MF^r_{d_{val}}$, is a polynomial problem.*

Proof. $MF^r_{d_{val}}$ is a convex cost flow problem since the cost of an arc $a \in A$ is $|x^r_a - x_a|$. Since a max-flow problem is a particular min-cost flow problem, according to Theorem 3.2, $MF^r_{d_{val}}$ is a polynomial problem. □

4.2 Complexity of the robust max-flow with distance d_{struct}

We show that for distance d_{struct} both the proactive and the reactive problems are $\mathcal{NP}$-hard.

Theorem 4.2. *The proactive $MF^p_{d_{struct}}$ is strongly $\mathcal{NP}$-hard.*

Proof. Similar to the proof of $MCF^p_{d_{val}}$, we use a reduction from problem 3-SAT.

Let I_{SAT} be an instance of the feasibility problem 3-SAT with m clauses and n variables. As represented in Figure 4, we create an instance I_{MF} of the optimization problem $MF^p_{d_{struct}}$ with one unique scenario ξ. We prove that an optimal solution of I_{MF} leads to a recovery cost of $3n+2m$ if and only if I_{SAT} is a yes-instance.

Construction of I_{MF}

In I_{MF}, the set of nodes V of graph $G = (V, A)$ is composed of:

- two nodes s and t;
- $2n$ *literal nodes* $\{l_i\}_{i=1}^n$ and $\{\bar{l}_i\}_{i=1}^n$;
- $3n$ *variable nodes* $\{x_i^1, x_i^2, x_i^3\}_{i=1}^n$;
- $2m$ *clause nodes* $\{C_p^1, C_p^2\}_{p=1}^m$.

The arc set A is composed of the following subsets:

- $A_{s,l}$ which contains one arc from s to each literal node;
- $A_{s,x}$ which contains one arc from s to each first variable node x_i^1;
- $A_{l,x}$ which contains one arc from each literal node to its corresponding first variable node x^1;
- A_x which contains arcs (x_i^1, x_i^2) and (x_i^2, x_i^3) for each i;
- $A_{x,t}$ which contains one arc from each last variable node x_i^3 to t;

Arc sets	Capacities	
	ξ^{nom}	ξ
$A_{s,l}$	0	$m+1$
$A_{s,x}$, $A_{s,C}$	1	0
$A_{l,x}$, $A_{l,C}$	0	1
A_x, $A_{x,t}$, A_C, $A_{C,t}$	1	1

Table 5: Capacities of each set of arcs for the nominal case and scenario ξ in instance I_{MF}.

- $A_{s,C}$ which contains one arc from s to each first clause node C_p^1;
- $A_{l,C}$ which contains one arc (l, C_p^1) for each clause C_p and each literal l in C_p. Hence, 3 such arcs exist for each clause;
- A_C which contains one arc (C_p^1, C_p^2) for each clause C_p;
- $A_{C,t}$ which contains one arc from each clause node C_p^2 to t.

Note that the size of the constructed graph is polynomial in n and m. It has $5n + 2m + 2$ nodes and $8n + 6m$ arcs.

As proved below, the value of c^* is equal to $n + m$. The nominal capacities ξ^{nom} and the scenario capacities ξ for each arc subset are presented in Table 5.

Solving I_{SAT} from a solution of I_{MF}

Let f be a nominal flow of I_{MF}. As exemplified in Figure 4b, the maximal nominal flow is $c^* = n+m$ and it can only be reached by a unique flow which sends:

- one unit for each clause C_p through path (s, C_p^1, C_p^2, t); and
- one unit for each variable x_i through path $(s, x_i^1, x_i^2, x_i^3, t)$.

Let f^ξ be the flow of the scenario. An empty flow f^ξ would lead to a recovery cost of $4n + 3m$ (i.e., the number of arcs used by the nominal flow). This can be improved by sending, for each variable x, one unit of flow along one of the two following paths: $P_x = (s, l, x^1, x^2, x^3, t)$ or $\overline{P}_x = (s, \bar{l}, x^1, x^2, x^3, t)$. Let us assume without loss of generality that only P_x is used. This leads to a reduction of 1 of the recovery cost. Indeed, three arcs of the path are also used in the nominal flow ((x^1, x^2), (x^2, x^3), (x^3, t)) and two are not ((s, l) and (l, x^1)). Note that both P_x and $\overline{P}_x$ cannot be used simultaneously as the capacity of (x^1, x^2) is equal to 1. Thus, the recovery cost of f^ξ is reduced by n by sending one unit of flow on either P_x or $\overline{P}_x$ for each variable x (see Figure 4d). Let A_{sl}^ξ be the subset of arcs of A_{sl} now used in f^ξ.

To further reduce the recovery cost, f^ξ must use arcs from A_C and $A_{C,t}$. The clause nodes associated with a clause C can only be reached by f^ξ through a path (s, l, C^1, C^2, t) with l one of the three literals included in C. If $(s, l) \in A_{sl}^\xi$ the use of this path reduces by 1 the recovery cost (since (C^1, C^2) and (C^2, t) are also used in f and (l, C^1) is not), otherwise the recovery cost remains the same. Thus, the recovery cost can be further reduced by at most m if each clause node is reached through a path which includes an arc in A_{sl}^ξ.

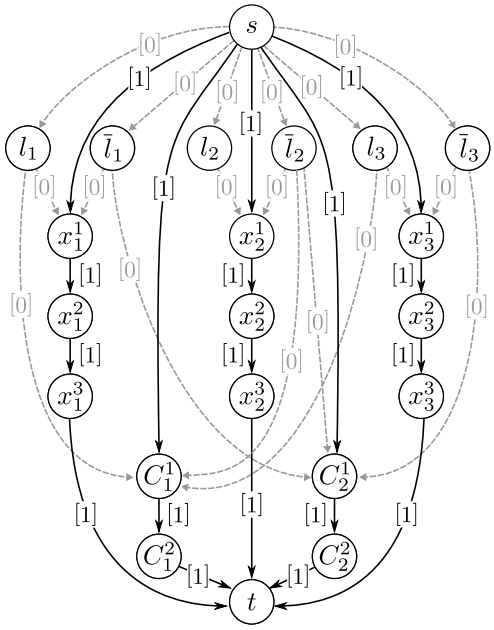
In conclusion, there exists an optimal solution to I_{MF} with recovery cost equal to $3n+2m$ if and only if I_{SAT} is satisfiable. In that case, the solution of I_{SAT} consists in fixing to true the variables x_i such that $f_{s,l_i}^\xi > 0$ and to false the others.

□

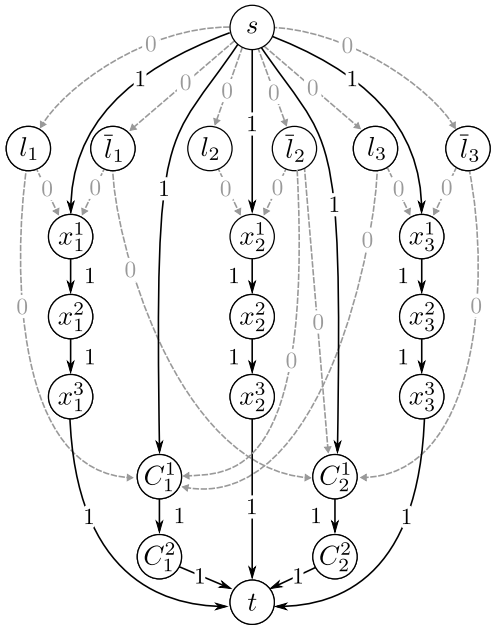
Corollary 4.1. *The reactive max-flow problem with distance d_{struct}, $MF^r_{d_{struct}}$, is strongly $\mathcal{NP}$-hard.*

Proof. The reduction used in the proof of Theorem 4.2 can also be used to prove this theorem since it only requires one scenario. The main difference is that the only feasible nominal solution (exemplified in Figure 4b) is given as an input of the reactive problem.

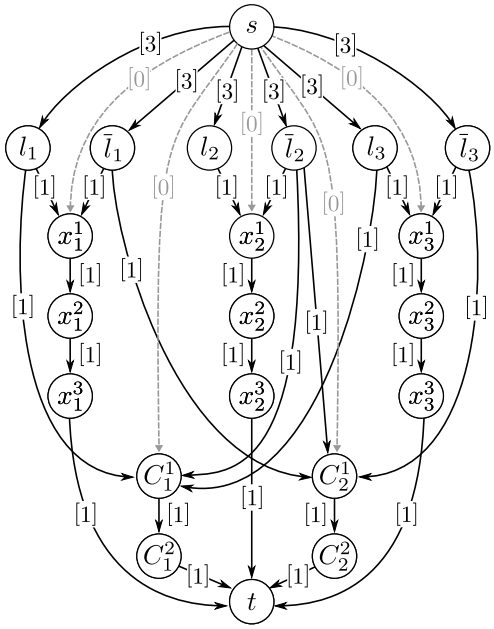
□



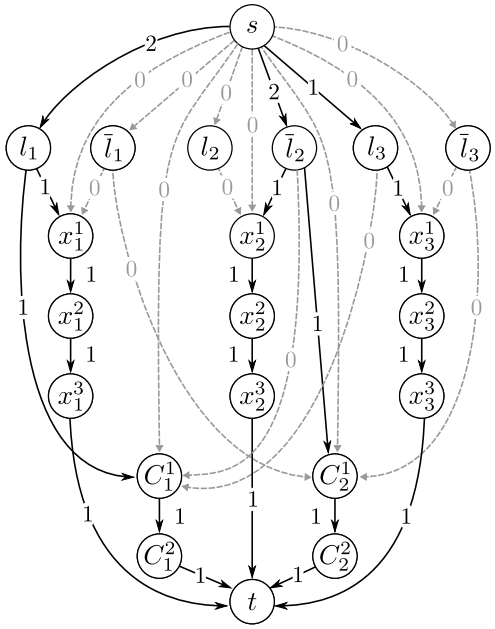
(a) Nominal capacities.



