

**A practicable robust counterpart
formulation for decomposable functions:
A network congestion case study**

E. Delage, L.G. Gianoli,
B. Sansò

G-2015-27

March 2015

Les textes publiés dans la série des rapports de recherche *Les Cahiers du GERAD* n'engagent que la responsabilité de leurs auteurs.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2015.

The authors are exclusively responsible for the content of their research papers published in the series *Les Cahiers du GERAD*.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2015.

A practicable robust counterpart formulation for decomposable functions: A network congestion case study

Erick Delage^a

Luca G. Gianoli^b

Brunilde Sansò^c

^a GERAD & HEC Montréal, Department of Decision Sciences, Montréal (Québec) Canada, H3T 2A7

^b GERAD & Polytechnique Montréal, Department of Electrical Engineering, Montréal (Québec) Canada, H3C 3A7 & Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Italy

^c GERAD & Polytechnique Montréal, Department of Electrical Engineering, Montréal (Québec) Canada, H3C 3A7

erick.delage@hec.ca

luca-giovanni.gianoli@polymtl.ca

brunilde.sanso@polymtl.ca

March 2015

Les Cahiers du GERAD

G-2015-27

Copyright © 2015 GERAD

Abstract: Robust optimization (RO) is a powerful mean to handle optimization problems where there is a set of parameters that are uncertain. The effectiveness of the method is especially noticeable when these parameters are only known to lie inside some uncertainty region. Unfortunately, there are important computational considerations that have prevented the methodology from being fully adopted in fields of practice where the cost function that needs to be *robustified* is nonlinear with respect to such parameters. In this paper, we propose a new robust optimization formulation that circumvent the computational burden in problems where the cost decomposes as the sum of convex costs for each decision variable. This is done by exploiting the fact that in this formulation the worst-case cost function can be expressed as a convex combination between a nominal and an upper-bound cost function. One can still control the conservatism of the robust solution by adjusting how many terms of the total cost function can simultaneously reach their respective most pessimistic value. In order to demonstrate the potential of our “practicable robust counterpart” formulation, we present how it can be employed on the robust optimization of packet routing on a telecommunication network with congestion. In such problems, an important source of uncertainty stems from the queueing delay, which is perhaps best approximated by a nonlinear convex function using the theory of M/M/1 queueing systems. Computational results on a large number of problem instances of realistic size confirm that it is possible to identify robust solutions that significantly outperform a deterministic approach in terms of both the amount of congestion and the risks of excess congestion. Moreover, our proposed method also improves significantly the quality of solutions that are obtained compared to an approximation scheme that is naively based on linearizing the delay function.

Key Words: Robust optimization, nonlinear uncertainty, queueing delay, convex combination, budget uncertainty

Acknowledgments: The first author gratefully acknowledges the support of the Canadian Natural Sciences and Engineering Research Council under [Grant 386416-2010] . The second and third authors are also supported by the Canadian Natural Sciences and Engineering Research Council under [Grant RGPIN123655-12].

1 Introduction

There is no doubt that most optimization problems that are encountered in modern practice are to some extent affected by the imperfect knowledge of the parameters that define them. While stochastic programming efficiently allows to account for such uncertainty when one can commit to a distribution that captures it well, such a distribution can often be unavailable. This is potentially the main argument that motivated the recent surge in interest for robust optimization (RO) models which only require one to identify an uncertainty set known to contain the unknown realization of the parameters (see Ben-Tal et al. (2009) and Bertsimas et al. (2011) for some seminal work on this topic). Another popular argument that promoters of the robust optimization framework have emphasized is the fact that these models benefit from having solutions that are more tractable than their stochastic programming version. This is for example the case for many linear programming and mixed integer linear programming models (see for instance Gabrel et al. (2014) and Ben-Tal et al. (2009)).

Unfortunately, in recent years, the scientific community has been confronted to real computational challenges when attempting to apply the RO framework to optimization problems involving functions that are non-concave in the uncertain parameters. This difficulty can be observed for instance in facility location problems where theoreticians struggle at improving the efficiency of exact and approximate solution methods (see Zeng and Zhao (2013) and Ardestani-Jaafari and Delage (2014) respectively). From a broader perspective, similar issues arise the moment that one wishes to account for the fact that decisions might not be implemented exactly as prescribed (see Ben-Tal and den Hertog (2011) for recent progress in context of conic quadratic problems). Although some significant progress has been made in the development of tractable solution methods for problems with non-concave objective functions, it appears that such methods must be identified on a case by case basis (see the tractable special cases identified in Ben-Tal et al. (2012)). Overall, these difficulties have prevented, or at least severely retarded, the adoption of the methodology in a number of fields of practice.

In this paper, we will overcome most of the computational challenges described above by proposing a new way of formulating robust optimization problems in situations where the function that needs to be “robustified” decomposes as the sum of convex costs for each decision variable. Specifically, the robust optimization model we are interested in takes the form:

$$\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}_\xi} \sum_{i=1}^n h_i(x_i, \xi), \quad (1)$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is a convex set of feasible decision vectors, $\xi \in \mathbb{R}^m$ is a vector of uncertain parameters, and $\mathcal{U}_\xi \subseteq \mathbb{R}^m$ is an arbitrary closed uncertainty set known to contain the true realization of ξ and “centered” at some nominal version $\hat{\xi}$ of the parameters. Compared to the classical robust optimization framework, we do not make any assumption about how $h_i(x_i, \xi)$ behaves in terms of ξ and about the geometry of \mathcal{U}_ξ , but instead assume that one is able for each term of the sum to formulate a convex function $f_i^+(\cdot)$ that provides a uniform upper bound for the cost that might be incurred through this term, i.e. $f_i^+(x_i) \geq \max_{\xi \in \mathcal{U}_\xi} h_i(x_i, \xi) \forall x_i \in \mathbb{R}$.

Since solving problem (1) exactly is known to be generally intractable, we propose to reformulate the robust optimization problem using the following formulation:

$$\min_{x \in \mathcal{X}} \max_{\alpha \in \mathcal{U}_\alpha} \sum_{i=1}^n \sup_{f_i: \mathcal{D}_i(f_i) \leq \alpha_i} f_i(x_i), \quad (2)$$

where $\mathcal{D}_i(f_i)$ is a measure of how much $f_i(\cdot)$ “deviates” from the nominal cost function $h_i(\cdot, \hat{\xi})^1$, and where $\mathcal{U}_\alpha \subseteq \mathbb{R}_+^n$ is a convex set of permissible joint deviations for the different terms of the summation. Intuitively, our hope is that in many situations it is possible to approximate closely the worst-case analysis of such a robust optimization problem by making the model focus on how much each cost function $h_i(\cdot, \xi)$ might diverge from its nominal form and on controlling, with the use of \mathcal{U}_α , whether large deviations can occur simultaneously over different cost terms. Note that one can easily define each member of the pair $(\mathcal{D}, \mathcal{U}_\alpha)$

such that problem (2) necessarily provides a solution that is guaranteed to achieve an actual objective value that is smaller than the value that is optimized.

While many definitions are possible for $\mathcal{D}_i(f_i)$, our proposed definition will compare the size of the deviation of $f_i(\cdot)$ relatively to the upper bound $f_i^+(\cdot)$. This will conveniently allow problem (2) to reduce to a convex optimization problem which structure is nearly more complex than the problem where ξ is well-known. In particular, it will take the form of what will be referred as the Practicable Robust Counterpart (PRC) formulation:

$$\text{(PRC)} \quad \underset{x \in \mathcal{X}}{\text{minimize}} \quad \max_{\alpha \in \mathcal{U}_\alpha} \sum_{i=1}^n \alpha_i f_i^+(x_i) + (1 - \alpha_i) h_i(x_i, \hat{\xi}). \quad (3)$$

One might recognize here that $\alpha \in \mathcal{U}_\alpha \subseteq [0, 1]^n$ can be interpreted as a parameter that identifies which convex combination of $h_i(\cdot, \hat{\xi})$ and $f_i^+(\cdot)$ captures the cost of x_i in the worst-case total cost. Moreover, the most important advantage of this formulation will be that if \mathcal{U}_α has a favorable representation, then problem (3) can be reformulated as a convex optimization problem of reasonable size. In order to allow some control over how conservative the robust solution obtained with problem (2) is, \mathcal{U}_α will take the shape of the budgeted uncertainty set popularized in Bertsimas and Sim (2004). Effectively, this uncertainty set will allow us to control how many functions $f_i(x_i)$ simultaneously take on their most pessimistic realizations $f_i^+(x_i)$.

In order to demonstrate the potential of the PRC formulation, we briefly present how it might be employed on the robust optimization of packet routing on a telecommunication network with congestion. Namely, while delays of packets that travel on different links of such networks are typically modeled using M/M/1 queuing system, there are strong empirical evidence indicating that a number of factors cause the realized transmission time to often suffer substantial deviations from these theoretical estimates (see Eijl (2002) as portrayed in Figure 1). In this context, RO has the potential to help Internet Service Providers (ISPs) to provide the requested *Quality of Service* (QoS) to incoming traffic. To the best of our knowledge, our application of robust optimization to this family of problems is both the first to account for the nonlinear effect of perturbations on link delays and the first with a computational complexity that is invariant to the number of trajectories that packets are allowed to take. We exploit both of these strengths to produce a numerical experiment that illustrates, in an exhaustive set of realistic network management conditions, what type of trade-off can be achieved between the amount of total congestion that is expected versus the amount that is at risk of being achieved under less favorable conditions.

Overall, we believe the contribution of this article to be three-fold:

1. To the best of our knowledge, we propose for the first time the use of robust optimization problem (3) to address parameter uncertainty in an optimization problem where the objective function decomposes as a sum of functions over each decision variable. We expect such a model to be especially powerful in practice when information about each function is limited to a series of historical measurements (see Remark 1).
2. We discuss for the first time how problem (3) can be used to approximate intractable instances of robust optimization problems that take the shape of problem (1). In particular, we show how it is always possible to design \mathcal{U}_α in order for our novel PRC formulation to provide a conservative solution, and in some case with careful engineering even provide an exact solution. Furthermore, we describe a simple procedure that can be used to calibrate the robust optimization model based on historical data.
3. Using the PRC formulation, we identify for the first time robust packet routing strategies that capture transmission delay uncertainty in a nonlinear fashion.

The remainder of the article is organized as follows. In Section 2 we look at the state-of-the-art literature on general RO and on RO application for the Telecom domain. We then present in Section 3 our PRC formulation, while in Section 4 we discuss a case of study application for a robust congestion minimization problem in telecommunication networks. Two illustrative examples for this application are presented in Section 5. Finally, Section 6 contains a discussion on computational results obtained by experimenting with both our PRC formulation and an approximation method based on a first-order Taylor expansion of the total cost function. Conclusions and final remarks are given in Section 7.

Remark 1 It is worth observing that problem (2) is a model that is also interesting in its own right in many practical situations where one has access to historical data about each terms of the cost function, i.e. historical pairs $(x_i, f_i(x_i))$ and yet cannot identify (or does not wish to assume there exists) relevant observable primitives that can explain the relation between x_i and $f_i(x_i)$. Take for instance the case presented in Figure 1 presenting historical observations of the relation between the percentage of utilization of a communication link that participates to the world wide web and the queuing delay that is incurred by packets traversing it. Based on this graph, the most transparent way of defining uncertainty about queuing delays might simply be to assume that the delay function deviates from the mean curve (expressed by the series of diamonds) by up to some amount that should reasonably not extend beyond the 99th percentile curve (or actually a convex approximation of it). Furthermore, we would not expect the delay curves for all links on a routing path to be at their respective 99th percentile concurrently.

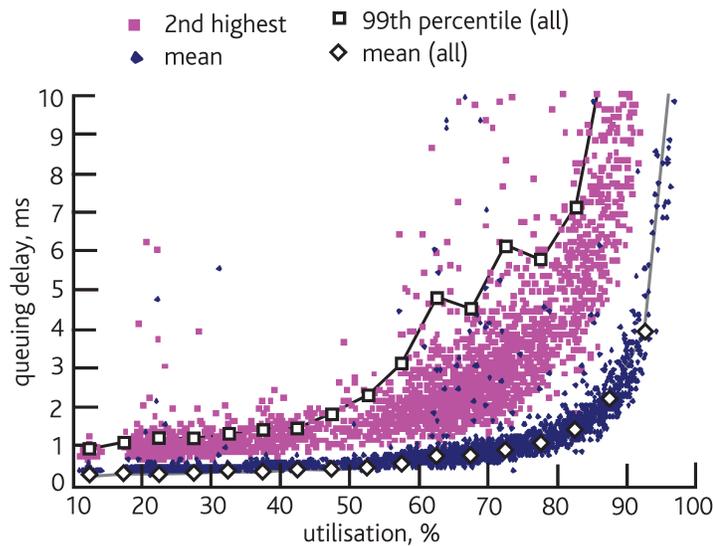


Figure 1: Delay measurements of C. Van Eijl Eijl (2002). The figure, which is taken from Hijazi et al. (2013), was originally published in Eijl (2002). With kind permission from Springer Science and Business Media.

2 Related work

Already in the fifties in the preliminary work of Dantzig (1955), and some years later in the study presented in Soyster (1973), *Robust Optimization* (RO) was identified as a powerful modeling approach to explicitly handle data uncertainty. Driven by work like Ben-Tal and Nemirovski (2000) that demonstrated how small data perturbation could seriously affect the quality of optimal solution, a comprehensive literature on RO emerged in the late 1990s and early 2000s and shaped its founding principles: see for instance El Ghaoui and Lebret (1997), Ben-Tal et al. (1998), Ben-Tal and Nemirovski (1998, 1999, 2000, 2002, 2003), and Bertsimas and Sim (2003, 2004) to name a few.

To be more specific, the robust version of a problem where all parameters are considered to be known (a.k.a the nominal problem) is technically referred as the *robust counterpart*. In particular, consider a convex optimization problem

$$(Nominal) \quad \underset{x \in \mathcal{X}}{\text{minimize}} \quad h(x, \hat{\xi}),$$

where $h(x, \xi)$ captures a total cost function that is parameterized with $\xi \in \mathbb{R}^m$, while x and \mathcal{X} are as before. When the parameters of this problem are considered to be uncertain, the robust counterpart can be

formulated as follows:

$$(RC) \quad \underset{x \in \mathcal{X}}{\text{minimize}} \quad \max_{\xi \in \mathcal{U}_\xi} h(x, \xi),$$

which can also be presented as

$$\begin{aligned} & \underset{t, x \in \mathcal{X}}{\text{minimize}} && t \\ & \text{subject to} && t \geq h(x, \xi), \forall \xi \in \mathcal{U}_\xi, \end{aligned}$$

where t captures a guaranteed bound on total cost. In other words, the RC formulation seeks a solution that has the lowest guaranteed total-cost under any realization of the parameters.

It is generally considered that the complexity of the RC formulation depends (i) on the structure of the nominal problem, with more complexity arising when passing from a linear program, to a conic program or even general convex optimization problem; and (ii) on the structure of the uncertainty region, with polyhedral (see in particular the budgeted uncertainty set presented in Bertsimas and Sim (2004)) and ellipsoidal regions being the most commonly used (c.f. Ben-Tal et al. (2009) and reference therein for an exhaustive overview). Yet, most tractable formulations that are currently discussed in the literature assume that the uncertain parameters are linearly involved in the objective function of the nominal problem.