(b) Unique optimal nominal solution with flow value $n + m$.



(c) Capacities of the scenario.



(d) Flow of the scenario of an optimal solution.

Figure 4: Example of the reduction to $MF^p_{d_{struct}}$ of an instance of the 3-SAT problem with three variables and two clauses $C_1 = (x_1 \vee \overline{x}_2 \vee x_3)$ and $C_2 = (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3)$. The arcs with a capacity or a flow of value 0 are colored in gray.

5 A case study - The line optimization Problem with uncertain demands

In this section, we apply our robust approach to a railroad planning problem inspired from [14, 17]. We compare our approach with two approaches from the literature and highlight their similarities and differences. We then assess the recovery cost reduction when relaxing the optimality constraint on the nominal objective of our approach. Finally, we show the efficiency of the proactive approach over the reactive approach.

The line optimization Problem occurs in a railway system with periodic timetables. A line in an urban transportation network is a path from a departure station to an arrival station with stops in intermediary stations. The frequency of a line is the number of times a train has to be operated on this line in a given time interval in order to cover passenger demands. We consider the problem of determining the lines deployment and frequencies in order to minimize the line costs.

5.1 Deterministic problem description

We are given the set $\mathcal{V}$ of stations and the network $\mathcal{N} = (\mathcal{V}, E, d)$ where edges of E represent the undirected links between stations and d_e is the length of $e \in E$. A symmetric origin-destination (OD) matrix $\xi^{nom} \in \mathbb{N}^{|\mathcal{V}|\times|\mathcal{V}|}$ provides the number of passengers that must be able to travel from any station s to any other station s', within one hour.

We consider a set L of potential lines which contains one line ℓ_{ij} for each pair of stations $i, j \in \mathcal{V}$ such that $i < j$ and $\xi^{nom}_{ij} > 0$. A line ℓ is characterised by a pair (s^1_ℓ, s^2_ℓ) of departure and arrival stations together with a shortest path in $\mathcal{N}$ linking the two stations. Trains are planned to run on the deployed lines and all trains are supposed to have the same carriage capacity of C passengers. Therefore, if a line ℓ is scheduled with frequency f, then up to $C \times f$ passengers per hour may be carried on each edge of line ℓ. Due to technical considerations, the frequency of any line is limited by a given value $\text{MF} \in \mathbb{N}$. We are also given the cost K_ℓ of deploying line ℓ and the unitary cost K'_ℓ associated with frequencies on line ℓ.

We introduce a binary variable x_ℓ equal to 1 if and only if line $\ell \in L$ is deployed and 0 otherwise. When ℓ is deployed, variable f_ℓ indicates the hourly frequency of line ℓ. A solution (x, f) is feasible for a given OD matrix ξ if it is included in set $X(\xi)$ defined by the following constraints:

$$X(\xi)\begin{cases} C \times \sum_{\ell\in L: e\in\ell} f_\ell \geq \sum_{\ell\in L: e\in\ell} \xi_{s^1_\ell, s^2_\ell} & e \in E \quad (4)\\ f_\ell \leq \text{MF}\, x_\ell & \ell \in L. \quad (5)\\ x_\ell \leq f_\ell & \ell \in L \quad (6)\\ f_\ell \in \mathbb{N} & \ell \in L\\ x_\ell \in \{0, 1\} & \ell \in L \end{cases}$$

Constraints (4) force the demand satisfaction of each edge. Indeed, their left-hand side corresponds to the sum of the capacities of trains that use edge e. Moreover, since a line is associated to each positive OD value, the right-hand side of these constraints is equal to the number of passengers whose shortest path goes through e. Constraints (5) and (6) link variables x_ℓ and f_ℓ. If line ℓ is not deployed then x_ℓ and f_ℓ are equal to zero, otherwise f_ℓ is not larger than MF.

The nominal objective of a solution (x, f) is defined as

$$\mathcal{N}O(x, f) = \sum_{\ell\in L}(K_\ell x_\ell + K'_\ell f_\ell) \quad (7)$$

The deterministic model is defined as

$$P\begin{cases} \min\ \mathcal{N}O(x, f)\\ \text{s.t. } (x, f) \in X(\xi^{nom}) \end{cases}$$

We now assume that the positive values of the OD matrix are uncertain and that this matrix can take any value within a set U of scenarios. We are given the set $\{\xi^1, ..., \xi^{|U|}\}$ of alternative matrices. To handle the uncertainty, this case study compares three robust approaches: our proactive approach, the anchored approach [9] and the k-distance approach [13]. Each approach either constrains or optimizes differences between the nominal solution and the solution of the scenarios. Our proactive approach minimizes an average distance between the nominal and the scenario solutions. The anchored approach maximizes the

number of variables whose value is identical in all solutions while the k-distance approach requires that the value of at most k variables can differ in the nominal solution and the solution of each scenario.

We consider two cases in which the differences between solutions are either measured on the frequency variables f or on the line deployment variables x. For the proactive approach this can be viewed as considering the distance in values d_{val} over f in the first case and the distance in structure d_{struct} over the lines deployment variables x in the second case.

We show how this problem can be modeled for each approach as a compact integer linear program. The compactness of the formulations is made possible by the fact that the uncertainty is represented by a finite set of scenarios U rather than by a continuous uncertainty set.

5.2 Recovery cost based on frequencies

In this section, the differences between solutions are measured through the frequency variables f^0 of the nominal solution and $\{f^\xi\}_{\xi\in U}$ of the scenario solutions.

Proactive approach

As presented previously, the proactive approach minimizes the average recovery cost between the nominal solution (x^0, f^0) and the scenario solutions $\{(x^\xi, f^\xi)\}_{\xi\in U}$ while setting the nominal objective $\mathcal{NO}(x^0, f^0)$ to its optimal value c^*. The recovery cost based on frequencies corresponds to the distance in values d_{val} from Definition 2.1 applied to the frequency variables f^0 and $\{f^\xi\}_{\xi\in U}$ (i.e., $\sum_{\xi\in U}\sum_{\ell\in L}|f^0_\ell - f^\xi_\ell|$). We consider variable df^ξ_ℓ equal to $|f^\xi_\ell - f^0_\ell|$ for each $\ell \in L$ and $\xi \in U$:

$$P^p_f(U, c^*)\begin{cases} \min & \sum_{\xi\in U}\sum_{\ell\in L} df^\xi_\ell & \\ \text{s.t.} & (x^0, f^0) \in X(\xi^{nom}) & \\ & (x^\xi, f^\xi) \in X(\xi) & \xi \in U \\ & \mathcal{NO}(x^0, f^0) = c^* & \\ & f^\xi_\ell - f^0_\ell \le df^\xi_\ell & \xi \in U,\ \ell \in L \\ & -f^\xi_\ell + f^0_\ell \le df^\xi_\ell & \xi \in U,\ \ell \in L \end{cases}$$

Anchored approach

A nominal variable is said to be *anchored* if its value in the nominal solution and in all scenario solutions are equal. For each line $\ell \in L$, variable a_ℓ is equal to 1 if f^0_ℓ is anchored and 0 otherwise. This approach maximizes the number of anchored variables while ensuring that the nominal objective of the nominal solution is optimal:

$$P^a_f(U, c^*)\begin{cases} \max & \sum_{\ell\in L} a_\ell & \\ \text{s.t.} & (x^0, f^0) \in X(\xi^{nom}) & \\ & (x^\xi, f^\xi) \in X(\xi) & \xi \in U \\ & \mathcal{NO}(x^0, f^0) = c^* & \\ & f^\xi_\ell - f^0_\ell \le \mathrm{MF}(1 - a_\ell) & \xi \in U,\ \ell \in L \\ & -f^\xi_\ell + f^0_\ell \le \mathrm{MF}(1 - a_\ell) & \xi \in U,\ \ell \in L \\ & a_\ell \in \{0, 1\} & \ell \in L \end{cases}$$

In any optimal solution, a_ℓ is equal to 1 unless $(f^\xi_\ell - f^0_\ell)$ is different from 0 for some scenario ξ.

k-distance approach

In this approach, for each scenario $\xi \in U$, the number of variables f^0_ℓ and f^ξ_ℓ which have different values is limited to $k \in \mathbb{N}$ and the nominal objective of the nominal solution x^0 is minimized. Variable δ^ξ_ℓ is equal to 1 if and only if the values of f^0_ℓ and f^ξ_ℓ are different and 0 otherwise.

$$P_f^d(U,k)\begin{cases}\min & \mathcal{NO}(x^0,f^0)\\ \text{s.t.} & (x^0,f^0)\in X(\xi^{nom})\\ & (x^\xi,f^\xi)\in X(\xi) & \xi\in U\\ & f_\ell^\xi - f_\ell^0 \le \text{MF }\delta_\ell^\xi & \xi\in U,\ \ell\in L\\ & -f_\ell^\xi + f_\ell^0 \le \text{MF }\delta_\ell^\xi & \xi\in U,\ \ell\in L\\ & \sum_{\ell\in L}\delta_\ell^\xi \le k & \xi\in U\\ & \delta_\ell^\xi\in\{0,1\} & \xi\in U,\ \ell\in L\end{cases}$$

5.3 Recovery cost based on lines deployment

In this section, the differences between the solutions are measured through the lines deployment variables x^0 of the nominal solution and $\{x^\xi\}_{\xi\in U}$ of the scenario solutions.

Proactive approach

The recovery cost based on lines deployment corresponds to the distance in structure d_{struct} from Definition 2.2 applied to the lines deployment variables x^0 and $\{x^\xi\}_{\xi\in U}$ (i.e., $\sum_{\xi\in U}\sum_{\ell\in L}|\mathbb{1}_{x_\ell^0>0}-\mathbb{1}_{x_\ell^\xi>0}|$). We consider variable dx_ℓ^ξ equal to $|\mathbb{1}_{x_\ell^\xi>0}-\mathbb{1}_{x_\ell^0>0}|$ for each $\ell\in L$ and $\xi\in U$.

$$P_x^p(U,c^*)\begin{cases}\min & \sum_{\xi\in U}\sum_{\ell\in L} dx_\ell^\xi\\ \text{s.t.} & (x^0,f^0)\in X(\xi^{nom})\\ & (x^\xi,f^\xi)\in X(\xi) & \xi\in U\\ & \mathcal{NO}(x^0,f^0)=c^*\\ & x_\ell^\xi - x_\ell^0 \le dx_\ell^\xi & \xi\in U,\ \ell\in L\\ & -x_\ell^\xi + x_\ell^0 \le dx_\ell^\xi & \xi\in U,\ \ell\in L\end{cases}$$

Anchored approach

In this version, the frequency of an anchored line ℓ can vary from one solution to another but if ℓ is opened in the nominal solution x^0, it cannot be closed in the scenario solutions and vice versa. To minimize the number of anchored variables, we introduce for each line $\ell\in L$, a binary variable a_ℓ equal to 1 if and only if x_ℓ^0 is anchored, that is, $x_\ell^\xi = x_\ell^0\ \forall\xi\in U$.

$$P_x^a(U,c^*)\begin{cases}\max & \sum_{\ell\in L} a_\ell\\ \text{s.t.} & (x^0,f^0)\in X(\xi^{nom})\\ & (x^\xi,f^\xi)\in X(\xi) & \xi\in U\\ & \mathcal{NO}(x^0,f^0)=c^*\\ & x_\ell^\xi - x_\ell^0 \le 1-a_\ell & \xi\in U,\ \ell\in L\\ & -x_\ell^\xi + x_\ell^0 \le 1-a_\ell & \xi\in U,\ \ell\in L\\ & a_\ell\in\{0,1\} & \ell\in L\end{cases}$$

k-distance approach

In this k-distance approach, the number of lines which deployment may vary between the nominal solution and the solution of a scenario is bounded by k and the nominal objective is still minimized. Variable δ_ℓ^ξ is equal to 1 if and only if the values of x_ℓ^0 and x_ℓ^ξ are different and 0 otherwise.

$$P_x^d(U,k)\begin{cases}\min & \mathcal{NO}(x^0,f^0)\\ \text{s.t.} & (x^0,f^0)\in X(\xi^{nom})\\ & (x^\xi,f^\xi)\in X(\xi) & \xi\in U\\ & x_\ell^\xi - x_\ell^0 \le \delta_\ell^\xi & \xi\in U,\ \ell\in L\\ & -x_\ell^\xi + x_\ell^0 \le \delta_\ell^\xi & \xi\in U,\ \ell\in L\\ & \sum_{\ell\in L}\delta_\ell^\xi \le k & \xi\in U\\ & \delta_\ell^\xi\in\{0,1\} & \xi\in U,\ \ell\in L\end{cases}$$

5.4 Numerical results

We adapt an instance from the data considered in [14][1] which provides an OD matrix ξ^{nom} and a network $\mathcal{N} = (\mathcal{V}, E, d)$ with $|V| = 23$, $|E| = 31$ and $|L| = 210$.

The maximum frequency MF is set to 6, which corresponds to 1 train on the line every 10 minutes when considering a time interval of 1 hour. The line opening costs $\{K_\ell\}_{\ell \in L}$ are deduced from the costs provided in the instance. A fixed frequency cost $\{K'_\ell\}_{\ell \in L}$ of 100 is considered. The capacity of each train is set to $C = 200$.

We consider a discrete uncertainty set consisting of 10 OD matrices $\{\xi^i\}_{i=1}^{10}$. To create a reasonable deviation from the nominal OD matrix ξ^{nom}, the value ξ^s_{ij} has a 20% chance of being equal to ξ^{nom}_{ij}. Otherwise, it is uniformly drawn from $[0.9 \times \xi^{nom}_{ij}, 1.1 \times \xi^{nom}_{ij}]$. Consequently, $\xi^s_{ij} = 0$ whenever $\xi^{nom}_{ij} = 0$.

The adjacency matrix of the network $\mathcal{N}$, the costs $\{K_\ell\}_{\ell \in L}$, and ξ^{nom} are explicitly given in Appendix B.

5.4.1 Comparison of the anchored and the k-distance approaches with our proactive approach

The results obtained for the robustness in frequencies and the robustness in lines deployment for the considered approaches are presented in Tables 6 and 7, respectively.

In these tables, three instances with the first two, the first six, and all ten scenarios are considered and this number of scenarios is reported in the first column. The k-distance approach requires us to fix the value of parameter $k \in \mathbb{N}$ which represents the maximal number of differences between the nominal solution and any scenario solution. Since there is no a priori relevant choice for this parameter we set its value to 0, 1, 2, 4 and 10. Consequently, seven lines are associated to each instance, one for the proactive approach, one for the anchored approach, and five for the k-distance approach. The proactive and the anchored approaches both ensure that the nominal objective of the nominal solution is equal to its optimal value c^*. This value is obtained by the resolution a priori of the nominal problem P. Note that P is only solved once as $c^* = 81742$ does not depend on the number of scenarios nor the recovery cost.

For each line of Table 6, a nominal solution (x^0, f^0) as well as scenario solutions $\{(x^\xi, f^\xi)\}_{\xi \in U}$ are obtained by solving $P^p_f(U, c^*)$, $P^a_f(U, c^*)$, or $P^k_f(U, k)$. From these solutions, we afterwards compute the distance $d_{val} = \sum_{\xi \in U} \sum_{\ell \in L} |f^\xi_\ell - f^0_\ell|$, the number of anchored lines frequency $\sum_{\ell \in L} \mathbb{1}_{f^0 = f^\xi \, \forall \xi \in U}$ and the nominal objective $\mathcal{NO}(x^0, f^0)$ through (7). The nominal objective is reported in the third column. Since its optimality is imposed by the proactive and the anchored approach, in both cases it is equal to $c^* = 81742$ regardless of the number of scenarios. This objective is larger for the k-distance approach whenever k is small. This highlights a drawback of the k-distance approach which does not allow us to constrain the nominal objective of the nominal solution. In other applications, a low value of k could even lead to infeasibilities. The number of differences in frequencies d_{val} is reported in the fourth column of Table 6. Among all nominal solutions with an optimal value of the nominal objective (81742), the proactive approach returns one whose number of differences is minimal (e.g., 50 for two scenarios). The number of anchored lines frequency is reported in the fifth column. Among all nominal solutions of nominal objective value 81742, the anchored approach returns one which maximizes the number of anchored lines (e.g., 200 over 210 for two scenarios). Table 7 contains similar results. Nominal and scenario solutions are computed by solving $P^p_x(U, c^*)$, $P^a_x(U, c^*)$, or $P^d_x(U, k)$. The fourth column contains the number of differences in lines deployment $d_{struct} = \sum_{\xi \in U} \sum_{\ell \in L} |x^\xi_\ell - x^0_\ell|$ while the fifth column contains the number of lines which are anchored $\sum_{\ell \in L} \mathbb{1}_{x^0 = x^\xi \, \forall \xi \in U}$.