When attempting to survey applications of RO in nonlinear contexts, one might identify the following relevant piece of work. In Li et al. (2011), the authors deal with convex and non-convex quadratic problems with interval uncertainty sets, while ellipsoidal uncertainty regions are considered in Takeda et al. (2010). In Hsiung et al. (2008), the authors tackle a geometric program where they exploit a piece-wise linear approximation of the nonlinear constraints in order to derive a tractable approximation. One can also find some work in Bertsimas et al. (2010) on “black-box” optimization methods that are adapted to RO problems where only the value and gradient of the objective function are available. Note that this particular method can get stuck with a solution that has arbitrarily poor worst-case performance simply because it is unable to reach beyond a locally optimal worst-case scenario. More recently, a compelling general framework was proposed in Ben-Tal et al. (2012) that can be used to construct tractable reformulations for a wide range of robust nonlinear optimization problems. This framework is however mostly limited to cost functions that are concave with respect to the uncertain parameters although some interesting truly non-linear special cases could also be addressed. To the best of our knowledge, there is at this time no known way of employing any of these recent development to address general cost function that decompose as in problem (1).

In telecommunication networks, there are different types of problems that can be affected by data uncertainty, such as routing, network design and capacity assignment, to name the most important. Uncertainty impacts different factors in those problems, for instance, traffic, delay, jitter or energy consumption. Only traffic uncertainty, however, has been largely explored in the robust or stochastic optimization literature, where one can find different methods to address it. First, the well-known cardinality-constrained approach of Bertsimas and Sim (2003) was implemented in Addis et al. (2013) and Addis et al. (2014) to model traffic variations on the network links in a multi period network design context related to energy-awareness. A variant of this approach is found in Coudert et al. (2013) where the uncertainty relates to the amount of traffic that is considered redundant.

A second strategy to model traffic uncertainty is the “hose-model” according to which only the overall quantity of traffic generated by (or destined to) a node is known, but the traffic requirements of a single traffic demand may vary while respecting a constraint on the total outbound (or inbound) traffic. The hose-model leads toward a polyhedral uncertainty structure. Telecommunication problems that deal with this approach are Mulyana and Killat (2005), Ben-Ameur and Kerivin (2005), Altın et al. (2007), and Altın et al. (2012), all working on different versions of routing optimization.

When we deal with network design or capacity assignment, the type of routing that is considered in the procedure can have a great impact on the overall robustness of the designed network. On that respect, static or dynamic routing have been taken into account as subproblems of network design or capacity assignment.

In static routing, a single routing realization must respect the capacity constraints for all the traffic matrices belonging to the uncertainty space whereas in dynamic routing there must be at least a feasible routing solution for each possible traffic matrix. Static routing produces more conservative solutions while dynamic routing allows to further optimize the installation costs. However, dynamic routing is way more complex to implement Mattia (2013). The new concept of “affine routing” has then been introduced in the network design or capacity assignment literature to add flexibility to static routing without getting to the complexity of dynamic routing. Notable literature in this area are Ouorou (2011, 2013), Poss and Raack (2013).

To the best of our knowledge, the only work dealing with uncertain nonlinear functions, namely uncertain queueing delay function in the Telecommunications literature, is presented in Hijazi et al. (2013). The authors propose a robust counterpart for a nonlinear network design problem with hard constraints on the maximum delay experienced by each traffic demand on its routing path. Link queueing delay is approximated by the well known delay function for M/M/1 systems:

$$\text{delay}_l(x_l) := \frac{1}{c_l - x_l}, \quad (4)$$

where the delay on link l is asymptotic to infinite when flow value x_l approaches the link capacity c_l . Delay uncertainty is modeled by adding to $\text{delay}_l(x_l)$ an uncertain parameter a_l which varies into a close symmetric interval. Note that, even though the introduction of a_l pushes the model to keep a certain delay margin for each demand, it fails to catch the nonlinear nature of delay variability: when utilization approaches 100%, queueing delay is naturally expected to suffer higher deviation. Differently from what is done in Hijazi et al. (2013), we model delay uncertainty by adopting a perspective that specifically grasp the nonlinearities of traffic variability: we address delay function uncertainty by approximating delay deviation as the convex combination between an upper bound and a nominal queueing delay function.

3 A practicable robust counterpart formulation

Let us recall the general optimization problem (1):

$$\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}_\xi} \sum_{i=1}^n h_i(x_i, \xi),$$

where each cost function $h_i(x_i, \xi)$ is convex in x_i and possibly non-concave in the perturbation ξ , which deviates within the uncertainty region \mathcal{U}_ξ known to contain some nominal version of the parameters $\hat{\xi}$. While one might easily establish that this is a convex optimization problem in x , it is important to realize that it still is an intractable problem to solve in general as shown in the following proposition which proof is deferred to Appendix A.

Proposition 3.1 *Evaluating the optimal value of the inner problem of problem (1) is NP-hard even when \mathcal{U}_ξ is a polyhedron.*

By the equivalence of optimization and validity problems (see Grötschel et al. (1981)), for the epigraph reformulation of problem (1), this negative result indicates that it is very unlikely that a general tractable method exists for solving problem (1). We are therefore reduced to deriving efficient approximation schemes or heuristics for this problem.

The idea behind the PRC formulation relies on replacing problem (1) with problem (2) with $\mathcal{U}_\alpha \subseteq [0, 1]^n$ and the deviation measure defined as follows.

Definition 3.1 *Given some set $\mathcal{U}_\xi \subset \mathbb{R}^n$ describing the joint uncertainty about each term $h_i(x_i, \xi)$ of the total cost function and some nominal version $\hat{\xi} \in \mathcal{U}_\xi$ of the uncertain vector ξ , for some term i of the total cost expression, let the upper semi-deviation of the cost function $f_i : \mathbb{R} \rightarrow \mathbb{R}$ compared to the nominal cost function be defined as*

$$\mathcal{D}_i(f_i) := \inf\{\alpha \in \mathbb{R} \mid \alpha f_i^+(x) + (1 - \alpha)\hat{f}_i(x) \geq f_i(x), \forall x \in \mathbb{R}\},$$

where $f_i^+(x) := \max_{\xi \in \mathcal{U}_\xi} h_i(x_i, \xi)$ and $\hat{f}_i(x) := h_i(x_i, \hat{\xi})$.

The main reason for using a robust optimization formulation in the form of problem (2) instead of problem (1) is that although the former might be less accurate, it is amenable to be solved exactly and effectively by off the shelf optimization software. We also establish below that the choice of \mathcal{U}_α gives enough flexibility to provide conservative approximations of the robust nonlinear optimization problem and possibly in some cases an exact solution. We finally present a natural polyhedral form for \mathcal{U}_α and propose a procedure for calibrating this set based on historical data which we believe should be very useful in practice.

3.1 Computational complexity of the PRC formulation

Perhaps the most important characteristic of a ‘‘practicable’’ optimization framework is the potential of applying it to problems that are of size encountered in practice. The following reformulation of problem (2) plays a critical role in achieving this objective.

Proposition 3.2 *Given that there exists some α in the relative interior of \mathcal{U}_α , problem (2) is equivalent to the PRC formulation (3) and to the following finite convex optimization problem:*

$$\underset{x \in \mathcal{X}, y, z}{\text{minimize}} \quad \sum_i y_i + \delta^*(z | \mathcal{U}_\alpha) \quad (5a)$$

$$\text{subject to} \quad y_i \geq \hat{f}_i(x_i), \forall i \quad (5b)$$

$$y_i + z_i \geq f_i^+(x_i), \forall i, \quad (5c)$$

where $y \in \mathbb{R}^n$, $z \in \mathbb{R}^n$, and $\delta^*(z | \mathcal{U}_\alpha)$ is the conjugate function of the indicator function for the set \mathcal{U}_α , i.e.

$$\delta^*(z | \mathcal{U}_\alpha) := \sup_{\alpha \in \mathcal{U}_\alpha} \sum_i \alpha_i z_i.$$

One might realize that the equivalent reformulation presented above is amenable to convex optimization tools as long as one can represent the conjugate function $\delta^*(v | \mathcal{U}_\alpha)$ in a tractable form. As shown in Ben-Tal et al. (2012) (see Table 1), this is known to be the case for many natural convex uncertainty sets one might think of. In particular, when \mathcal{U}_α is a polyhedron, i.e. $\{\alpha \in \mathbb{R}^n \mid B\alpha \leq b\}$ with $B \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$, the problem further reduces to

$$\underset{x \in \mathcal{X}, y, z, u}{\text{minimize}} \quad \sum_i y_i + b^T u \quad (6a)$$

$$\text{subject to} \quad y_i \geq \hat{f}_i(x_i), \forall i \quad (6b)$$

$$y_i + z_i \geq f_i^+(x_i), \forall i \quad (6c)$$

$$B^T u = z \quad (6d)$$

$$u \geq 0, \quad (6e)$$

where $u \in \mathbb{R}^p$. Next, we present a short proof of this result.

Proof of Proposition 3.2: Our first step will be to show that problem (2) is equivalent to problem (3). We will then employ Fenchel duality similarly as was done in Ben-Tal et al. (2012) to get a reduction of the problem that resembles to problem (5) and apply a clever replacement of variable to make the problem amenable for convex optimization tools.

Step 1: Let us consider the operation $\sup_{f_i: \mathcal{D}_i(f_i) \leq \alpha_i} f_i(x_i)$ for some fixed values of x_i and α_i . Given our definition of $\mathcal{D}_i(\cdot)$, we must have that the supremum is attained at $f_i(x) = \alpha_i f_i^+(x) + (1 - \alpha_i) \hat{f}_i(x)$. It is therefore the case that our problem reduces to

$$\min_{x \in \mathcal{X}} \max_{\alpha \in \mathcal{U}_\alpha} \sum_{i=1}^n \alpha_i f_i^+(x_i) + (1 - \alpha_i) \hat{f}_i(x_i).$$

Note that now the objective that needs to be robustified is convex in x , since $\mathcal{U}_\alpha \subseteq [0, 1]^n$, and linear in the uncertain parameters α .

Step 2: We can apply Theorem 2 of Ben-Tal et al. (2012) to generate an equivalent reformulation of the PRC formulation that takes the form:

$$\min_{x \in \mathcal{X}} \min_{z \in \mathbb{R}^n} \delta^*(z | \mathcal{U}_\alpha) - f_*(z, x),$$

where $\delta^*(z | \mathcal{U}_\alpha)$ is the conjugate of the support function defined in the theorem and $f_*(z, x)$ is the partial concave conjugate dual of

$$f(x, \alpha) := \sum_{i=1}^n \alpha_i f_i^+(x_i) + (1 - \alpha_i) \hat{f}_i(x_i) - \delta(\alpha | [0, 1]^n),$$

where $\delta(\alpha | [0, 1]^n)$ is the indicator function that returns 0 if $\alpha \in [0, 1]^n$ and infinity otherwise, and where we made explicit what is the domain that needs to be considered in terms of α in order for $f(x, \alpha)$ to be convex in x . We can confirm that Fenchel duality guarantees strict duality since we assumed that the relative interior of \mathcal{U}_α was not empty and $\mathcal{U}_\alpha \subseteq [0, 1]^n$. Next, one can actually get a closed form reformulation for $f_*(v, x)$ using Lagrangian duality:

$$\begin{aligned} f_*(z, x) &= \inf_{\alpha} z^T \alpha - \left(\sum_{i=1}^n \alpha_i f_i^+(x_i) + (1 - \alpha_i) \hat{f}_i(x_i) \right) + \delta(\alpha | [0, 1]^n) \\ &= \inf_{\alpha} \sup_{\lambda, \lambda'} z^T \alpha - \left(\sum_{i=1}^n \alpha_i f_i^+(x_i) + (1 - \alpha_i) \hat{f}_i(x_i) \right) - \lambda'^T \alpha - \lambda^T (1 - \alpha) \\ &= \sup \left\{ - \left(\sum_i \hat{f}_i(x) + \lambda_i \right) \mid \lambda \geq 0, \lambda_i \geq f_i^+(x_i) - \hat{f}_i(x_i) - z_i \right\}. \end{aligned}$$

Hence, we obtain the reformulation

$$\begin{aligned} &\text{minimize}_{x, z, \lambda} \quad \sum_i \hat{f}_i(x_i) + \lambda_i + \delta^*(z | \mathcal{U}_\alpha) \\ &\text{subject to} \quad \lambda_i \geq f_i^+(x_i) - \hat{f}_i(x_i) - z_i, \forall i \\ &\quad \lambda \geq 0 \\ &\quad x \in \mathcal{X}. \end{aligned}$$

Step 3: Note that because of the first constraint in the reformulation above, it is not clear anymore whether this is indeed a convex optimization problem. Fortunately, a clever change of variable brings back this property: i.e. $y_i = \lambda_i + \hat{f}_i(x_i)$.

$$\begin{aligned} &\text{minimize}_{x, y, z} \quad \sum_i y_i + \delta^*(z | \mathcal{U}_\alpha) \\ &\text{subject to} \quad y_i \geq f_i^+(x_i) - z_i, \forall i \\ &\quad y_i \geq \hat{f}_i(x_i), \forall i \\ &\quad x \in \mathcal{X}. \end{aligned}$$

This completes our proof. □

In what follows, we provide theoretical validation regarding the quality of the solution that might be obtained through such a tractable approximation problem.

3.2 Descriptive power of the PRC formulation

A second valuable property for a practical robust optimization framework would be its descriptive power. In other words, how well can a formulation like problem (2) allow us to approach the solution of problem (1). In this regard, while we can certainly not provide a recipe for constructing \mathcal{U}_α so that problem (2) generally provides solutions that are exactly optimal, we can actually confirm that by designing \mathcal{U}_α carefully, one is at least able to obtain a conservative approximation of this exact solution, and in some case for a “well crafted” choice of \mathcal{U}_α even retrieve an optimal solution.

Proposition 3.3 *Let*

$$\hat{\mathcal{U}}_\alpha := \text{Co}(\{\alpha \in [0, 1]^n \mid \exists \xi \in \mathcal{U}_\xi, \alpha_i = \mathcal{D}_i(h_i(\cdot, \xi))\}),$$

where $\text{Co}(\mathcal{S})$ refers to the convex hull of \mathcal{S} , then problem (2) with $\mathcal{D}_i(\cdot)$ as defined in Definition 3.1 and $\hat{\mathcal{U}}_\alpha$ (or equivalently the PRC formulation (3)) returns a feasible solution that is guaranteed to achieve a worst-case total cost that is lower than the approximate optimal value obtained from its objective function.

Proof of Proposition 3.3: One can see that by the definition of $\hat{\mathcal{U}}_\alpha$, when it is used, the condition

$$\forall \xi \in \mathcal{U}_\xi, \exists \alpha \in \hat{\mathcal{U}}_\alpha, \mathcal{D}_i(h_i(\cdot, \xi)) \leq \alpha_i \forall i$$

is satisfied. Hence,

$$\max_{\xi \in \mathcal{U}_\xi} \sum_i h_i(x_i, \xi) \leq \max_{\xi \in \mathcal{U}_\xi} \sum_i \sup_{f_i: \mathcal{D}_i(f_i) \leq \mathcal{D}_i(h_i(\cdot, \xi))} f_i(x_i) \leq \max_{\alpha \in \hat{\mathcal{U}}_\alpha} \sum_i \sup_{f_i: \mathcal{D}_i(f_i) \leq \alpha_i} f_i(x_i).$$

□

In some sense, $\hat{\mathcal{U}}_\alpha$ provides the tightest conservative approximation that problem (2) can achieve with a convex uncertainty set. In practice however, we expect that it might be impossible to characterize $\hat{\mathcal{U}}_\alpha$ with a tractable representation and might need to settle with some \mathcal{U}_α of simpler form that respects

$$\hat{\mathcal{U}}_\alpha \subseteq \mathcal{U}_\alpha \subseteq [0, 1]^n$$

in order to minimize an upper bound on worst-case cost that is a little less tight yet more tractable

$$\max_{\xi \in \mathcal{U}_\xi} \sum_i h_i(x_i, \xi) \leq \max_{\alpha \in \mathcal{U}_\alpha} \sum_i \sup_{f_i: \mathcal{D}_i(f_i) \leq \alpha_i} f_i(x_i) \leq \max_{\alpha \in \hat{\mathcal{U}}_\alpha} \sum_i \sup_{f_i: \mathcal{D}_i(f_i) \leq \alpha_i} f_i(x_i).$$

Alternatively, while most approximation schemes developed for robust optimization are concerned with obtaining solutions that embody the principle of conservativeness, we believe that from a modeling standpoint it is also valuable to confirm that it is theoretically possible to obtain the true optimal solution by solving a well constructed approximation model. The following result confirms this property in some fairly general circumstances for the PRC formulation (3). We refer the readers to Appendix B for a complete proof.

Proposition 3.4 *Let the following four conditions be satisfied:*

1. The sets \mathcal{X} and \mathcal{U}_ξ are compact and each term $h_i(x_i, \xi)$ is bounded above over the Cartesian product $\mathcal{X} \times \mathcal{U}_\xi$.
2. For all $\xi \in \mathcal{U}_\xi$, the functions $h_i(x_i, \xi)$ are differentiable and strictly convex in x_i .
3. For all $x \in \mathcal{X}$, $h_i(x_i, \xi)$ is upper semi-continuous in ξ .
4. For all $x \in \mathcal{X}$, $\frac{df_i^+(x_i)}{dx_i} \geq \frac{dh_i(x_i, \xi)}{dx_i} \geq \frac{df_i(x_i)}{dx_i}$ for all $\xi \in \mathcal{U}_\xi$ and all i .

Then, there exists a polyhedral set $\mathcal{U}_\alpha \subseteq [0, 1]^n$ such that problem (2) with $\mathcal{D}_i(\cdot)$ as defined in Definition 3.1 (or equivalently the PRC formulation (3)) returns an optimal solution for problem (1).

This result confirms that the approximation scheme that is proposed is well motivated. In simple words, it states that if problem (1) satisfies the mild conditions 1 to 4 (see Section 4.3 for a practical example where these are met), then it is theoretically possible to identify a set \mathcal{U}_α under which the approximate solution returned by the PRC formulation ends up being the actual optimal solution of problem (1). Unfortunately, the proof does not provide a recipe for designing such a \mathcal{U}_α .

3.3 Employing a budgeted uncertainty set for \mathcal{U}_α

The results presented above indicate that the choice of \mathcal{U}_α plays a critical role in the quality of the solution that is obtained from the PRC formulation. To this end, we now propose a version of \mathcal{U}_α which is inspired from the budgeted uncertainty set introduced in Bertsimas and Sim (2003).

Definition 3.2 *Given some $\Gamma \geq 0$ capturing the total number of terms of the total cost function $\sum_i h_i(x_i, \xi)$ that could stray away from their respective nominal version $h_i(x_i, \hat{\xi})$, the “budgeted uncertainty set” takes the form:*

$$\mathcal{U}_\alpha(\Gamma) := \left\{ \alpha \in [0, 1]^n \mid \sum_{i=1}^n \alpha_i \leq \Gamma \right\}.$$

Indeed, the budgeted uncertainty set $\mathcal{U}_\alpha(\Gamma)$ has the potential to capture the fact that each term of the cost function is not expected to reach their respective worst-case value simultaneously. This intuition might be especially fruitful in situations where each term of the cost function involves mutually exclusive subsets of perturbation parameters considered independent from each other. For example, in the case where $f_i(x_i, \xi) = \sum_i h_i(x_i, \xi_i)$ with independent ξ_i 's. The use of a budgeted uncertainty set also gains much of its practical relevance from the ease with which one can adjust the budget Γ in order to study in a particular problem instance the inherent tradeoff that needs to be made between total cost under the nominal problem and total cost under potential worst-case scenarios, a.k.a. the “price of robustness”. This feature will be demonstrated empirically in Section 6.

It follows naturally from our results about computational efficiency that using $\mathcal{U}_\alpha(\Gamma)$ allows the PRC formulation (3) to reduce to a convex optimization problem of relatively small size. The following reformulation can easily be obtained by applying Proposition 3.2 and in particular by exploiting the structure presented in problem (6). The details can be found in Appendix C.

Corollary 3.1 *Problem (2) with \mathcal{D} defined as in Definition 3.1 (or equivalently the PRC formulation (3)) and with $\mathcal{U}_\alpha = \mathcal{U}_\alpha(\Gamma)$ reduces to the following convex optimization problem:*

$$\underset{x, y, u}{\text{minimize}} \quad \sum_i y_i + \Gamma u \tag{7a}$$

$$\text{subject to} \quad y_i \geq \hat{f}_i(x_i), \forall i \tag{7b}$$

$$y_i + u \geq f_i^+(x_i), \forall i \tag{7c}$$

$$x \in \mathcal{X}, \tag{7d}$$

where $u \in \mathbb{R}$.

Given that in practice, information about ξ might be based on historical data, we conclude this section by describing a procedure that can be followed to calibrate Γ based on such data in order for the robust solution that is obtained using problem (7) to have interesting statistical properties.

Proposition 3.5 *Let $\{\xi^k\}_{k=1}^K$ be a set of independently and identically distributed random vector based on distribution F and let, for each k , $\Gamma^k := \sum_i \mathcal{D}_i(h_i(\cdot, \xi^k))$ capture the total deviation observed over the terms of the total cost under realization ξ^k . Then, one can consider that*

1. For any $K \geq 1$, with probability larger than $1 - \epsilon$

$$\forall x \in \mathcal{X}, \quad \mathbb{P} \left(\sum_i h_i(x_i, \xi) \leq \sup_{\alpha \in \mathcal{U}_\alpha(\bar{\Gamma})} \sum_i \alpha_i f_i^+(x_i) + (1 - \alpha_i) \hat{f}_i(x_i) \right) \geq 1 - \beta,$$

with $\beta = \frac{1}{\epsilon(K+1)}$, when $\bar{\Gamma}$ is set to be equal to $\max_k \Gamma^k$.

2. For K large enough, with probability larger than $1 - \epsilon$

$$\forall x \in \mathcal{X}, \quad \mathbb{P} \left(\sum_i h_i(x_i, \xi) \leq \sup_{\alpha \in \mathcal{U}_\alpha(\bar{\Gamma})} \sum_i \alpha_i f_i^+(x_i) + (1 - \alpha_i) \hat{f}_i(x_i) \right) \geq 1 - \beta,$$

when $\bar{\Gamma}$ is set to be equal to the $\lceil K(1 - \beta) + \Phi^{-1}(1 - \epsilon/2) \sqrt{K\beta(1 - \beta)} \rceil$ -th term in the ordered list (from smallest to largest) of Γ^k 's, with $\Phi^{-1}(\cdot)$ the inverse function of the cumulative density function of a standard normal distribution.

In particular, this proposition implies that there are simple ways of choosing Γ in order to obtain a robust solution that has only β chances of incurring a cost that is larger than the optimal value of problem (7). Hence, one can see the optimal solution of this problem as an approximate solution to the problem of minimizing the “value at risk” of the total cost:

$$\begin{aligned} & \underset{s \in \mathbb{R}, x \in \mathcal{X}}{\text{minimize}} && s \\ & \text{subject to} && \mathbb{P}_F \left(\sum_i h_i(x_i, \xi) \leq s \right) \geq 1 - \beta, \end{aligned}$$

without making any assumption about the distribution of ξ beside that it is the same distribution that generated each sample of the historical data set.

Proof of Proposition 3.5: In both cases, we will demonstrate that the probability $\mathbb{P}_F(\sum_i \mathcal{D}_i(h_i(\cdot, \xi)) \leq \bar{\Gamma}) \geq 1 - \beta$. Given that this is true, the rest follows simply since given any $x \in \mathcal{X}$ and any ξ that satisfies $\sum_i \mathcal{D}_i(h_i(\cdot, \xi)) \leq \bar{\Gamma}$, the pair (x, ξ) will also satisfy

$$\sum_i h_i(x_i, \xi) \leq \max_{\alpha \in \mathcal{U}_\alpha(\bar{\Gamma})} \sum_i \sup_{f_i: \mathcal{D}_i(f_i) \leq \alpha_i} f_i(x_i) = \sup_{\alpha \in \mathcal{U}_\alpha(\bar{\Gamma})} \sum_i \alpha_i f_i^+(x_i) + (1 - \alpha_i) \hat{f}_i(x_i).$$

In Case 1, the result follows from Corollary 1 in Calafiore and Campi (2005) since $\bar{\Gamma} \in \mathbb{R}$ can be seen as the optimal solution of

$$\begin{aligned} & \underset{\Gamma}{\text{minimize}} && \Gamma \\ & \text{subject to} && \sum_i \mathcal{D}_i(h_i(\cdot, \xi)) \leq \Gamma, \forall \xi \in \{\xi^k\}_{k=1}^K, \end{aligned}$$

where all ξ 's in $\{\xi^k\}_{k=1}^K$ are drawn from F .

To obtain the guarantee presented in Case 2, one can use a well known result in statistics about establishing a confidence interval for a percentile of some random variable Z based on K i.i.d. samples. In particular, it is known that when K is large enough one can have $1 - \epsilon$ confidence that the $1 - \beta$ percentile is smaller than the k^* -th term of the ordered list of $\{Z^k\}_{k=1}^K$, where k^* is the smallest integer above $K(1 - \beta) + \Phi^{-1}(1 - \epsilon/2) \sqrt{K\beta/(1 - \beta)}$. \square

A data-driven context that is a bit more complicated to handle occurs when one only has access to the historical costs that were paid for the specific decisions that were implemented without any information about the realized ξ 's. In particular, let $\{(x^k, h^k)\}_{k=1}^K$ be a set of historical data point capturing at K different

points of time k which were the decisions $(x_1^k, x_2^k, \dots, x_n^k)$ that were made and which were the individual costs that were paid $(h_1^k, h_2^k, \dots, h_n^k)$. When each cost term of the total cost expression are independent from each other, this information might for instance take the shape of figures like Figure 1. Indeed, one might be tempted in this situation to follow a similar data analysis as above yet this must be done with care in order for the chosen $\bar{\Gamma}$ to be insensitive to the historical distribution of implemented decisions x^k . To resolve this issue, one might be able to employ the result of Proposition 3.5 after conservatively estimating each Γ^k as the largest value that can be motivated by the pair (x^k, h^k) , namely,

$$\begin{aligned} \Gamma^k &:= \max_{\xi \in \mathcal{U}_\xi} \sum_i \mathcal{D}_i(h_i(\cdot, \xi)) \\ &\text{s.t.} \quad h_i(x_i^k, \xi) = h_i^k, \forall i. \end{aligned}$$

Alternatively, it might also be possibly to recuperate information about ξ^k , and implicitly about its distribution, through Bayesian inference. Recognizing that there are serious difficulties that need to be addressed in both cases, we leave both directions of solutions open to further study.

4 A network congestion case study

There are different performance measures that are key for telecommunication networks planning and evaluation. They typically depend on the type of technologies that are implemented. IP networks are based on packet-switching, which is a store-and-forward method that depends on buffering packets, on network devices along the route, up to their final destination. For those networks, congestion measures depending on the queuing behaviour of the traffic are important to evaluate network performance (see Fortz and Thorup (2004) and Hijazi et al. (2013)). Telephone networks and some other access technologies, deal instead with circuit-switching, a method that implies the reservation of circuits for the whole duration of a call. The key performance measures in those cases is the probability of blocking (see Plante and Sansò (2002)). In recent years, there has been another aspect taken into account in network planning, the energy consumption (as in Restrepo et al. (2009)) that yields a trade-off with congestion in Sansò and Mellah (2009).

In all three cases, we are dealing with non-linear functions that are used in network design, routing and/or capacity assignment. Those non-linearities approximate complex phenomena that cannot be perfectly known a-priori and generate uncertainty that impacts performance.

This paper deals exclusively with routing IP networks, for which congestion delay is major issue. Note, however, that congestion delay is widely used not only in routing (also called *traffic engineering* (TE)) but also in other *network design* (ND) problems either as an objective function or as part of explicit problem constraints. In most optimization models for those problems, the classical M/M/1 mean delay function is typically embedded even though it is well known that, given the nature of the traffic, it may not be appropriate in some cases (see Gendreau et al. (1996)). In fact, as shown in Figure 1 obtained from Eijl (2002), the delay curve on a real link may experience significant deviations from the mean.

This deviation becomes the core of our example since we want to show how the proposed routing robustness method can be applied in this context. For this, we start from the nominal problem, based on the classical mean delay formulation, and show how to apply the PRC formulation to a more general version of such congestion minimization problem.