The proactive and the anchored approaches do not lead to the same solutions as the recovery cost is not optimal in the anchored approach and the number of anchored lines is not optimal in the proactive approach. These differences are increased with the number of scenarios as it becomes harder to satisfy both objectives simultaneously. Note that the increase in d_{val} and d_{struct} provided by the anchored approach is often greater than the decrease in number of variables anchored in the proactive approach. This is due to the fact that anchoring a variable is quite constraining since it requires its value to be identical in the nominal solution and in all the scenario solutions. The advantage is that it guarantees that parts of the nominal solution will not be disrupted. However, this comes at a price in terms of flexibility which is reflected by an increase of the recovery costs d_{val} and d_{struct}. In the k-distance approach, an increase of the nominal objective for low values of k enables us to obtain better recovery costs than the proactive approach and more anchored variables than the anchored approach. In particular, when $k = 0$ the scenario solutions are necessarily identical to the nominal solution which leads to a recovery cost of 0 and an anchoring of all the 210 lines. However, we observe a quick deterioration of these two objectives when k increases showing once again that the choice of parameter k can be challenging. Moreover, the

[1] `https://www.gams.com/latest/gamslib_ml/libhtml/gamslib_lop.html`

value k is shared by all scenarios while some of them may require fewer changes than others which can lead to less suitable scenario solutions.

In most cases, resolutions are completed within seconds. However, we notice that the computation time increases with the number of scenarios and even becomes prohibitive for k-distances with distance d_{val}. A resolution time of more than 39 hours is even reached for $k = 2$.

$\|U\|$	**Method**	$\mathcal{NO}(x^0, f^0)$	**Lines frequency differences** d_{val}	**Number of anchored lines**	**Time (s)**
2	Proactive	**81742**	**50**	192 (-4%)	9
	Anchored	81742	78 (+56%)	**200**	18
	0-distance	89028 (+9%)	0 (-100%)	210 (+5%)	0
	1-distance	86663 (+6%)	12 (-76%)	208 (+4%)	7
	2-distance	84740 (+4%)	19 (-62%)	206 (+3%)	8
	4-distance	82156 (+1%)	43 (-14%)	202 (+1%)	7
	10-distance	81742	96 (+92%)	194 (-3%)	3
6	Proactive	**81742**	**120**	185 (-7%)	22
	Anchored	81742	189 (+58%)	**199**	80
	0-distance	91998 (+13%)	0 (-100%)	210 (+6%)	1
	1-distance	87274 (+7%)	27 (-78%)	204 (+3%)	23
	2-distance	85178 (+4%)	56 (-53%)	201 (+1%)	503
	4-distance	82158 (+1%)	122 (+2%)	195 (-2%)	75
	10-distance	81742	248 (+107%)	170 (-15%)	18
10	Proactive	**81742**	**196**	181 (-9%)	25
	Anchored	81742	327 (+67%)	**198**	500
	0-distance	94308 (+15%)	0 (-100%)	210 (+6%)	0
	1-distance	89733 (+10%)	39 (-80%)	202 (+2%)	4654
	2-distance	86428 (+6%)	80 (-59%)	197 (-1%)	141417
	4-distance	83124 (+2%)	153 (-22%)	190 (-4%)	4718
	10-distance	81742	400 (+104%)	158 (-20%)	27

Table 6: Results of each approach when considering the recovery cost based on frequencies. For a given number of scenarios $|U|$, the percentage in a cell corresponds to the relative change between the cell value and the value in bold in the same column.

5.4.2 Flexibility of the nominal objective

Imposing an optimal nominal objective value on the nominal solution of the proactive approach may be too restrictive as small increases of the nominal objective may lead to significant decreases of the recovery cost. As presented in Inequality (3), a parameter $\varepsilon \geq 0$ which corresponds to an acceptable percentage of increase of the nominal objective can be introduced for this purpose. Consequently, the constraint in P_f^p and P_x^p which ensures that the nominal objective is equal to c^* can be replaced by

$$\mathcal{NO}(x^0, f^0) \leq c^*(1 + \varepsilon). \tag{8}$$

Tables (8a) and (8b) present the results obtained for four values of ε (0%, 1%, 2% and 5%) for the robustness in frequencies and the robustness in lines deployment, respectively. In both cases, allowing an increase of the nominal objective enables us to significantly reduce the recovery cost and increase the number of anchored variables. In particular, for ε equal to 5%, the recovery cost is at most equal to 1 and the number of anchored variables is at least equal to 209 over 210 lines.

Similar results have been observed in Tables 6 and 7 with the recoverable robustness approach for small values of k. The major difference with the proactive approach is that it seems easier for a decision maker to choose an acceptable percentage ε of deterioration of the nominal objective rather than a maximal number k of differences allowed in terms of variables values.

$\|U\|$	Method	$\mathcal{NO}(x^0, f^0)$	Lines deployment differences d_{struct}	Number of anchored lines	Time (s)
2	Proactive	**81742**	**10**	201 (>-1%)	4
	Anchored	81742	15 (+50%)	**202**	7
	0-distance	85709 (+5%)	0 (-100%)	210 (+4%)	0
	1-distance	84486 (+3%)	2 (-80%)	208 (+3%)	1
	2-distance	83340 (+2%)	4 (-60%)	206 (+2%)	1
	4-distance	81950 (<1%)	8 (-20%)	203 (<1%)	1
	10-distance	81742	20 (+100%)	193 (-4%)	2
6	Proactive	**81742**	**23**	193 (-4%)	14
	Anchored	81742	41 (+78%)	**202**	1
	0-distance	87598 (+7%)	0 (-100%)	210 (+4%)	4
	1-distance	84774 (+4%)	6 (-74%)	204 (+1%)	1
	2-distance	83621 (+2%)	12 (-48%)	200 (-1%)	4
	4-distance	81952 (<1%)	22 (-4%)	194 (-4%)	2
	10-distance	81742	52 (+126%)	175 (-13%)	2
10	Proactive	**81742**	**37**	191 (-4%)	84
	Anchored	81742	71 (+92%)	**200**	8
	0-distance	88907 (+9%)	0 (-100%)	210 (+5%)	0
	1-distance	85921 (+5%)	10 (-73%)	203 (+2%)	5
	2-distance	84045 (+3%)	18 (-51%)	200	2
	4-distance	82634 (+1%)	37	191 (-4%)	13
	10-distance	81742	87 (+135%)	166 (-17%)	8

Table 7: Results of each approach when considering the recovery cost based on lines deployment. For a given number of scenarios $|U|$, the percentage in a cell corresponds to the relative change between the cell value and the value in bold in the same column.

Once again the resolution times are generally small. The only exception is observed for distance d_{val} for 10 scenarios and epsilon equal to 5% which is solved in almost 3 hours. This could be due to a sudden increase in the size of the feasible solution set.

5.5 Proactive and reactive approaches comparison

The reactive approach is considered when the uncertainty has not been anticipated and the nominal solution (x^0, f^0) is infeasible for a scenario s. In that context, the reactive problem P^r provides a reactive solution (x^r, f^r) feasible for s and whose recovery cost with (x^0, f^0) is minimal. Problem P^r is similar to the proactive problem P^p except that only one scenario $\xi \in U$ is considered and that the nominal solution (x^0, f^0) is given as an input rather than an output. The reactive line optimization problem associated with the recovery cost based on frequencies can be modelled by considering a binary variable d_ℓ equal to 1 if and only if f^r_ℓ is equal to f^0_ℓ for each $\ell \in L$:

$$P^r(\xi, x^0, f^0) \begin{cases} \min & \sum_{\ell \in L} d_\ell \\ \text{s.t.} & (x^r, f^r) \in X(\xi) \\ & f^r_\ell - f^0_\ell \leq d_\ell \quad \ell \in L \\ & -f^r_\ell + f^0_\ell \leq d_\ell \ell \in L \end{cases}$$

We do not present the reactive model for the recovery cost based on lines deployment as it only requires in the last two sets of constraints replacing f^r_ℓ and f^0_ℓ by x^r_ℓ and x^0_ℓ, respectively.

Let Q^* be the set of nominal solutions with an optimal nominal objective c^*. The aim of our proactive approach can be viewed as finding a solution in Q^* which minimizes the sum of the recovery costs over the scenarios in set U. To assess the efficiency of the proactive solution, we compare its recovery cost to

$\|U\|$	ε	$\mathcal{NO}(x^0, f^0)$	**Lines frequency differences** d_{val}	**Time (s)**
2	0%	81742	50	9
	1%	82558	29	2
	2%	83348	24	10
	5%	85797	11	10
6	0%	81742	120	22
	1%	82558	71	52
	2%	83338	55	12
	5%	85802	32	35
10	0%	81742	196	25
	1%	82558	121	140
	2%	83338	93	61
	5%	85730	59	9978

(a) Recovery cost based on the frequencies.