4.1 The nominal congestion minimization problem

Let $G = (V, \mathcal{A})$ be a directed graph representing an IP network. V is the set of network nodes (routers), while \mathcal{A} is the set of unidirectional network links. The capacity available on a link $(i, j) \in \mathcal{A}$ is defined as c_{ij} . Given the set of traffic demands \mathbf{D} , where each demand $\mathbf{d} \in \mathbf{D}$ is characterized by a source node $o^{\mathbf{d}}$, a destination node $t^{\mathbf{d}}$ and a request of $\rho^{\mathbf{d}}$ units of traffic, our aim is to adjust the multiple routing paths used by each demand $\mathbf{d} \in \mathbf{D}$ so as to minimize a measure of network congestion based on the queuing delay

observed on each link. In particular, let $x_{ji}^{\mathbf{d}}$ be the positive real variable which represents the amount of traffic generated by demand $\mathbf{d} \in \mathbf{D}$ and routed through link $(i, j) \in \mathcal{A}$, and x_{ij} the total traffic carried by link $(i, j) \in \mathcal{A}$, we define the total congestion as

$$\frac{1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}} \sum_{(i,j) \in \mathcal{A}} \frac{x_{ij}}{c_{ij} - x_{ij}}, \quad [s/Mbit]$$

which corresponds to the weighted sum of the M/M/1 queuing delay function $\frac{1}{c_{ij} - x_{ij}}$ computed on each link. Note that the total congestion decomposes as the sum of congestion of each of the links of the network, each being affected by the amount of traffic that traverses it through the expression:

$$h_{ij}(x_{ij}) := \frac{x_{ij}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - x_{ij})}.$$

Network routing is considered multi-commodity and fully splittable. Given all the previous assumptions, let us define the following nonlinear convex programming formulation for the nominal congestion minimization problem:

$$\begin{aligned} & \underset{x, \{x^{\mathbf{d}}\}_{\mathbf{d} \in \mathbf{D}}}{\text{minimize}} && \frac{1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}} \sum_{(i,j) \in \mathcal{A}} \frac{x_{ij}}{c_{ij} - x_{ij}} && (8a) \end{aligned}$$

$$\text{subject to} \quad x_{ij} + \leq c_{ij}, \forall (i, j) \in \mathcal{A} \quad (8b)$$

$$\sum_{\mathbf{d} \in \mathbf{D}} x_{ji}^{\mathbf{d}} \leq x_{ij}, \forall (i, j) \in \mathcal{A} \quad (8c)$$

$$\sum_{(i,j) \in \mathcal{A}} x_{ji}^{\mathbf{d}} - \sum_{(j,i) \in \mathcal{A}} x_{ji}^{\mathbf{d}} = \begin{cases} \rho^{\mathbf{d}} & \text{if } i = o^{\mathbf{d}} \\ -\rho^{\mathbf{d}} & \text{if } i = t^{\mathbf{d}} \\ 0 & \text{else} \end{cases}, \forall i \in V, \mathbf{d} \in \mathbf{D} \quad (8d)$$

$$x_{ji}^{\mathbf{d}} \geq 0, \forall (i, j) \in \mathcal{A}, \mathbf{d} \in \mathbf{D}. \quad (8e)$$

In this model, constraint (8b) imposes that the total traffic on each link respects the maximum capacity. Note that in the classical nonlinear splittable flow formulation such constraints would be implied by the domain of the objective function, however, we make them explicit to facilitate the transition towards the robust model. Constraint (8c) allows to correctly compute the total amount of traffic x_{ij} carried by link $(i, j) \in \mathcal{A}$. Constraint (8d) routes each demand $\mathbf{d} \in \mathbf{D}$ from source $o^{\mathbf{d}}$ to destination $t^{\mathbf{d}}$ by splitting the traffic over multiple paths, when required. Finally, variable domains are reported by constraint (8e).

4.2 The robust counterpart

The classical M/M/1 formulation provided above is well known to be optimistic, as both empirical and simulation methods have shown that delay values can significantly deviate from it (as argued in Eijl (2002) and Hijazi et al. (2013)). In fact, given the non-Poisson nature of most Internet traffic, the delay values provided by the M/M/1 function are, in many cases, lower bounds on measured delay.

This motivates the use of a parametrized model in order to study the effect of perturbations of the delay functions at different links on the total congestion and possibly design routing strategies that are immunized against such perturbations. Practically speaking, one could argue that the following parametrized delay model might be the most flexible one to use for this purpose:

$$h_{ij}(x_{ij}, a_{ij}, b_{ij}, d_{ij}, e_{ij}) := d_{ij} \cdot \left(\frac{(1 + x_{ij})^{1+a_{ij}} - 1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - x_{ij} - b_{ij})} \right) + e_{ij} \frac{x_{ij}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}}, \quad (9)$$

where $a_{ij} \geq 0$, $b_{ij} \geq 0$, $d_{ij} \geq 0$, and $e_{ij} \geq 0$ are the parameters that define the shape of the delay function and which might be difficult to estimate precisely. Intuitively, e_{ij} might capture some additive delay (as was used

in Hijazi et al. (2013)), b_{ij} might account for temporary capacity variation due, for instance, to interference phenomena in radio link or to instantaneous increment of traffic, finally a_{ij} and d_{ij} might together serve the purpose of accounting for some additional nonlinearities that have been observed (see for instance the calibrated version of this model in Section 5.2). Note that $h_{ij}(x_{ij}, 0, 0, 1, 0)$ reduces to the M/M/1 weighted delay function $h_{ij}(x_{ij})$. Another important property of this parametrized delay function is that it is always convex in x_{ij} and increasing in its four parameters.

In what follows, for simplicity of exposure and to better grasp both the congestion uncertainty and the nonlinear nature of congestion, we apply the PRC formulation in a context where there is no uncertainty on d_{ij} and e_{ij} (i.e. the parameters that affect the delay linearly) so that $d_{ij} = 1$ and $e_{ij} = 0$. Thus, we will assume instead that the following parametrized weighted congestion cost function is used :

$$h_{ij}(x_{ij}, a_{ij}, b_{ij}) := \frac{(1 + x_{ij})^{1+a_{ij}} - 1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c_{ij} - x_{ij} - b_{ij})},$$

and that perturbations can cause each parameter to vary in its respective closed interval $[\hat{a}_{ij} - \bar{a}_{ij}, \hat{a}_{ij} + \bar{a}_{ij}]$ and $[\hat{b}_{ij} - \bar{b}_{ij}, \hat{b}_{ij} + \bar{b}_{ij}]$, where $\hat{a}_{ij} \geq 0$, $\bar{a}_{ij} < 1 + \hat{a}_{ij}$ and $\bar{b}_{ij} < c_{ij} - \hat{b}_{ij}$.

In this context, a robust optimization approach would therefore recommend retrieving the solution to the following robust nonlinear optimization problem:

$$\begin{aligned} & \underset{x, \{x^d\}_{\mathbf{d} \in \mathbf{D}}}{\text{minimize}} && \max_{(a,b) \in \mathcal{U}_{ab}} \sum_{(i,j) \in \mathcal{A}} h_{ij}(x_{ij}, a_{ij}, b_{ij}) \end{aligned} \quad (10a)$$

$$\text{subject to} \quad x_{ij} + \hat{b}_{ij} + \bar{b}_{ij} \leq c_{ij} \quad (10b)$$

$$(8c) - (8e), \quad (10c)$$

where the first constraint ensures that the traffic can be routed under any realization of b , and where

$$\mathcal{U}_{ab} \subseteq \left\{ (a, b) \in \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|} \mid \begin{array}{l} \hat{a}_{ij} - \bar{a}_{ij} \leq a_{ij} \leq \hat{a}_{ij} + \bar{a}_{ij}, \forall (i, j) \in \mathcal{A} \\ \hat{b}_{ij} - \bar{b}_{ij} \leq b_{ij} \leq \hat{b}_{ij} + \bar{b}_{ij}, \forall (i, j) \in \mathcal{A} \end{array} \right\}.$$

In particular, we will focus on the budgeted uncertainty set which captures the fact that only a limited number of parameters among all the a_{ij} 's and b_{ij} 's can take an extreme value:

$$\mathcal{U}_{ab}(\Gamma) := \left\{ (a, b) \mid \begin{array}{l} \exists (\delta^a, \delta^b) \in \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|}, \\ a_{ij} = \hat{a}_{ij} + \bar{a}_{ij} \delta_{ij}^a, \forall (i, j) \in \mathcal{A} \\ b_{ij} = \hat{b}_{ij} + \bar{b}_{ij} \delta_{ij}^b, \forall (i, j) \in \mathcal{A} \\ \|\delta^a\|_\infty \leq 1, \|\delta^b\|_\infty \leq 1 \\ \|\delta^a\|_1 + \|\delta^b\|_1 \leq \Gamma \end{array} \right\}.$$

4.3 The practicable robust counterpart formulation

As argued in Section 3, problem (10) is potentially intractable which motivates the use of the PRC formulation which consists in considering instead the following optimization problem:

$$\begin{aligned} & \underset{x, \{x^d\}_{\mathbf{d} \in \mathbf{D}}}{\text{minimize}} && \max_{\alpha \in \mathcal{U}_\alpha} \sum_{(i,j) \in \mathcal{A}} \sup_{f_{ij}: \mathcal{D}_{ij}(f_{ij}) \leq \alpha_{ij}} f_{ij}(x_{ij}) \end{aligned} \quad (11a)$$

$$\text{subject to} \quad (10b), (8c) - (8e), \quad (11b)$$

where $\mathcal{D}_{ij}(\cdot)$ follows Definition 3.1. To complete this model, we need to characterize $f_{ij}^+(x_{ij})$ using

$$f_{ij}^+(x_{ij}) := h_{ij}(x_{ij}, \hat{a}_{ij} + \bar{a}_{ij}, \hat{b}_{ij} + \bar{b}_{ij}) \geq \max_{(a,b) \in \mathcal{U}_{ab}} h_{ij}(x_{ij}, a_{ij}, b_{ij}).$$

Based on Section 3, we know that it is possible to select \mathcal{U}_α so that optimization model (11) returns a conservative approximation of the original problems (see Proposition 3.3). Furthermore, looking back at the

conditions of Proposition 3.4, we can establish that there exists some \mathcal{U}_α that will allow this model to return the truly optimal robust traffic assignment. We refer the reader to Appendix D for a detailed validation of the conditions.

Finally, based on Corollary 3.1, we get that the PRC formulation can be reduced to the following convex optimization model when $\mathcal{U}_\alpha(\Gamma)$ is employed:

$$\begin{aligned}
& \underset{x, \{x^d\}_{d \in \mathbf{D}}, y, u}{\text{minimize}} && \sum_{ij} y_{ij} + \Gamma u \\
& \text{subject to} && y_{ij} \geq \frac{(1 + x_{ij})^{1+\hat{a}_{ij}} - 1}{\sum_{d \in \mathbf{D}} \rho^d (c_{ij} - x_{ij} - \hat{b}_{ij})}, \forall (i, j) \in \mathcal{A} \\
& && y_{ij} + u \geq \frac{(1 + x_{ij})^{1+\hat{a}_{ij}+\hat{a}_{ij}} - 1}{\sum_{d \in \mathbf{D}} \rho^d (c_{ij} - x_{ij} - \hat{b}_{ij} - \bar{b}_{ij})}, \forall (i, j) \in \mathcal{A} \\
& && \text{(8c) - (8e)},
\end{aligned}$$

where $y \in \mathbb{R}^{|\mathcal{A}|}$ and $u \in \mathbb{R}$.

4.4 A Taylor series expansion reformulation

One can actually think of an alternative scheme for solving approximately problem (10) that would rely on first-order Taylor series expansion to get rid of the nonlinearities in (a, b) . Indeed it is possible to approximate each congestion terms $h_{ij}(x_{ij}, a_{ij}, b_{ij})$ with the following expansion:

$$\tilde{h}_{ij}(x_{ij}, a_{ij}, b_{ij}) = h_{ij}(x_{ij}, \hat{a}_{ij}, \hat{b}_{ij}) + (a_{ij} - \hat{a}_{ij}) \left. \frac{dh_{ij}}{da_{ij}} \right|_{(x_{ij}, \hat{a}_{ij}, \hat{b}_{ij})} + (b_{ij} - \hat{b}_{ij}) \left. \frac{dh_{ij}}{db_{ij}} \right|_{(x_{ij}, \hat{a}_{ij}, \hat{b}_{ij})},$$

where

$$\begin{aligned}
\left. \frac{dh_{ij}}{da_{ij}} \right|_{(x_{ij}, \hat{a}_{ij}, \hat{b}_{ij})} &= \frac{\ln(1 + x_{ij})(1 + x_{ij})^{1+a_{ij}}}{\sum_{d \in \mathbf{D}} \rho^d (c_{ij} - b_{ij} - x_{ij})} \Big|_{(x_{ij}, \hat{a}_{ij}, \hat{b}_{ij})} = \frac{\ln(1 + x_{ij})(1 + x_{ij})^{1+\hat{a}_{ij}}}{\sum_{d \in \mathbf{D}} \rho^d (c_{ij} - \hat{b}_{ij} - x_{ij})} \\
\left. \frac{dh_{ij}}{db_{ij}} \right|_{(x_{ij}, \hat{a}_{ij}, \hat{b}_{ij})} &= \frac{(1 + x_{ij})^{1+a_{ij}} - 1}{\sum_{d \in \mathbf{D}} \rho^d (c_{ij} - b_{ij} - x_{ij})^2} \Big|_{(x_{ij}, \hat{a}_{ij}, \hat{b}_{ij})} = \frac{(1 + x_{ij})^{1+\hat{a}_{ij}} - 1}{\sum_{d \in \mathbf{D}} \rho^d (c_{ij} - \hat{b}_{ij} - x_{ij})^2}.
\end{aligned}$$

Hence, we have that problem (10) can be approximated by the formulation below

$$\underset{x}{\text{minimize}} \quad \max_{(a,b) \in \mathcal{U}_{ab}} \sum_{(i,j) \in \mathcal{A}} \tilde{h}_{ij}(x_{ij}, a_{ij}, b_{ij}) \tag{12a}$$

$$\text{subject to} \quad \text{(10b), (8c) - (8e)}, \tag{12b}$$

where $\tilde{h}_{ij}(x_{ij}, a_{ij}, b_{ij})$ is equal to

$$\left(\frac{(1 + x_{ij})^{1+\hat{a}_{ij}} - 1}{c_{ij} - \hat{b}_{ij} - x_{ij}} + (a_{ij} - \hat{a}_{ij}) \frac{\ln(1 + x_{ij})(1 + x_{ij})^{1+\hat{a}_{ij}}}{c_{ij} - \hat{b}_{ij} - x_{ij}} + (b_{ij} - \hat{b}_{ij}) \frac{(1 + x_{ij})^{1+\hat{a}_{ij}} - 1}{(c_{ij} - \hat{b}_{ij} - x_{ij})^2} \right) \frac{1}{\sum_{d \in \mathbf{D}} \rho^d}.$$

If it so happens that for all $(a, b) \in \mathcal{U}_{ab}$ there is actually a pair $(a', b') \in \mathcal{U}_{ab}$ with $a' \geq \hat{a}$, $b' \geq \hat{b}$, $a' \geq a$ and $b' \geq b$, as is the case for $\mathcal{U}_{ab}(\Gamma)$, then this reformulation is a convex optimization problem since the objective function reduces to

$$\begin{aligned}
& \underset{(a,b) \in \mathcal{U}_{ab}}{\text{maximize}} && \sum_{(i,j) \in \mathcal{A}} \tilde{h}_{ij}(x_{ij}, a_{ij}, b_{ij}) \\
& \text{subject to} && a \geq \hat{a}, b \geq \hat{b},
\end{aligned}$$

and one can confirm that each $\tilde{h}_{ij}(x_{ij}, a_{ij}, b_{ij})$ is convex in x_{ij} for such values of a 's and b 's. In fact, when using $\mathcal{U}_{ab}(\Gamma)$, the problem can be reformulated as the following tractable convex optimization problem:

$$\begin{aligned}
& \underset{x, u, v, w}{\text{minimize}} && \sum_{(i,j) \in \mathcal{A}} \frac{(1+x_{ij})^{1+\hat{a}_{ij}} - 1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c_{ij} - \hat{b}_{ij} - x_{ij})} + v_{ij} + w_{ij} + \Gamma u \\
& \text{subject to} && u + v_{ij} \geq \bar{a}_{ij} \frac{\ln(1+x_{ij})(1+x_{ij})^{1+\hat{a}_{ij}}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c_{ij} - \hat{b}_{ij} - x_{ij})} \\
& && u + w_{ij} \geq \bar{b}_{ij} \frac{(1+x_{ij})^{1+\hat{a}_{ij}} - 1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c_{ij} - \hat{b}_{ij} - x_{ij})^2} \\
& && (10b), (8c) - (8e),
\end{aligned}$$

The downside of this approximate reformulation is that it is not clear whether it provides a good solution in terms of x . In fact, one can show that for all $1 \leq x_{ij} < c_{ij} - \hat{b}_{ij} - \bar{b}_{ij}$, the function $h_{ij}(x_{ij}, a_{ij}, b_{ij})$ is a convex function of a_{ij} and b_{ij} over the region that interests us² thus that $\tilde{h}_{ij}(x_{ij}, a_{ij}, b_{ij})$ provides a lower bound to $h_{ij}(x_{ij}, a_{ij}, b_{ij})$. Effectively, this means that problem (12) will most likely underestimate the worst-case total cost for such levels of traffic and it is likely that the solution that is returned tries to exploit this inaccuracy. We believe this might explain the somewhat erratic performance observed in the numerical experiments of Section 6 for the solutions obtained using this approximation scheme.