$\|U\|$	ε	$\mathcal{NO}(x^0, f^0)$	**Lines deployment differences** d_{struct}	**Time (s)**
2	0%	81742	10	4
	1%	82450	6	1
	2%	83340	4	1
	5%	85812	0	0
6	0%	81742	23	14
	1%	82349	16	3
	2%	83374	11	9
	5%	85809	2	1
10	0%	81742	37	84
	1%	82449	25	12
	2%	83166	19	47
	5%	85734	6	2

(b) Recovery cost based on the lines deployment.

Table 8: Influence of parameter ε of our proactive approach on the recovery cost.

the ones of four other solutions from Q^*. These four solutions have been selected to be as different as possible in order to be representative of Q^* (see Appendix A for more details on the selection process).

Tables 9a and 9b present the results obtained when considering the recovery cost based on frequencies and lines deployment, respectively. For each considered nominal solution (x, f), the recovery cost $v(P^r(\xi, x, f))$ of any scenario $\xi \in U$ is obtained by solving the reactive problem $P^r(\xi, x, f)$. The sum of the recovery costs over all scenarios in U is then obtained by computing $\sum_{\xi \in U} v(P^r(\xi, x, f))$ and this value is represented in the third column of both tables. The proactive solution necessarily returns the optimal recovery cost and the others lead to a mean increase of 8% with a maximal increase of 24%. This significant variability of the recovery cost among solutions from Q^* highlights the relevance of considering the proactive approach. In Table 9a, the first reactive solution is very similar to the proactive solution as its recovery cost is identical for 2 and 6 scenarios and it is only incremented for 10 scenarios. In Table 9b one reactive solution always leads to an optimal recovery cost but it is not always the same depending on the number of scenarios. This shows that the choice of the scenarios is a sensitive task that may significantly impact the proactive solution.

The resolution times of the proactive problems are significantly higher. This was expected since the reactive problem contains the variables of only one solution while the proactive problem contains $|U| + 1$.

$\|U\|$	**Solution**	**Lines frequency differences** d_{val}	**Time (s)**
2	Proactive	**50**	9
	Reactive (x^A, f^A)	50	0
	Reactive (x^B, f^B)	57 (+14%)	0
	Reactive (x^C, f^C)	62 (+24%)	0
	Reactive (x^D, f^D)	56 (+12%)	0
6	Proactive	**120**	22
	Reactive (x^A, f^A)	120	0
	Reactive (x^B, f^B)	123 (+2%)	0
	Reactive (x^C, f^C)	133 (+11%)	0
	Reactive (x^D, f^D)	121 (+1%)	0
10	Proactive	**196**	25
	Reactive (x^A, f^A)	197 (+1%)	0
	Reactive (x^B, f^B)	211 (+8%)	0
	Reactive (x^C, f^C)	220 (+12%)	0
	Reactive (x^D, f^D)	204 (+4%)	0

(a) Recovery cost based on frequencies.

$\|U\|$	**Solution**	**Lines deployment differences** d_{struct}	**Time (s)**
2	Proactive	**10**	4
	Reactive (x^A, f^A)	10	0
	Reactive (x^B, f^B)	12 (+20%)	0
	Reactive (x^C, f^C)	11 (+10%)	0
	Reactive (x^D, f^D)	11 (+10%)	0
6	Proactive	**23**	14
	Reactive (x^A, f^A)	26 (+13%)	0
	Reactive (x^B, f^B)	28 (+22%)	0
	Reactive (x^C, f^C)	24 (+4%)	0
	Reactive (x^D, f^D)	23	0
10	Proactive	**37**	84
	Reactive (x^A, f^A)	38 (+3%)	0
	Reactive (x^B, f^B)	43 (+16%)	0
	Reactive (x^C, f^C)	39 (+5%)	0
	Reactive (x^D, f^D)	37	0

(b) Recovery cost based on lines deployment.

Table 9: Comparison of the proactive and the reactive approaches with four different optimal nominal solutions (x^A, f^A), (x^B, f^B), (x^C, f^C) and (x^D, f^D). For each solution (x^i, y^i) and each scenario $\xi \in U$ the reactive problem $P^r(\xi, x^i, y^i)$ is solved. The values in the table correspond to the sum of the recovery costs obtained over all scenarios. For a given number of scenarios $|U|$, the percentage in a cell corresponds to the relative change between the cell value and the value in bold.

6 A min-max robustness version of the proactive problem

In this section, we introduce a version P^w of the proactive problem P^p which is closer to classical robust approaches. Instead of minimizing the weighted sum of the recovery costs over all scenarios, we minimize the recovery cost of the worst-case scenario:

$$P^w(U, c^*)\begin{cases} \min\limits_{x^p \in X(\xi^{nom})} \max\limits_{\xi_i \in U} \min\limits_{x^i \in X(\xi_i)} d(x^p, x^i). & (9) \\ \text{s.t. } f(x^p) = c^* & (10) \end{cases}$$

For a given distance d, we denote by MCF_d^w and MF_d^w the proactive problems MCF_d^p and MF_d^p in which the sum over the scenarios in the objective is replaced by a maximum.

In this section, we study the complexity of $MCF_{d_{val}}^w$, $MCF_{d_{struct}}^w$, $MF_{d_{val}}^w$ and $MF_{d_{struct}}^w$. Then, we show conditions on distance d and function $X(\xi)$ under which the set of extreme points of the convex hull $conv(U)$ of a discrete uncertainty set U always contain a worst-case scenario. Finally, we prove that this result is no longer true for three other classes deduced from the first one: when the variables must be integer, when distance d is non-convex or when the uncertainty is on the left-hand side of the constraints rather than their right-hand side.

6.1 Complexity results

As summarized in the following two propositions, we prove that the complexity results obtained previously hold for this version of the proactive problem.

Proposition 6.1. *$MCF_{d_{struct}}^w$ and $MF_{d_{struct}}^w$ are strongly $\mathcal{NP}$-hard.*

Proof. When only one scenario ξ is considered, the proactive problems P^p and P^w are equivalent since the recovery cost of ξ is both equal to the sum and to the maximum of the recovery costs over all scenarios. The proofs of Theorems 3.3 and 4.2 only consider one scenario. Consequently, these proofs also provide valid reductions from problem 3-SAT to problems $MCF_{d_{struct}}^w$ and $MF_{d_{struct}}^w$. □

Proposition 6.2. *$MCF_{d_{val}}^w$ and $MF_{d_{val}}^w$ are strongly $\mathcal{NP}$-hard.*

Proof. In the proof of Theorem 3.1, for any instance I_{SAT} of the 3-SAT problem with n boolean variables and m clauses, we construct an instance I_{MCF} for the proactive problem $MCF_{d_{val}}^p$. We prove that all I_{SAT} clauses are satisfied if and only if the recovery cost of each scenario is equal to its lowest possible value $4n$. Thus, we finally conclude that I_{SAT} is a yes-instance if and only the optimal value of $MCF_{d_{val}}^p$ is at most $4nm$.

The same reduction can be considered to prove the $\mathcal{NP}$-hardness of $MCF_{d_{val}}^w$. Since $4n$ is the lowest possible recovery cost of a scenario, the maximal recovery cost over all scenarios is equal to $4n$ if and only if the recovery cost of all scenarios is equal to $4n$. Consequently, I_{SAT} is a yes-instance if and only if the objective value of $MCF_{d_{val}}^w$ is at most $4n$.

A similar reasoning can prove the $\mathcal{NP}$-hardness of $MF_{d_{val}}^w$ from the proof of Theorem 4.1. □

6.2 Equivalence of optimization over the discrete set of scenarios set and its convex hull

For many robust optimization problems it has been proved that considering a discrete set of uncertain scenarios is equivalent to considering its convex hull (i.e., $v(P(U)) = v(P(conv(U)))$ [4, 11, 29]. Such results can make solving $P(conv(U))$ easier by restricting the set of scenarios considered during the resolution.

We first prove that this result holds for a class $\mathcal{P}_0^w$ of proactive problems. Let $\mathcal{P}_0^w$ be the set of proactive problems P^w in which the distance d is convex, each scenario $\xi \in U$ is a vector of $\mathbb{R}^m$, and its feasibility set is $X(\xi) = \{x \in \mathbb{R}^n \mid h(x) \leq \xi\}$, with $h : \mathbb{R}^n \to \mathbb{R}^m$ a convex function. The class $\mathcal{P}_0^w$ for example includes the proactive problems which consider distance d_{val} and linear constraints with uncertain right-hand sides. Note that this class does not include $MF_{d_{val}}^w$ and $MCF_{d_{val}}^w$ as the variables can be fractional.

Subsequently, we show that this result does not hold if the variables are imposed to be integers, if the distance is non-convex or if the uncertainty is on the left-hand side of the linear constraints.