5 Two illustrative examples

This section is devoted to two examples that illustrate some of the modeling and resolution choices that are made in this paper. The first, presented in Subsection 5.1 shows the impact of applying the robustness method to a very simple network. The second, explained in Subsection 5.2 provides some insights about nonlinearity issues and their implication in the robust congestion minimization problem.

5.1 The value of being robust

Let us consider the 5-node network of Figure 3. There is a source node, `src`, a destination node, `dest`, and three transit nodes, `transit_1`, `transit_2` and `transit_3` that give rise to three different routes. The links capacity is 2 Gbps. The demand is such that 3 Gbps of traffic must be routed from `source` to `dest`. The three paths available are portrayed in three shades of black and grey, and their links are characterized by different delay profiles based on the a and b parameters. Nominal and upper bounds of a and b are reported in Table 1.

Examining those parameters, it can be seen that, in terms of nominal functions (\hat{a} and \hat{b}), black links experience the lowest delay. On the other hand, light-grey links are those with the worst latency. From the perspective of the upper bound function the situation is, however, reversed. In fact, black links present the worst upper bound delay profile whereas light grey links show almost no uncertainty, i.e. with upper bound values very close to nominal. Practically speaking, those links that experience the lowest delay, on average, are also those characterized by the highest variability, while those that show the highest delay, always on average, show the most stable performance. Since a robust approach is designed to optimize the worst-case scenario, it should prefer sending traffic on the grey links (the less uncertain), while a typical nominal method would give priority to black links (the fastest, on average). The above explanation can be graphically perceived by examining Figure 2 where both nominal (dashed lines) and upper bound (continuous lines) congestion functions are shown. The reader can appreciate that the black pair of lines are the widest apart, showing great uncertainty, whereas the light grey pair of lines are very close together.

To show the effect of robustness on the optimal solutions, Table 2 contains the solution results of the PRC formulation when the budget parameter Γ is varied from 0 to $|\mathcal{A}|$. Since each route is characterized by

Table 1: Set of values for the illustrative example (Figure 3).

<i>Links</i>	\hat{a}	\hat{b}	$\hat{a} + \bar{a}$	$\hat{b} + \bar{b}$
<i>Black</i>	0.4	300	0.80	600
<i>Medium – grey</i>	0.5	400	0.65	550
<i>Light – grey</i>	0.6	500	0.61	510

Links definition		
Type	Names	Route
<i>Black</i>	src-transit_1; transit_1-dest	1
<i>Medium – grey</i>	src-transit_2; transit_2-dest	2
<i>Light – grey</i>	src-transit_3; transit_3-dest	3

Other problem parameters		
Parameter	domain	Value
c_{ij}	$\forall (i, j) \in \mathcal{A}$	2 Mb/s
ρ^d	$\forall d \in \mathcal{D}$	3 Mb/s

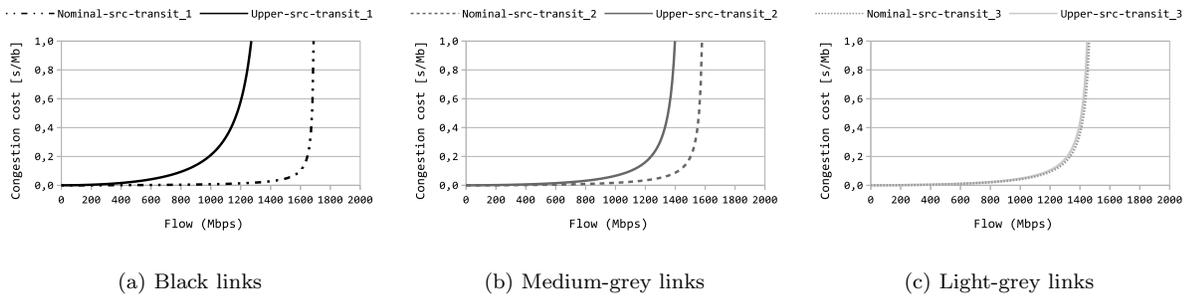


Figure 2: Nominal and upper bound functions for the links of the illustrative example (Figure 3).

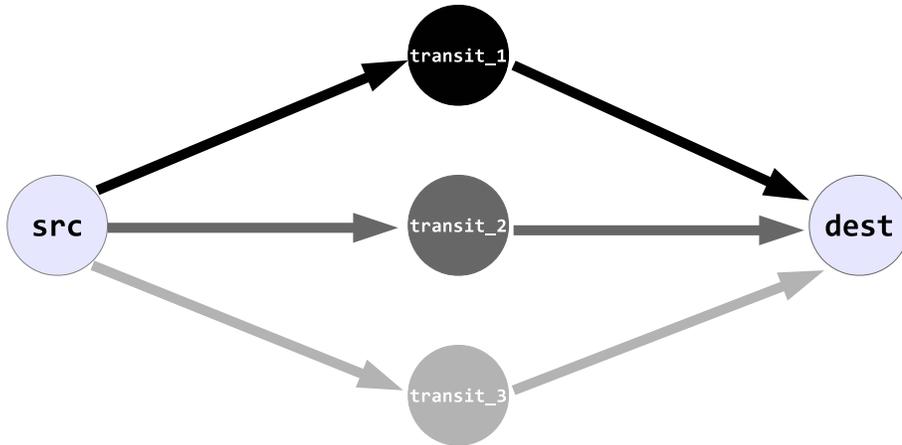


Figure 3: Three-route network illustrative example.

the “colour” of their links, from now on we assign that colour to the name of the route. As expected, the flow values (see Table 2) obtained for each value of Γ , show how traffic is progressively shifted from the black link route to the light-grey route as the solution becomes more robust. From a robustness perspective, it is more convenient to use links which are more expensive, on average, but less uncertain.

Table 2: Routing solution of the illustrative example of Figure 3. The table represents the solution of the PRC formulation when employing the values of Table 1, in terms of routed flows per route (in Mb/s) when budget Γ is varied.

Γ	Routed flows (Mb/s)		
	<i>Black route</i>	<i>Medium – grey route</i>	<i>Light – grey route</i>
0.0000	1264.19	1007.99	727.82
0.0060	1221.77	1027.55	750.68
0.0108	1201.44	1036.88	761.68
0.0192	1174.66	1049.13	776.22
0.0336	1140.13	1064.85	795.02
0.0600	1096.84	1084.46	818.70
0.1068	1044.05	1108.22	847.73
0.1896	981.30	1136.27	882.43
0.3372	909.07	1168.31	922.62
0.6000	853.84	1176.75	969.40
1.0668	822.07	1153.00	1024.93
1.8972	788.30	1126.39	1085.31
3.3738	759.05	1094.43	1146.52
6.0000	790.22	1063.55	1146.23

The robustness of the computed solutions is now assessed by evaluating the congestion values that are obtained when 1000 different random cost scenarios are produced. Each random scenario is represented by a set of a and b parameters generated within the uncertainty region defined by nominal (\hat{a} and \hat{b}) and upper bound ($\hat{a} + \bar{a}$ and $\hat{b} + \bar{b}$) parameters. The random generation process is precisely described in Section 6.2.

For each random scenario and each value of Γ , we consider the corresponding robust solution and compute its actual congestion cost according to the realizations of a and b . Thus, for each Γ -solution we obtain 1000 different congestion costs. For those 1000 congestion values, we can find statistics like average and percentiles, portrayed in plotting form in Figure 4(a) and Figure 4(b). In both figures, the X axis represents the average congestion costs over the 1000 scenarios whereas the Y axis shows the 100th percentile, to which in the remainder of the paper we often refer as *worst-case scenario or value* (see Figure 4(a)) or 95th percentile (see Figure 4(b)), when Γ is increased from 0 to $|\mathcal{A}|$. All values on both axis are normalized with respect to those of the nominal problem solution ($\Gamma = 0$), which is always represented by the point (1, 1). In this way the relative variation between robust and nominal solutions is clearly pointed out: $(X, Y) = \left(\frac{X^\Gamma}{X^0}, \frac{Y^\Gamma}{Y^0} \right)$.

The idea behind the plots of Figure 4 is to show the trade-off between worst-case values and average values when the robustness parameter Γ is varied. It is perhaps surprising to observe that already with a small values of robustness, the solution that is returned by the PRC formulation outperforms the solution of the nominal problem both in terms of expected total cost and in terms of its 95th percentile (or even 100th percentile). For larger values of robustness, the PRC solutions seem to identify a Pareto frontier of compromises that needs to be made between expected total cost and its larger percentiles. Indeed, the expected cost increases as the solution becomes more robust and worst-case congestion is further minimized. This behavior was actually confirmed on an extensive set of experiments presented in Section 6.

5.2 The relevance of non-linear perturbations

This section is devoted to providing insights about the use of the parametric delay function presented in Equation (9), and how it permits to better characterize link delay uncertainty when integrated to the PRC formulation. In particular, we provide evidence that an additive perturbation model is insufficient and

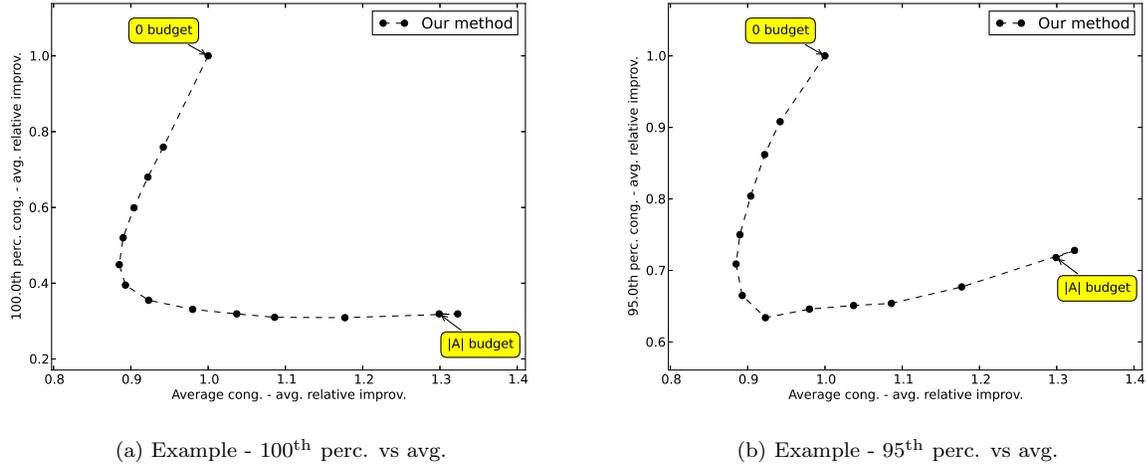


Figure 4: Results with example network topology.

illustrate the main shortcoming associated to linearization schemes as the TSEA approach presented in Section 4.4 when approximating a robust optimization that involve nonlinear perturbation effects.

This illustrative example is inspired by the delay uncertainty that can be perceived in the measurements presented in Figure 1. We actually refer the reader to Figure 5(a) where the same measurements are presented using units that are more closely aligned with our application of interest: namely, in our framework a delay function maps the amount of flow (in Mb/s) to the delay that is experienced per Megabit of data.³ Based on this figure, it is possible to calibrate our nominal weighted delay function $h(x, \hat{a}, \hat{b}, \hat{d}, \hat{e})$ using the diamond shaped points. These points were visually estimated to the 10 pairs of values reported in Table 3, which are plotted in Figure 5(b) under the label “Nominal-figure”, and which led the following choice of parameters

$$\hat{a} := 2.65, \quad \hat{b} := 0, \quad \hat{d} := 0.14, \quad \hat{e} := 0.$$

Note that in order to perform a meaningful calibration we chose to minimize the sum of relative absolute deviation in delay predictions, namely the following optimization model:

$$\min_{a,b,d,e} \sum_{m=1}^M \frac{|y_m - h(x_m, a, b, d, e) \cdot (\sum_{d \in \mathbf{D}} \rho^d / x_m)|}{y_m},$$

where $h(x_m, a, b, d, e) \cdot (\sum_{d \in \mathbf{D}} \rho^d / x_m)$ converts our weighted link delay function into the actual link delay.

A similar procedure (see Table 3 for the points) was used to characterize an upper bound on the link delay function to obtain the following parameters:

$$\bar{a} := 0.5, \quad \bar{b} := 2, \quad \bar{d} := 0, \quad \bar{e} := 250.$$

This calibrated upper bounding function is presented in Figure 5(b) under the label “Upper-region” while the calibrated nominal function was labeled “Lower-region”. Both appear to capture fairly accurately the

Table 3: Points representing the *Nominal-figure* and the *Upper-figure* plots visually derived from Figure 5(a).

	$\{y_m\}_{m=1}^1$	0	4.5	9	13.5	18	22.5	27	31.5	36	40.5	44.5
<i>Nominal-figure</i>	$\{x_m\}_{m=1}^1$	0	25	25	25	25	25	50	100	250	500	2500
<i>Upper-figure</i>	$\{x_m\}_{m=1}^1$	0	250	300	325	350	550	825	1075	1500	7500	∞

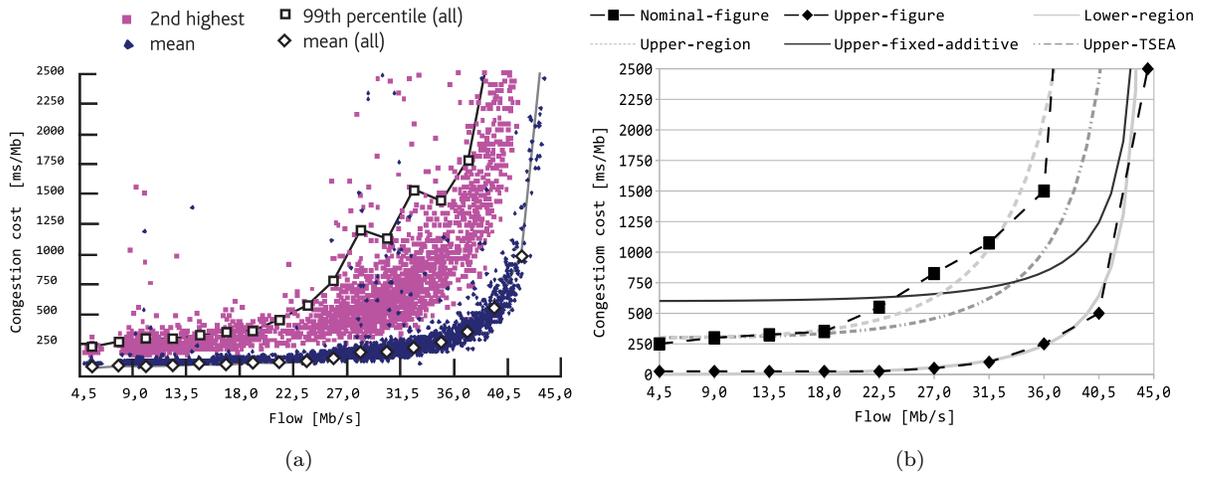


Figure 5: Comparison of uncertainty regions produced through different approaches. (a) Delay measurements of C. Van Eijl Eijl (2002), (b) uncertainty regions captured by different methods. Figure (a), which is taken from Hijazi et al. (2013), was originally published in Eijl (2002). With kind permission from Springer Science and Business Media.