Proposition 6.3. *Let $\mathcal{P}_0^w$ be the set of proactive problems P^w in which (i) the distance d is convex, and (ii) the feasibility set of any scenario $\xi \in \mathbb{R}^m$ is a convex set defined as $X(\xi) = \{x \in \mathbb{R}^n \mid h(x) \leq \xi\}$, where $h : \mathbb{R}^n \to \mathbb{R}^m$ is a convex function. It holds that $v(P_0^w(U, c^*)) = v(P_0^w(conv(U), c^*))$ for all $P_w^0 \in \mathcal{P}_w^0$ and all discrete uncertainty sets U.*

Proof. Let x^p be a nominal solution and $g_{x^p} : \xi \mapsto \min_{x \in X(\xi)} d(x^p, x)$ be the function which associates to any scenario $\xi \in conv(U)$ the distance between x^p and its closest solution in $X(\xi)$. So P_0^w can be re-written as

$$P_0^w(U, c^*) \begin{cases} \min\limits_{x^p \in X(\xi^{nom})} \max\limits_{\xi_i \in U} g_{x^p}(\xi_i). \\ \text{s.t. } f(x^p) = c^* \end{cases}$$

Let ξ_1, ξ_2 and ξ' be three scenarios in $conv(U)$ such that $\xi' = t\,\xi_1 + (1-t)\xi_2$ with $t \in [0, 1]$. To prove that g_{x^p} is convex, it is sufficient to show that there exists $x' \in X(\xi')$ such that

$$d(x^p, x') \leq t\, g_{x^p}(\xi_1) + (1-t) g_{x^p}(\xi_2). \tag{11}$$

Let $x_1^* = argmin_{x \in X(\xi_1)} d(x^p, x)$ (i.e., $g_{x^p}(\xi_1) = d(x^p, x_1^*)$) and $x_2^* = argmin_{x \in X(\xi_2)} d(x^p, x)$. We take $x' = t\, x_1^* + (1-t)x_2^*$. We first prove that x' is in $X(\xi')$ and then that it satisfies Inequality (11).

The convexity of h ensures that:

$$h(x') \leq t\, h(x_1^*) + (1-t)h(x_2^*) \leq t\, \xi_1 + (1-t)\xi_2 = \xi'$$

and therefore x' is in $X(\xi')$. Finally, Inequality (11) follows from the convexity of d.

Since $conv(U)$ is a compact set, the convexity of g_{x_p} ensures that it reaches a maximum on one of its extreme points which are all included in U.

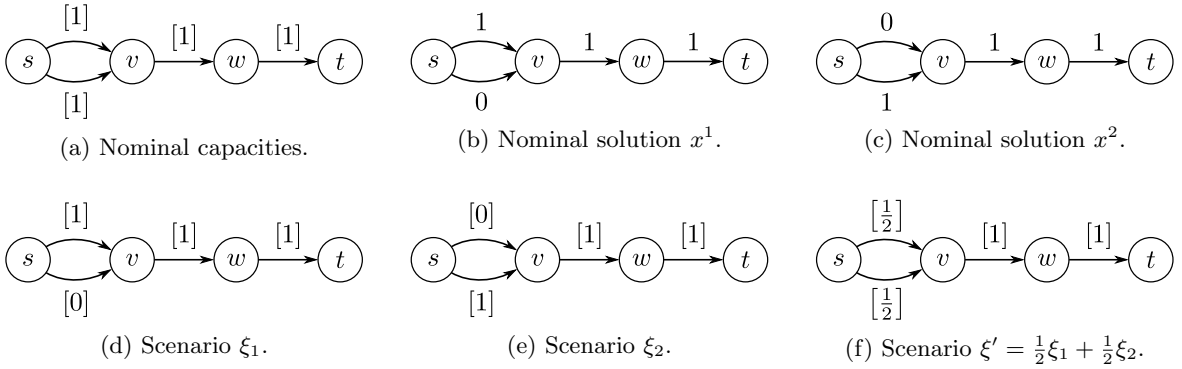
□

We now define three classes of min-max proactive problems P^w: (i) $\mathcal{P}_1^w$ is the same as $\mathcal{P}_0^w$ but solutions x are integers, (ii) $\mathcal{P}_2^w$ is the same as $\mathcal{P}_0^w$ but the distance d is no longer convex, and (iii) $\mathcal{P}_3^w$ is the same as $\mathcal{P}_0^w$ but the the feasibility sets are polyhedra with uncertain constraint matrix. More precisely, in $\mathcal{P}_3^w$, the scenarios are matrices in $\mathbb{R}^{n \times m}$ and the feasibility sets are $X : \xi \mapsto \{x \in \mathbb{R}^n \mid \xi x \leq b\}$ with $b \in \mathbb{R}^m$.

For each of these classes, we prove that Property 6.3 does not hold.

Proposition 6.4. *There exists an instance $P_1^w \in \mathcal{P}_1^w$ and a discrete uncertainty set U, such that $v(P_1^w(U, c^*)) < v(P_1^w(conv(U), c^*))$.*

Proof. We consider the nominal integer max-flow problem represented in Figure 5a. This problem has an optimal flow value c^* of 1 which can only be obtained by the two solutions x^1 and x^2 represented in Figures 5b and 5c, respectively.



(a) Nominal capacities. (b) Nominal solution x^1. (c) Nominal solution x^2.

(d) Scenario ξ_1. (e) Scenario ξ_2. (f) Scenario $\xi' = \frac{1}{2}\xi_1 + \frac{1}{2}\xi_2$.

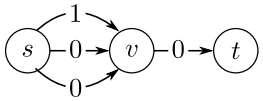
Figure 5: An instance of $MCF_{d_{val}}^w$ to illustrate Proposition 6.4

Let us consider a discrete uncertainty set $U = \{\xi_1, \xi_2\}$ whose two scenarios are represented in Figures 5d and 5e, respectively. The corresponding proactive problem is in $\mathcal{P}_1^w$ since its solutions must be integers, distance d_{val} is convex, each scenario $\xi \in U$ is a vector in $\mathbb{R}^m$ of uncertain capacities and the feasibility set $X(\xi)$ can be expressed as linear constraints with uncertain right-hand sides.

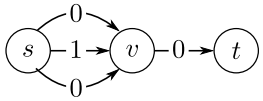
The only feasible integer flows for scenario ξ_1 are x^1 and the empty flow x^0. Similarly, x^2 and x^0 are the only feasible solutions for scenario ξ_2. Consequently, we deduce that

$$v(MF_{d_{val}}^w(U, 1)) = \min_{x^p \in \{x^1, x^2\}} \max \Big(\underbrace{\min_{x^{\xi_1} \in \{x^0, x^1\}} d_{val}(x^p, x^{\xi_1})}_{\text{Recovery cost of } \xi_1}, \underbrace{\min_{x^{\xi_2} \in \{x^0, x^2\}} d_{val}(x^p, x^{\xi_2})}_{\text{Recovery cost of } \xi_2} \Big) = 2.$$

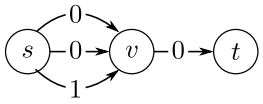
Indeed, if the proactive solution x^p is x^1, then the maximal recovery cost is equal to 2 and it is obtained for scenario ξ_2 with solution x^2. A symmetric reasoning with scenario ξ_1 leads to the same recovery cost if x^p is x^2.



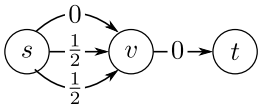
(a) Node and nominal arc demands.



(b) Demands of scenario ξ_1.



(c) Demands of scenario ξ_2.



(d) Demands of scenario $\frac{1}{2}\xi_1 + \frac{1}{2}\xi_2$.

Figure 6: Demands of an instance of $MCF^w_{d_{struct}}$.

We now consider scenario $\xi' = \frac{1}{2}\xi_1 + \frac{1}{2}\xi_2$ in $conv(U)$ represented in Figure 5f. The only feasible integer flow for ξ' is the empty flow x^0 and hence its recovery cost is 3.

□

We now prove that the same result holds if we consider continuous variables and a non-convex distance d.

Proposition 6.5. *There exists an instance $P_2^w \in \mathcal{P}_2^w$ and a discrete uncertainty set U, such that $v(P_2^w(U, c^*)) < v(P_2^w(conv(U), c^*))$.*

Proof. We consider a variation of the robust min-cost flow problem $MCF^w_{d_{struct}}$ in which the flows are not required to be integers. Such a problem is in $\mathcal{P}_2^w$ since the distance d_{struct} is non-convex and the feasibility set $X(\xi)$ can be expressed as linear constraints with uncertain right-hand side.

We consider an instance whose node and nominal arc demands are represented in Figure 6a and whose arcs capacities $\{u_a\}_{a\in A}$ and arcs unitary costs $\{c_a\}_{a\in A}$ are all equal to 1. This instance contains two scenarios $U = \{\xi_1, \xi_2\}$ represented in Figures 6b and 6c.

There exists only one feasible nominal solution x^p of cost $c^* = 2$. Similarly, scenarios ξ_1 and ξ_2 only each admit one feasible solution whose distance d_{struct} with x^p is 2. However, the only feasible solution of scenario $\frac{1}{2}\xi_1 + \frac{1}{2}\xi_2 \in conv(U)$, represented in Figure 6d, leads to a distance d_{struct} of 3.

□

Eventually, we prove that if we consider uncertainty in matrix coefficients, then Property 6.3 does not hold either.

Proposition 6.6. *There exists an instance $P_3^w \in \mathcal{P}_3^w$ and a discrete uncertainty set U, such that $v(P_3^w(U, c^*)) < v(P_3^w(conv(U), c^*))$.*

Proof. Let us consider the feasibility set $X(\xi^{nom}) = \{x \in \mathbb{R}^2 \mid \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} x \leq \begin{pmatrix} -1 \\ -1 \end{pmatrix}\}$. The nominal problem:

$$\begin{cases} \min & x_1 + x_2 \\ \text{s.t.} & x \in X(\xi^{nom}) \end{cases}$$

only has one optimal solution $x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ of value $c^* = 2$.

We consider the uncertainty set $U = \{\xi_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \xi_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\}$.

Since the only possible solution of the proactive problem is x^*, the recovery cost is:

- $\min_{x\in\mathbb{R}^2:\, x_1 \leq -1} d_{val}(x^*, x) = 2$ for scenario ξ_1, which is obtained by solution $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$;
- $\min_{x\in\mathbb{R}^2:\, x_2 \leq -1} d_{val}(x^*, x) = 2$ for scenario ξ_2, which obtained by solution $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$;
- $\min_{x\in X(\frac{1}{2}\xi_1 + \frac{1}{2}\xi_2)} d_{val}(x^*, x) = 6$ for scenario $\frac{1}{2}\xi_1 + \frac{1}{2}\xi_2$, which is obtained by solution $\begin{pmatrix} -2 \\ -2 \end{pmatrix}$.

□

7 Conclusion

We introduced a new robust approach which optimizes the recovery cost by minimizing the recovery cost over a discrete set of scenarios while ensuring the optimality of the nominal objective. We proved that the proactive counterparts of two polynomial network flow problems are $\mathcal{NP}$-hard and that their reactive counterparts are polynomial for d_{val} and $\mathcal{NP}$-hard for d_{struct}.

We show in a case study of a railroad planning problem that a proactive solution can significantly reduce the recovery cost compared to other solutions with an optimal nominal objective. Relaxing the optimality constraint on the nominal objective can also further reduce the recovery cost. Unlike the k-distance approach, the proactive approach does not require the definition of a parameter k which may prove difficult to fix. We also observed that the anchored approach tends to increase more significantly the recovery cost than the proactive approach decreases the number of anchored lines.

In future works it would be interesting to study the complexity of other problems or distances in this framework all the more so if it enables the identification of polynomial proactive integer problems. The discrete set of scenarios U could also be replaced by classical sets such as box, budgeted or polytope uncertainty sets. This may lead to more challenging problems from the numerical viewpoint as it would no longer be possible to define a compact formulation which associates a set of variables to each possible scenario. Finally, rather than imposing a bound on the nominal objective, the proactive problem could be solved as a bi-objective problem in which both the recovery cost and the nominal objective are minimized.

Acknowledgement

The beginnings of this work originate from discussions on Rémi Lucas's PhD thesis [24]. The authors thank François Ramond, Rémi Chevrier and Rémi Lucas for the fruitful collaboration during this thesis.

Appendix

A Nominal solutions generation

In Section 5.5 we consider four solutions from Q^*, the set of solutions which have an optimal nominal objective c^*. We show that their recovery cost vary significantly. We detail in this section how these solutions are obtained.

In order to be as representative as possible of Q^*, we want to generate iteratively solutions which are as distant from one another as possible (i.e., which have a recovery cost as high as possible with one another). More formally, if the recovery cost is defined by a distance d and if $Q \subset Q^*$ is the set of solutions already obtained, the solution x of the next iteration is one whose distance is farthest from its nearest solution in Q: $x \in argmax_{x \in Q^*} \min_{x^q \in Q} d(x, x^q)$.

This problem can be solved through a MILP for the line optimization problem when considering any of the two distances presented in Section 5.1. Since both models are similar we only present that for distance d_{struct} over the lines deployment variables x. In this model, variable z represents the objective value while variables x and f correspond to the new solution. Finally, for each solution $x^q \in Q$ and each line $\ell \in L$, the binary variable dx^q_ℓ is equal to 1 if and only if x_ℓ is equal to x^q_ℓ.

$$
P_x(Q, c^*) \begin{cases} \max & z & & \\ \text{s.t.} & z \leq \sum_{\ell \in L} dx^q_\ell & x^q \in Q & (12) \\ & (x, f) \in X(\xi^{nom}) & & (13) \\ & \mathcal{N}O(x, f) = c^* & & (14) \\ & x^q_\ell - x_\ell \leq dx^q_\ell & x^q \in Q,\ \ell \in L & (15) \\ & -x^q_\ell + x_\ell \leq dx^q_\ell & x^q \in Q,\ \ell \in L & (16) \end{cases}
$$

Constraints (12) ensure that the objective is at most the recovery cost between x and any $x^q \in Q$. Constraints (13) and (14) guarantee that $(x, f) \in Q^*$. Finally, the link between variables x and dx^q is ensured through constraints (15) and (16).

B Case study instance

This section gives more details on the case study instance considered in Section 5 which is inspired from [14, 17]. Table 10 contains the adjacency matrix and Table 11 the line deployment costs $\{K_\ell\}_{\ell \in L}$. The nominal OD matrix ξ^{nom} is presented in Table 12. The OD matrices of the scenarios are not presented for the sake of conciseness.

References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, *Network flows*, Cambridge, Mass.: Alfred P. Sloan School of Management, Massachusetts, 1988.

[2] H. Aissi, C. Bazgan, and D. Vanderpooten, *Min–max and min–max regret versions of combinatorial optimization problems: A survey*, Eur. J. Oper. Res. **197** (2009), 427–438.

[3] A. Atamtürk and M. Zhang, *Two-stage robust network flow and design under demand uncertainty*, Oper. Res. **55** (2007), 662–673.

[4] J. Ayoub and M. Poss, *Decomposition for adjustable robust linear optimization subject to uncertainty polytope*, Comput. Manage. Sci. **13** (2016), 219–239.

[5] W. Ben-Ameur and H. Kerivin, *New economical virtual private networks*, Commun. ACM **46** (2003), 69–69.

[6] A. Ben-Tal, B. Golany, A. Nemirovski, and J.P. Vial, *Retailer-supplier flexible commitments contracts: A robust optimization approach*, Manufacturing Service Oper. Manage. **7** (2005), 248–271.

[7] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, *Adjustable robust solutions of uncertain linear programs*, Math. Program. **99** (2004), 351–376.

[8] A. Ben-Tal and A. Nemirovski, *Robust convex optimization*, Math. Oper. Res. **23** (1998), 769–805.

	Ah	Apd	Asd	Asdz	Asn	Bd	Ehv	Gn	Gv	Gvc	Hgl	Hr	Lls	Lw	Mt	Odzg	Rsdg	Rtd	Shl	Std	Ut	Zl	Zvg
Ah																					1		1
Apd			1								1										1		
Asd		1											1						1		1		
Asdz													1						1		1		
Asn								1														1	
Bd																		1			1		
Ehv																				1	1		
Gn					1																		
Gv										1								1	1				
Gvc									1										1		1		
Hgl		1														1							
Hr														1								1	
Lls			1	1																		1	
Lw												1											
Mt																				1			
Odzg											1												
Rsdg																		1					
Rtd						1			1								1				1		
Shl			1	1					1	1													
Std							1								1								
Ut	1	1	1	1		1	1			1								1				1	
Zl					1							1	1								1		
Zvg	1																						

Table 10: Adjacency matrix.