Table 4: Functions used in Figure 5(b).

ID	Expression	Description
Nominal-figure	Data-driven ⁴	Series of points representing the typical amount of link delay
Upper-figure	Data-driven ⁴	Series of points representing an upper bound on the link delay
Lower-region	$\left(\frac{\sum_{d \in \mathbf{D}} \rho^d}{x}\right) \cdot \left(\hat{d} \frac{(1+x)^{1+\hat{a}} - 1}{\sum_{d \in \mathbf{D}} \rho^d (c-x-\hat{b})} + \hat{e} \frac{x}{\sum_{d \in \mathbf{D}} \rho^d}\right)$	Function that approximates the nominal link delay function according to our parameterized weighted delay function
Upper-region	$\left(\frac{\sum_{d \in \mathbf{D}} \rho^d}{x}\right) \cdot \left((\hat{d} + \bar{d}) \frac{(1+x)^{1+\hat{a}+\bar{a}} - 1}{\sum_{d \in \mathbf{D}} \rho^d (c-x-\hat{b}-\bar{b})} + (\hat{e} + \bar{e}) \frac{x}{\sum_{d \in \mathbf{D}} \rho^d}\right)$	Function that approximates the upper bound on link delay according to our parameterized weighted link delay function
Upper-fixed-additive	$\left(\frac{\sum_{d \in \mathbf{D}} \rho^d}{x}\right) \cdot \left(\hat{d} \frac{(1+x)^{1+\hat{a}} - 1}{\sum_{d \in \mathbf{D}} \rho^d (c-x-\hat{b})} + \hat{e} \frac{x}{\sum_{d \in \mathbf{D}} \rho^d}\right) + \kappa$	Function that approximates the upper bound on link delay when employing an additive delay perturbation as in Hijazi et al. (2013)
Upper-TSEA	$\left(\frac{\sum_{d \in \mathbf{D}} \rho^d}{x}\right) \cdot \tilde{h}(x, \hat{a} + \bar{a}, \hat{b} + \bar{b}, \hat{d} + \bar{d}, \hat{e} + \bar{e})$	Function that approximates the upper bound on link delay when employing TSEA
Parameter	Description	Value
\hat{a}	Nominal a parameter	2.65
\bar{a}	Maximum deviation for a parameter	0.5
\hat{b}	Nominal b parameter	0 Mbps
\bar{b}	Maximum deviation for b parameter	2 Mbps
\hat{d}	Nominal d parameter	0.14
\bar{d}	Maximum deviation for d parameter	0
\hat{e}	Nominal e parameter	0 s
\bar{e}	Maximum deviation for e parameter	0.3 s
r	Average number of packet in 1 Mb	250
c	Link capacity (as in Eijl (2002))	45 Mbps
κ	Maximum additive delay perturbation	0.6 s

range of values that the delay might take for any amount of flow that is passed through the link. Looking at the calibrated parameters, it is clear that the parameters that affect delay nonlinearly, namely a and b , play an important role in characterizing the uncertainty. Alternatively, one might try only using an additive noise model and obtain the upper bound function labeled “Upper-fixed-additive” which is clearly inadequate. Indeed, this upper bound for the uncertainty region overestimates the worst-case delay when the flow is small while underestimating it when the flow is large.

Finally, the curve labeled “Upper-TSEA” captures the upper bound of the uncertainty region that is used when employing a linearization scheme based on a first-order Taylor series expansion of $h(x, a, b, d, e)$. One might observe here that the upper bound of TSEA for the uncertainty region is more accurate than the additive perturbation model, especially for small values of x , but that it quickly ends up underestimating the worst-case delay as the amount of flow increases. It is worth noting however that compared to the PRC formulation such an approximation scheme might more appropriately capture the correlations that could exist between different link delay functions assuming that one can design an uncertainty set that captures well how all parameters in $\{(a_{ij}, b_{ij}, d_{ij}, e_{ij})\}$ behave jointly. For completeness, we conclude this discussion with the definition of a TSEA approximation of the weighted link delay function in this context:

$$\begin{aligned} \tilde{h}(x, a, b, d, e) := & \hat{d} \frac{(1+x)^{1+\hat{a}} - 1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c - x - \hat{b})} + \hat{e} \frac{x}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}} + (a - \hat{a}) \frac{\ln(1+x)(1+x)^{1+\hat{a}}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c - x - \hat{b})} + \\ & + (b - \hat{b}) \frac{(1+x)^{1+\hat{a}} - 1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c - x - \hat{b})^2} + (d - \hat{d}) \frac{(1+x)^{1+\hat{a}} - 1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c - x - \hat{b})} + (e - \hat{e}) \frac{x}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}}. \end{aligned}$$

6 Numerical experiments

This section presents a study involving thousands of problem instances (categorized into 18 different test conditions) of congestion minimization problems. We believe this study allowed us to identify some important strengths and limitations of the PRC formulation, in particular when compared to the nominal problem and the TSEA approach. We start by describing carefully how the problem instances of this study were generated and how the solutions’ performances were statistically evaluated through Monte-Carlo simulations. Next, we discuss the most relevant observations that were made about these results, while highlighting during this process the merits of the PRC formulation. The computing platform used for the tests was an Intel i7 PC with 4 core and multi-thread 8x, equipped with 8Gb of RAM. Non-linear convex programs were solved with MINOS 5.51 and the AMPL language.

The 18 different tests are summarized in Table 5. Each ID number on the table represents a set of *instances* that are characterized by their topology, the number of nodes, edges, robustness parameters, etc. For each ID, 1000 random instances were generated so that $|V^e|$ edge nodes were randomly selected, traffic matrices were chosen according to a random distribution (more details below) and link capacities were also randomly assigned to respect, on average, the following proportions: (i) 8 Gbps (30%), (ii) 5 Gbps (30%), (iii) 2.4 Gbps (20%) and (iv) 2 Gbps (20%). Both upper bound (f_{ij}^+) and nominal functions (f_{ij}) of each network link were also randomly selected. For this purpose, for each link $(i, j) \in \mathcal{A}$, both nominal, i.e., \hat{a}_{ij} and \hat{b}_{ij} , and maximum deviation values, i.e., \bar{a}_{ij} and \bar{b}_{ij} , were generated for bounding the a and b parameters. Now, for each of the 1000 random instances, the model was solved for $\Gamma = 0, \dots, |\mathcal{A}|$, which provided 13 different solution per test instance (from the least to the most robust). Two alternative solution methods were used for comparison (details below).

Once the random instances and their results were calculated, the question was how to assess the robustness of the method. For this, a post-processing simulation-like approach was used in which 999 realizations of the random parameters a and b of each link were generated for each of the previously described 1000 instances. We call a single realization a *scenario* and for each one of them, we determined how much both nominal and robust solutions would have costed, in terms of congestion, according to the current values of a and b . Both average and percentile costs over the set of 999 scenarios and for each value of Γ were evaluated to determine the robustness of the PRC formulation and the TSEA approach.

Table 5: Description of test instances.

ID	Net	$ V - V^e $	$ \mathcal{A} $	a_{max}^+	b_{max}^+	σ	λ	ϖ	$iter$	$Region$
1	nobel	28-14	82	1	0.5	0.1	2	0.8	1000	1
2	nobel	28-14	82	1	0.5	0.1	2	0.8	1000	2
3	nobel	28-14	82	1	0.5	0.1	2	0.5	1000	1
4	nobel	28-14	82	1	0.5	0.1	2	0.5	1000	2
5	nobel	28-14	82	0.6	0.5	0.1	2	0.8	1000	1
6	nobel	28-14	82	0.6	0.5	0.1	2	0.8	1000	2
7	nobel	28-14	82	0.6	0.5	0.1	2	0.5	1000	1
8	nobel	28-14	82	0.6	0.5	0.1	2	0.5	1000	2
9	nobel	28-14	82	0.2	0.5	0.1	2	0.8	1000	1
10	nobel	28-14	82	0.2	0.5	0.1	2	0.8	1000	2
11	nobel	28-14	82	0.2	0.5	0.1	2	0.5	1000	1
12	nobel	28-14	82	0.2	0.5	0.1	2	0.5	1000	2
13	germany	50-25	176	1	0.5	0.1	2	0.8	200	1
14	germany	50-25	176	1	0.5	0.1	2	0.8	200	2
15	germany	50-25	176	0.6	0.5	0.1	2	0.8	200	1
16	germany	50-25	176	0.6	0.5	0.1	2	0.8	200	2
17	germany	50-25	176	0.2	0.5	0.1	2	0.8	200	1
18	germany	50-25	176	0.2	0.5	0.1	2	0.8	200	2

6.1 Instance generation details

As previously explained, an *instance* is characterized by: 1) a randomly generated topology and link capacity, 2) uncertainty region and 3) traffic demand. More details on those elements are given in the following subsections.

6.1.1 Network topologies and link capacities

Two network topologies were extracted from the very popular SNDLibrary from Orlowski et al. (2010), i.e., **nobel-eu** and **germany**. Column $|V| - |V^e|$ of Table 5 reports the total number of nodes and pure edge nodes, i.e., those with either source or destination of traffic, while column $|\mathcal{A}|$ represents the cardinality of the link set. For each random instance, $|V^e|$ edge nodes and link capacities were randomly selected.

6.1.2 Uncertain region

For each instance, the upper bound (f_{ij}^+) and the nominal functions (\hat{f}_{ij}) of each network link were randomly determined in the following manner. For each link $(i, j) \in \mathcal{A}$, both nominal, i.e., \hat{a}_{ij} and \hat{b}_{ij} , and maximum deviation values, i.e., \bar{a}_{ij} and \bar{b}_{ij} , were generated for a and b . We experimented with two different ways of generating these values. This gives rise to two different types of uncertainty regions.

Region 1: nominal values $\hat{a}_{ij}, \hat{b}_{ij} \forall (i, j) \in \mathcal{A}$ respectively generated according to the uniform distributions $U(0.01, a_{max}^+)$ and $U(0.01, b_{max}^+)$, while the maximum deviation values $\bar{a}_{ij} \forall (i, j) \in \mathcal{A}$ and $\bar{b}_{ij} \forall (i, j) \in \mathcal{A}$ followed respectively two uniform distributions defined by

$$U(0, \min(a_{max}^+ - \hat{a}_{ij}, \hat{a}_{ij}))$$

and

$$U(0, \min(b_{max}^+ - \hat{b}_{ij}, \hat{b}_{ij})).$$

Region 2: nominal and maximum deviation values of b generated as for *Region 1*, while each nominal $\hat{a}_{ij}, \forall (i, j) \in \mathcal{A}$ was generated according to $U(0.01, a_{max}^+/2)$, and each maximum deviation $\bar{a}_{ij}, \forall (i, j) \in \mathcal{A}$ was set equal to \hat{a}_{ij} . The values of b_{max}^+ reported in Table 5 were normalized w.r.t. the link capacity.

6.1.3 Traffic matrices

The traffic matrices provided by the SNDLibrary (one for each network) were modified as follows. For each random instance, all the demands that were not Origin-Destination were discarded. Then, traffic request values $\rho^d \forall d \in \mathbf{D}$ were chosen according to a Uniform distribution $U(1, 10)$, and the resulting traffic matrix was scaled for the maximum parameters which allowed to route all traffic demands with fully splittable routing while not consuming more than $\varpi (c_{ij} - \hat{b}_{ij} - \bar{b}_{ij})$. The higher ϖ (never larger than 1), the more loaded the network. The scaling parameter was computed by means of an LP formulation.

Note that the number of traffic demands $|\mathbf{D}|$ was related to the composition of the edge node set V^e , which determines which demands had to be discarded from the reference SNDLibrary traffic demand set.

6.2 Postprocessing scenario generation details

As previously mentioned, the post processing *scenarios* were obtained by randomly generating values a and b for each link. This was done according to a multivariate normal distribution centered on the nominal values. The standard deviation of each parameter a (or b) was \bar{a}_{ij}/λ (see Table 5). If a parameter was generated outside the uncertainty region $[\hat{a}_{ij}, \hat{a}_{ij} + \bar{a}_{ij}]$, it was remapped within the desired interval by considering the equivalence between the CDFs of the corresponding Normal distribution and the Uniform distribution $U(\hat{a}_{ij}, \hat{a}_{ij} + \bar{a}_{ij})$. Note that with $\lambda = 3$, on average 99.9% of the random parameters would lay within the uncertainty region. The a and b parameters of links generated by the same node were assumed to have a correlation $\sigma = 0.1$ while other correlations were set to zero.

6.3 Computational results

Extensive testing was performed for networks `nobel-eu` and `germany50`. As shown in Table 5, the problem parameters were progressively adjusted to assess the performance, under different conditions, of both the PRC formulation and the TSEA approach.

For each of the 1000 instances, an X-Y plot similar to the one shown in Figure 4 was created, to analyze the trade-off between average and worst-case congestion as a function of robustness. Differently from the illustrative example of Section 5, results are now averaged over the 1000 instances of each class. In other words, the following X-Y plots were obtained by averaging over the individual X-Y plots of the 1000 instances.

In Table 6 the average results obtained over the 1000 instances are presented for each instance ID. In particular, the best improvement is reported in terms of (i) worst-case, (ii) 95th percentile, (iii) 90th percentile and (iv) average congestion costs observed when the robust solutions are applied. The improvement is computed with respect to the congestion costs obtained when $\Gamma = 0$ (nominal problem). In this table we report the best improvement obtained by varying Γ . Note that such an improvement might be achieved by different values of Γ .

The last row of Table 6 demonstrates how robustness allows to significantly improve worst-case, up to 62.4%, 95th percentile, up to 10.5%, 90th percentile, up to 5.5% and even average congestion costs, up to 4.7%. The PRC formulation clearly outperforms, on average, the TSEA approach: by 15.8%, 0.9%, 0.1% and 0.2% in terms of, respectively, worst-case, 95th percentile, 90th percentile and average congestion.

A clear trend emerges from Table 6: the larger and more nonlinear the uncertainty region is, the larger the relative improvement produced by the PRC formulation will be. Furthermore, in terms of 95th percentile and 90th percentile congestion costs, the difference between the PRC formulation and the TSEA approach is less significant than the one registered for worst-case costs.

Table 6: Best improvements achieved by robust approaches.

<i>ID</i>	Worst-case		95 th perc.		90 th perc.		Average	
	PRC	TSEA	PRC	TSEA	PRC	TSEA	PRC	TSEA
1	62.4%	26.0%	10.5%	9.4%	5.3%	6.2%	4.7%	3.7%
2	29.0%	0.0%	7.4%	4.5%	5.0%	3.8%	1.3%	2.1%
3	31.0%	20.0%	8.3%	8.2%	4.6%	5.7%	1.6%	1.2%
4	14.0%	5.2%	8.2%	5.9%	5.5%	4.5%	0.4%	2.4%
5	47.3%	23.7%	5.8%	5.0%	2.3%	2.4%	3.2%	1.6%
6	20.7%	2.3%	2.5%	2.2%	1.1%	1.5%	1.1%	0.5%
7	7.5%	6.7%	1.5%	1.6%	0.8%	0.9%	0.3%	0.1%
8	4.2%	3.6%	1.3%	1.8%	0.7%	1.3%	0.0%	0.1%
9	11.5%	11.0%	2.4%	2.3%	0.9%	0.9%	0.9%	0.8%
10	10.4%	9.8%	1.9%	1.7%	0.7%	0.6%	0.7%	0.5%
11	0.5%	0.6%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%
12	0.4%	0.4%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%
13	76.1%	30.4%	13.4%	11.5%	7.4%	7.6%	5.7%	4.6%
14	46.1%	0.0%	9.3%	5.4%	6.6%	4.7%	2.2%	2.8%
15	67.9%	31.8%	8.8%	6.9%	3.8%	3.1%	3.8%	1.9%
16	34.5%	11.5%	3.0%	2.2%	1.4%	1.4%	1.1%	0.2%
17	28.2%	26.6%	3.7%	3.5%	1.3%	1.2%	1.5%	1.2%
18	20.6%	19.1%	2.3%	2.0%	0.9%	0.7%	1.0%	0.8%
average	28.5%	12.7%	5.0%	4.1%	2.7%	2.6%	1.6%	1.4%

These and other relevant results emerging from the general table are discussed in the next sections with the aid of the corresponding X-Y plots.

6.3.1 Worst-case analysis

For this analysis we considered ID class 5, which is characterized by (i) uncertainty regions of moderate size, i.e., $a_{max}^+ = 0.6$, generated according to *Region 1*, (ii) moderate traffic load, i.e. fully splittable routing is achieved without exceeding 80% of the worst-case link capacity $c - \hat{b} - \bar{b}$. Three different subfigures can be found in Figure 6. For both the PRC formulation and the TSEA approach, the plots show the trade-off between average and worst-case congestion costs as the robustness budget is increased from 0 to $|\mathcal{A}|$ (13 points). To represent worst-case cost we report 100th (see Figure 6(a)), 95th (see Figure 6(b)) and 90th (see Figure 6(c)) percentile congestion values. The reader is reminded that all percentiles and average congestion values were taken by averaging over 1000 test instances of the same class for each one of the Γ points. Furthermore, to better highlight the improvement w.r.t. the nominal solutions, both percentiles and averages were normalized with respect to the values obtained with zero budget, that is, with $\Gamma = 0$.

Examining Figure 6(a), it can be seen that the PRC formulation substantially reduces the worst-case congestion outcome as Γ is increased. Most importantly, the PRC formulation considerably improves the figures obtained by TSEA, achieving a worst-case congestion improvement close to 45% (whereas TSEA can get up to 25%). Furthermore, with the PRC formulation a clear trade-off between worst-case and average values can be observed: reducing worst-case congestion naturally causes a degradation of the average. Note that the degradation is very small (namely less than 3%) in the case of the PRC formulation.

It is worth pointing out that, due to the particular shape of the congestion function, that tends to infinite when the traffic reaches the available capacity, being robust with a very small Γ allows, in a first moment, to reduce as well the average congestion.

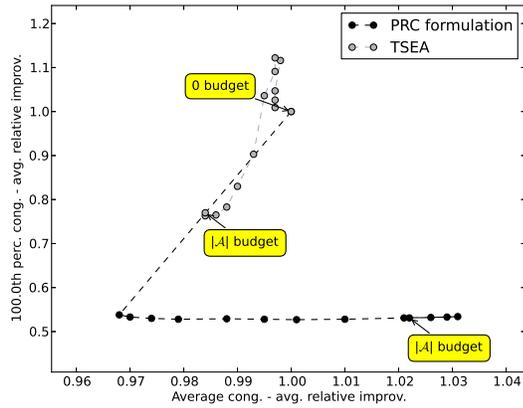
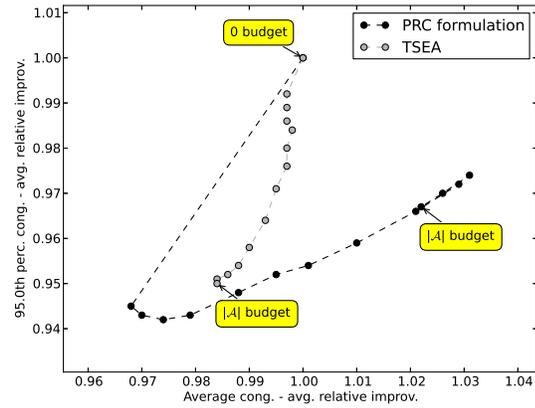
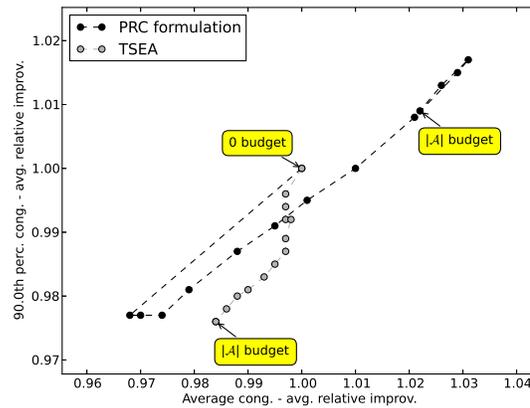
(a) Inst. 5 - 100th perc. vs avg.(b) Inst. 5 - 95th perc. vs avg.(c) Inst. 5 - 90th perc. vs avg.

Figure 6: Analyzing the impact of robustness on worst-case congestion. Uncertainty regions are of moderate size, i.e., $a_{max}^+ = 0.6$, a parameters are generated according to *Region 1*, traffic load is kept moderate, i.e. $\varpi = 0.8$. The notation *avg. relative improv.* used on both X and Y axis remarks that all X-Y coordinates represent the average variation (1000 test instances) with respect to both the X and Y normalized coordinates obtained with the nominal solutions. Nominal coordinates are always (1,1).

The 95th and 90th percentile plots for the PRC formulation show a very interesting behavior. At the beginning, as Γ increases from 0, the performance has a sharp improvement. However, as Γ moves towards its maximum value, a congestion deterioration is progressively observed. This phenomenon has a very simple explanation: to further reduce the worst-case figures (see Figure 6(a)), the model is forced to pay a price in terms of the 95th and 90th percentiles. This behavior is not observed with TSEA because the latter fails to improve the worst-case congestion up to a point where we can start to observe the aforementioned trade-off.

From these plots it can be affirmed that the PRC formulation clearly outperforms TSEA. What is more remarkable is that this is the case despite the fact that, unlike the TSEA approach, the PRC formulation does not require perfect knowledge of the parametric form of the delay function nor does it need detailed information about uncertainty in the parameters. This aspect is crucial because an exact characterization of a nominal function can be hardly derivable in many cases: the PRC formulation circumvent this issue by exploiting nominal and upper bound functions, both of which can be derived straightforwardly when analyzing historical data.

6.3.2 Impact of the uncertainty region nonlinearity

Here we want to show how the PRC formulation and the TSEA approach behave as uncertainty regions become wider and more nonlinear, i.e., by varying a_{max}^+ from 0.2 (instance 9) to 1 (instance 1) passing from 0.6 (instance 5). Link cost parameters were always generated according to *Region 1*. In Figure 7 the 95th percentile congestion values obtained while varying the robustness budget Γ from 0 to $|\mathcal{A}|$ are reported.

As expected, the wider and more nonlinear the uncertainty regions are, the higher the worst-case congestion improvement becomes. This trend is found for both methods: from 2.5% to 11% improvement with the PRC formulation and from 2% to 9.5% with the TSEA approach.

It is also worth pointing out that as a_{max}^+ grows and the uncertainty regions become more nonlinear, the gap between PRC and TSEA increases. This confirms that the first-order Taylor approximation works better when the uncertainty region is defined with less pronounced nonlinear terms. The PRC formulation,

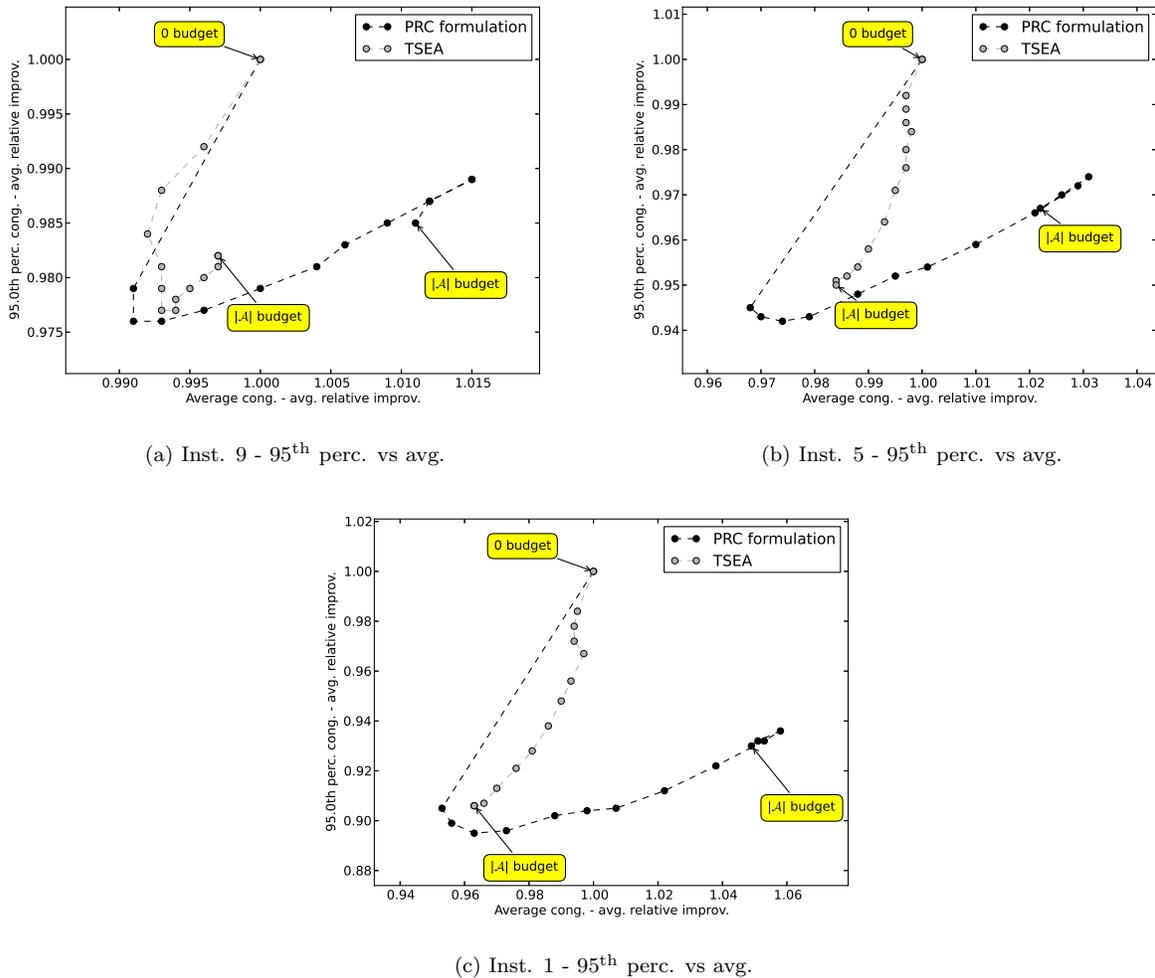


Figure 7: Analyzing the impact of the uncertainty region nonlinearity on congestion costs. Delay function parameters a_{max}^+ varying from 0.2 (a) to 0.6 (b) and again to 1 (c). Parameters are generated according to *Region 1*, traffic load is kept moderate, i.e. $\varpi = 0.8$. All figures show the 95th percentile versus average congestion trend. The notation *avg. relative improv.* used on both X and Y axis remarks that all X-Y coordinates represent the average variation (1000 test instances) with respect to both the X and Y normalized coordinates obtained with the nominal solutions. Nominal coordinates are always (1,1).

which is based on the convex combination of upper bound and nominal functions does not seem to be affected by such nonlinearities.

6.3.3 Sensitivity to network load

When the traffic load is higher, the link works close to its capacity limit $c - \hat{b} - \bar{b}$. Due to the nonlinearity of the delay function, the gap between nominal and real functions in this part of the uncertainty region can be much wider (see also Figure 5); thus, it is still more important to consider the possible deviation from the nominal curves, since a bad evaluation could lead, in this case, to a significant deterioration of the global congestion.

In Figure 8 the worst-case congestion curves obtained by varying the budget Γ for two different traffic scaling are compared. In Figures 8(a) and 8(c) traffic matrices were scaled to respect a capacity of $0.8 (c - \hat{b} - \bar{b})$ whereas Figures 8(b) and 8(d) show the results for a capacity scaling of $0.5 (c - \hat{b} - \bar{b})$. All parameters were generated according to *Region 1*.

As expected, when the traffic is higher, being robust is more important and may lead to a 95th percentile congestion improvement of 12% (e.g. in $a_{max}^+ = 1$, instance 1, Figure 8(a)) and 5.5% (e.e. in $a_{max}^+ = 0.6$, instance 5, Figure 8(c)). With a less congested network, the improvements are less significant and range from 8% (see Figure 8(b)) to 1.5% (see Figure 8(d)).

6.3.4 Sensitivity to the uncertainty region type

It is quite interesting to observe how the plot profiles change when the way the uncertainty region is generated changes. For *Region 1*, the generation scheme used to compute nominal and upper bound values for a and b provides a subset of link delay functions with very low nominal values but significant variability, and a second group where nominal delay is higher but variability is lower. In this case, the robust approaches achieve a clear trade-off between average and worst-case congestion cost, i.e., average costs are increased to further improve worst-case and 95th percentile costs (see the u-shape plots of Figure 6). In fact, as robustness (Γ) is increased, the robust approaches tend to prefer routing traffic on links with higher average delay but less uncertainty, rather than on those where delay is lower, on average, but more uncertain.