	Ah	Apd	Asd	Asdz	Asn	Bd	Ehv	Gn	Gv	Gvc	Hgl	Hr	Lls	Lw	Mt	Odzg	Rsdg	Rtd	Shl	Std	Ut	Zl	Zvg
Ah			923	892	1908	1253	1185		1066	1060		1727	1241	2014			1147	1035	1004		574	1210	224
Apd			767	929		1397	1222		1104	1097	642				2051	754	1185	1072	1041	1920	611		1197
Asd	923	767			1559	1135	1066	1733	599	599	1303	1484	549	1665	1896	1521	960	742	224	1659	455	966	1041
Asdz	892	929			1571	1147	1035	1746	592	592	1465	1496	561	1677	1864		954	736	218	1627	424	979	1010
Asn	1908		1559	1571		2120	2051	281	1933	1927			1116		2881		2014	1902	1684	2644	1440	698	2026
Bd	1253	1397	1135	1147	2120		1503	2401	661	667	1933	2045	1602	2332	2226	2045	630	517	1035	2095	892	1528	1478
Ehv	1185	1222	1066	1035	2051	1503		2226	1210	1203	1758	1977	1490	2157	829	1977	1291	1178	1147	698	717	1459	1303
Gn			1733	1746	281	2401	2226		2108	2101			1291		3055		2189	2076	1964	2818	1615	979	2201
Gv	1066	1104	599	592	1933	661	1210	2108		112	1640	1858	1047	2039	2039		362	249	480	1802	599	1235	
Gvc	1060	1097	599	592	1927	667	1203	2101	112		1634	1746	1047	2033	2033	1746	368	256	480	1796	592	1228	1178
Hgl		642	1303	1465		1933	1758		1640	1634					2588	218	1721	1609	1578	2457	1147		
Hr	1727		1484	1496		2045	1977		1858	1746			1041	287	2700		1939	1721	1609	2569	1260	623	1952
Lls	1241		549	561	1116	1602	1490	1291	1047	1047		1041		1222	2214		1303	1191	673	2083	773	524	
Lw	2014		1665	1677		2332	2157		2039	2033		287	1222		2987		2120	2008	1896	2750	1546	910	2133
Mt		2051	1896	1864	2881	2226	829	3055	2039	2033	2588	2700	2214	2987			2120	2008	1977	237	1546	2182	
Odzg		754	1521			2045	1977			1746	218							1721			1260		
Rsdg	1147	1185	960	954	2014	630	1291	2189	362	368	1721	1939	1303	2120	2120			218	736	1989	680	1316	
Rtd	1035	1072	742	736	1902	517	1178	2076	249	256	1609	1721	1191	2008	2008	1721	218		623	1771	567	1203	1154
Shl	1004	1041	224	218	1684	1035	1147	1964	480	480	1578	1609	673	1896	1977		736	623		1740	536	1091	
Std		1920	1659	1627	2644	2095	698	2818	1802	1796	2457	2569	2083	2750	237		1989	1771	1740		1309	2051	
Ut	574	611	455	424	1440	892	717	1615	599	592	1147	1260	773	1546	1546	1260	680	567	536	1309		742	692
Zl	1210		966	979	698	1528	1459	979	1235	1228		623	524	910	2182		1316	1203	1091	2051	742		1434
Zvg	224	1197	1041	1010	2026	1478	1303	2201		1178		1952		2133				1154			692	1434	

Table 11: Line deployment costs $\{K'_\ell\}_{\ell \in L}$.

	Ah	Apd	Asd	Asdz	Asn	Bd	Ehv	Gn	Gv	Gvc	Hgl	Hr	Lls	Lw	Mt	Odzg	Rsdg	Rtd	Shl	Std	Ut	Zl	Zvg
Ah			726	721	16	136	101		109	741		11	8	13			320	180	60		4244	56	602
Apd			917	76		57	62		21	202	468				5	71	47	143	32	10	1160		83
Asd	726	917			235	154	437	205	730	2540	287	115	2783	90	37	22	2258	1756	6469	155	4919	400	489
Asdz	721	76			58	138	542	88	461	207	24	16	819	21	149		6	369	1456	203	5826	171	155
Asn	16		235	58		42	33	1720	13	117			152		14		48	125	32	28	502	854	19
Bd	136	57	154	138	42		950	34	531	228	39	16	6	14	79	5	329	1829	7	157	1165	79	14
Ehv	101	62	437	542	33	950		28	35	335	28	11	8	9	404	3	75	569	99	936	3109	47	11
Gn			205	88	1720	34	28		12	73			200		13		33	75	48	29	331	720	14
Gv	109	21	730	461	13	531	35	12		785	20	10	29	6	8		890	4586	1339	22	225	33	
Gvc	741	202	2540	207	117	228	335	73	785		81	48	31	26	41	7	3	2829	1503	104	3138	163	229
Hgl		468	287	24		39	28		20	81					12	75	24	52	11	20	422		
Hr	11		115	16		16	11		10	48			77	478	4		10	58	20	8	209	511	19
Lls	8		2783	819	152	6	8	200	29	31		77		77	1		15	46	103	2	89	390	
Lw	13		90	21		14	9		6	26		478	77		4		7	36	20	9	145	380	14
Mt		5	37	149	14	79	404	13	8	41	12	4	1	4			22	73	29	863	359	21	
Odzg		71	22			5	3			7	75							11			21		
Rsdg	320	47	2258	6	48	329	75	33	890	3	24	10	15	7	22			1077	164	19	325	100	
Rtd	180	143	1756	369	125	1829	569	75	4586	2829	52	58	46	36	73	11	1077		509	179	2260	182	157
Shl	60	32	6469	1456	32	7	99	48	1339	1503	11	20	103	20	29		164	509		44	278	64	
Std		10	155	203	28	157	936	29	22	104	20	8	2	9	863		19	179	44		720	46	
Ut	4244	1160	4919	5826	502	1165	3109	331	225	3138	422	209	89	145	359	21	325	2260	278	720		1112	996
Zl	56		400	171	854	79	47	720	33	163		511	390	380	21		100	182	64	46	1112		32
Zvg	602	83	489	155	19	14	11	14		229		19		14				157			996	32	

Table 12: Nominal OD matrix ξ^{nom}.

[9] P. Bendotti, P. Chrétienne, P. Fouilhoux, and A. Pass-Lanneau, *The anchor-robust project scheduling problem*, Oper. Res. (2022).

[10] D. Bertsimas, E. Nasrabadi, and S. Stiller, *Robust and adaptive network flows*, Oper. Res. **61** (2013), 1218–1242.

[11] D. Bertsimas and M. Sim, *The price of robustness*, Oper. Res. **52** (2004), 35–53.

[12] C. Buchheim and J. Kurtz, *Robust combinatorial optimization under convex and discrete cost uncertainty*, EURO J. Comput. Optim. **6** (2018), 211–238.

[13] C. Büsing, *Recoverable robust shortest path problems*, Networks **59** (2012), 181–189.

[14] M.R. Bussieck, P. Kreuzer, and U.T. Zimmermann, *Optimal lines for railway systems*, Eur. J. Oper. Res. **96** (1997), 54–63.

[15] M.L. Ginsberg, A.J. Parkes, and A. Roy, *Supermodels and robustness*, AAAI/IAAI, 1998, pp. 334–339.

[16] M. Goerigk and A. Schöbel, *Recovery-to-optimality: A new two-stage approach to robustness with an application to aperiodic timetabling*, Comput. Oper. Res. **52** (2014), 1–15.

[17] J.W. Goossens, S. Van Hoesel, and L. Kroon, *A branch-and-cut approach for solving railway line-planning problems*, Transp. Sci. **38** (2004), 379–393.

[18] A. Gupta, J. Kleinberg, A. Kumar, R. Rastogi, and B. Yener, *Provisioning a virtual private network: A network design problem for multicommodity flow*, Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing, 2001, pp. 389–398.

[19] W. Herroelen and R. Leus, *The construction of stable project baseline schedules*, Eur. J. Oper. Res. **156** (2004), 550–565.

[20] R. Leus, *The generation of stable project plans: Complexity and exact algorithms*, Ph.D. thesis, Katholieke University, 2003.

[21] R. Leus and W. Herroelen, *A branch-and-bound algorithm for stable scheduling in single-machine production systems*, Tech. report, KU Leuven-Departement toegepaste economische wetenschappen, 2004.

[22] R. Leus and W. Herroelen, *The complexity of machine scheduling for stability with a single disrupted job*, Oper. Res. Lett. **33** (2005), 151–156.

[23] C. Liebchen, M. Lübbecke, R. Möhring, and S. Stiller, *The concept of recoverable robustness, linear programming recovery, and railway applications*, Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 1–27.

[24] R. Lucas, *Planification adaptative des ressources ferroviaires*, Ph.D. thesis, ENSTA Paris, 2020.

[25] R. Lucas, Z. Ales, S. Elloumi, and F. Ramond, *Reducing the adaptation costs of a rolling stock schedule with adaptive solution: The case of demand changes*, RailNorrköping 2019. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA), Norrköping, Sweden, June 17th–20th, 2019, Number 069 in Linköping University Electronic Press pp. 857–876.

[26] S.V. Mehta and R.M. Uzsoy, *Predictable scheduling of a job shop subject to breakdowns*, IEEE Trans. Robotics Automation **14** (1998), 365–378.

[27] R. O'Donovan, R. Uzsoy, and K.N. McKay, *Predictable scheduling of a single machine with breakdowns and sensitive jobs*, Int. J. Production Res. **37** (1999), 4217–4233.

[28] A. Ouorou and J.P. Vial, *A model for robust capacity planning for telecommunications networks under demand uncertainty*, 2007 6th International Workshop on Design and Reliable Communication Networks, 2007, pp. 1–4.

[29] M. Poss and C. Raack, *Affine recourse for the robust network design problem: Between static and dynamic routing*, Networks **61** (2013), 180–198.

[30] M. Sevaux and K. Sörensen, *A genetic algorithm for robust schedules in a one-machine environment with ready times and due dates*, Q. J. Belgian, French Italian Oper. Res. Societies **2** (2004), 129–147.

[31] A.L. Soyster, *Convex programming with set-inclusive constraints and applications to inexact linear programming*, Oper. Res. **21** (1973), 1154–1157.

[32] S. Van de Vonder, E. Demeulemeester, W. Herroelen, and R. Leus, *The use of buffers in project management: The trade-off between stability and makespan*, Int. J. Production Econ. **97** (2005), 227 – 240.