On the other hand, when uncertainty regions are generated according to *Region 2*, the situation is somehow different. Since \bar{a} is now equal to \hat{a} , the less uncertain links are also those with the lowest average delay. The X-Y plots drawn for *Region 2* instances show that as the uncertainty regions become larger (Instance 2 with $a_{max}^+ = 1$) the typical u-shape observed with *Region 1* is observed only for the first values of Γ . Starting from a certain $\Gamma = 0$, the trend previously observed is inverted such that increasing Γ allows the PRC formulation to improve both worst-case and average performance (see Figure 9). Note that if we consider Instance 2, the worst-case solution ($\Gamma = |\mathcal{A}|$) is even the best in terms of both average and 95th percentile congestion. Among the multiple interpretations which can be given to explain this behavior we provide the following two:

1. To achieve robustness in this case, it is necessary to find the best trade-off between the shortest paths in terms of hops and average delay and those shortest in terms of worst-case delay. When the budget Γ is small, the PRC formulation cannot distribute the uncertainty over the whole set of links, and the situation leads it to a traffic that is less splitter, which is less efficient in congestion terms. Once $\Gamma = |\mathcal{A}|$, traffic is maximally splitted due to the higher convexity of the delay function, and this routing configuration seems to perform quite well in terms of average delay costs too: a high splitting ratio results in a lower aggregation of traffic, so that a delay variation on a link has a lower impact in terms of overall congestion.
2. There is an approximation gap when a convex combination is used to represent real delay profiles. With this specific setting, this may prevent the PRC formulation from achieving the classical U-shape plot in the resulting X-Y graphs.

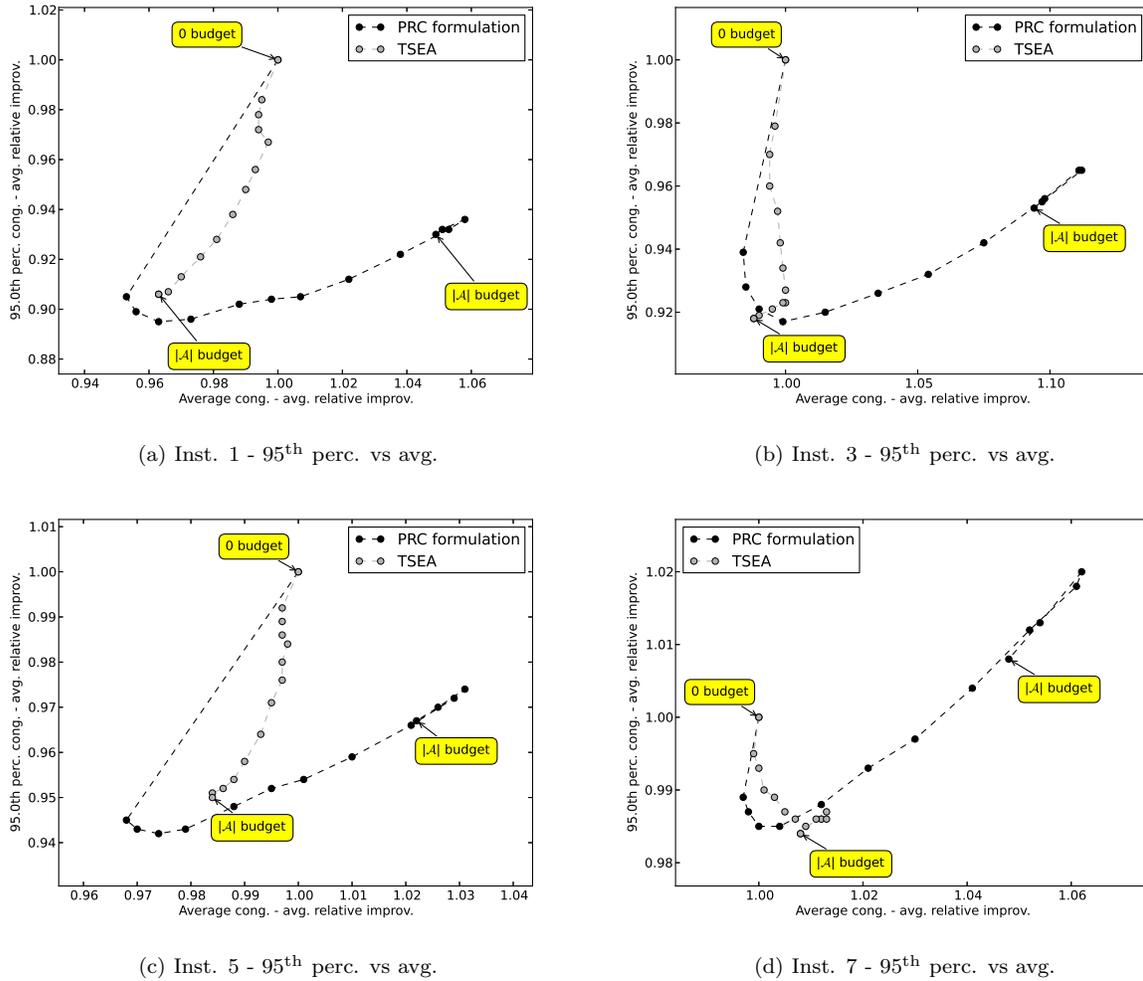


Figure 8: Analyzing the sensitivity to network load. Traffic load varies from moderate in plots (a) and (c), i.e., $\varpi = 0.8$, to low in plots (b) and (d), i.e., $\varpi = 0.5$. Delay function parameters a_{max}^+ are varied from 1 in plots (a) and (b) to 0.6 in plots (c) and (d). Delay parameters are generated according to *Region 1*. All plots show the 95th percentile versus average congestion trend. The notation *avg. relative improv.* used on both X and Y axis remarks that all X-Y coordinates represent the average variation (1000 test instances) with respect to both the X and Y normalized coordinates obtained with the nominal solutions. Nominal coordinates are always (1,1).

It is worth pointing out that, although X-Y plots are different according to the uncertainty region types, the PRC formulation always allows to substantially reduce worst-case and even average congestion. We thus believe that it can be adopted independently of the uncertainty region shapes. However, for each uncertainty setting, it is important to adjust $|\Gamma|$ to maximize the method performance.

6.3.5 Results with a larger network

To verify whether the considerations made for *nobel-eu* remained valid with a larger and different network, in Figure 10 we report six plots obtained using *germany50*, a network of 50 nodes and 176 links. From top down, $a_{max}^+ = 0.2$, $a_{max}^+ = 0.6$ and $a_{max}^+ = 1$, while from left to right the 95th percentile-average and the 90th percentile-average trade-offs are shown. The uncertainty regions were generated by according to *Region*

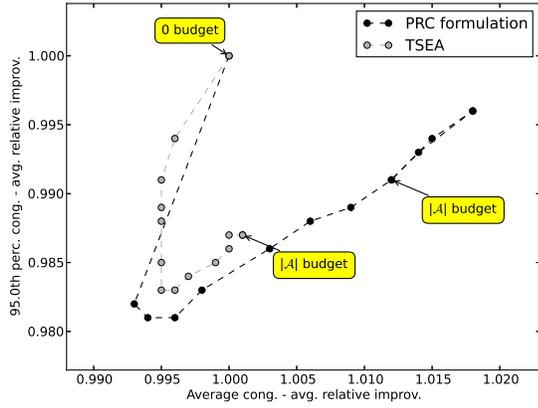
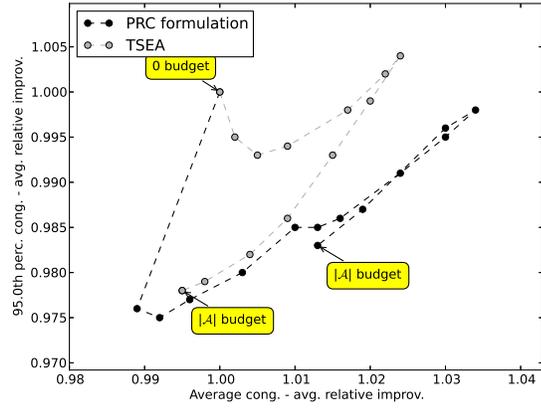
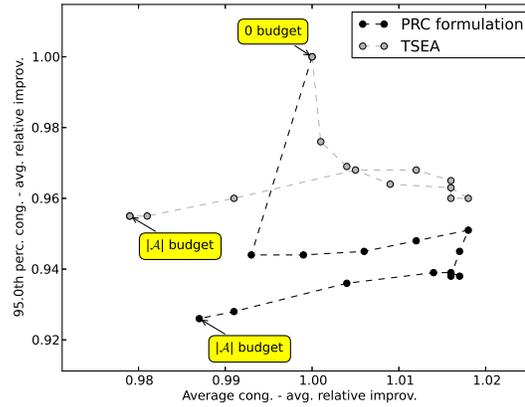
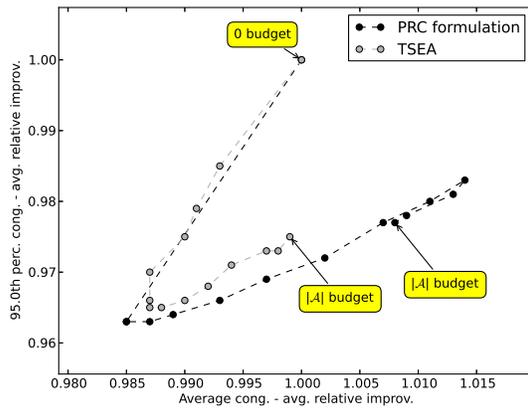
(a) Inst. 10 - 95th perc. vs avg.(b) Inst. 6 - 95th perc. vs avg.(c) Inst. 2 - 95th perc. vs avg.

Figure 9: Analyzing the worst-case versus average congestion trend when delay function parameters are generated according to *Region 2*. Delay function parameters a_{max}^+ is varying from 0.2 (a) to 0.6 (b) and again to 1 (c). Traffic load is kept moderate, i.e. $\varpi = 0.8$. All figures show the 95th percentile versus average congestion trend. The notation *avg. relative improv.* used on both X and Y axis remarks that all X-Y coordinates represent the average variation (1000 test instances) with respect to both the X and Y normalized coordinates obtained with the nominal solutions. Nominal coordinates are always (1,1).

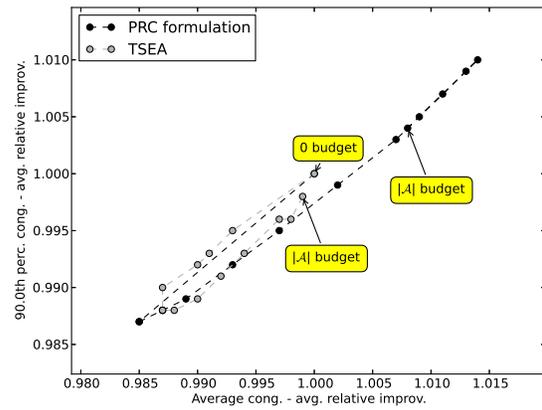
1. The plots reveal that the same trends are also observed with this network, which seems to confirm the quality of the PRC formulation.

7 Conclusion

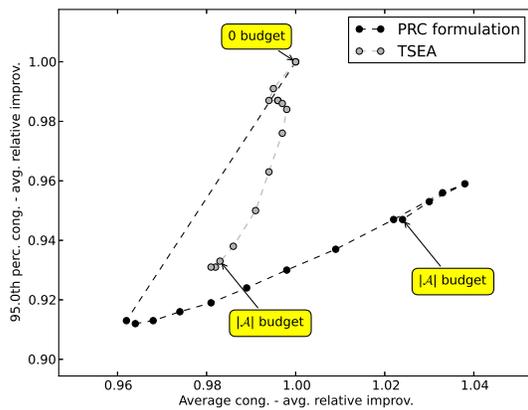
In this article, we have proposed a novel robust optimization formulation to handle optimization problems characterized by cost functions that decompose over each decision variables and behave nonlinearly with respect to both the parameters and the decisions. Our formulation, which we named Practicable Robust Counterpart formulation, only assumes that each term of the nominal cost function is convex with respect to the decision variable and can be bounded above by another convex function. The PRC formulation has the big advantage of reducing to a robust optimization problem that can easily be calibrated using historical data, and that ends up not being much harder to solve than the deterministic version of the problem. We apply the PRC formulation to the question of how to minimize network congestion through



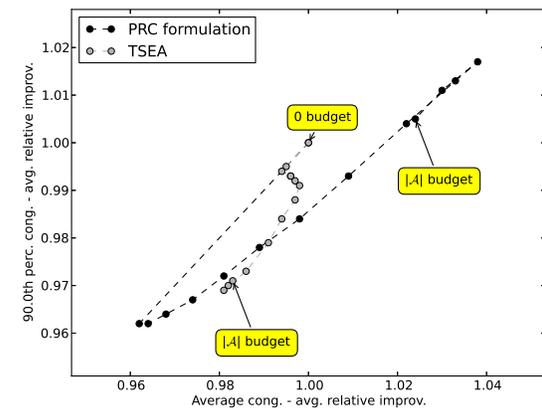
(a) Inst. 17 - 95th perc. vs avg.



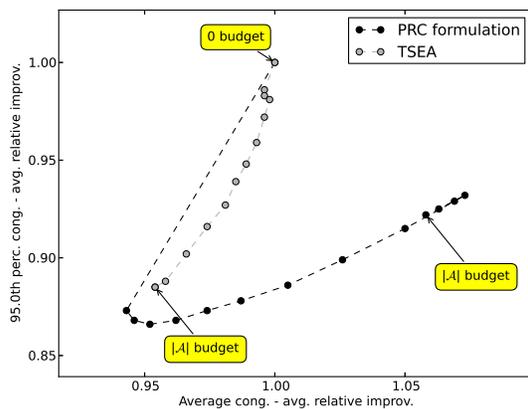
(b) Inst. 17 - 90th perc. vs avg.



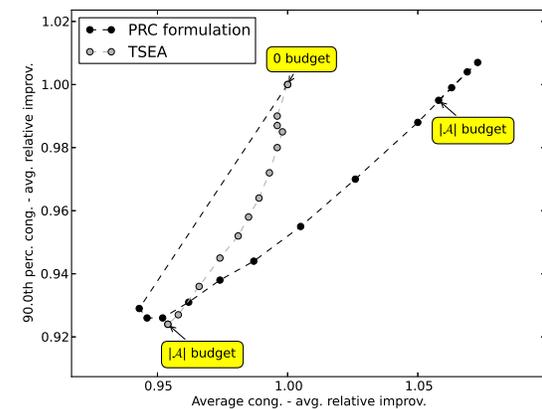
(c) Inst. 15 - 95th perc. vs avg.



(d) Inst. 15 - 90th perc. vs avg.



(e) Inst. 13 - 95th perc. vs avg.



(f) Inst. 13 - 90th perc. vs avg.

Figure 10: Analyzing the overall results with a network of 50 nodes, i.e., **germany50**. Delay function parameters a_{max}^+ varying from 0.2 (a-b) to 0.6 (c-d) and again to 1 (e-f). Traffic load is kept moderate, i.e. $\varpi = 0.8$. Parameters are generated according to *Region 1*. Plots (a-c-e) show the 95th percentile versus average congestion trend, while plots (b-d-f) the 90th percentile one. The notation *avg. relative improv.* used on both X and Y axis remarks that all X-Y coordinates represent the average variation (1000 test instances) with respect to both the X and Y normalized coordinates obtained with the nominal solutions. Nominal coordinates are always (1,1).

multi-commodity fully splittable routing when transmission delays on each link are considered uncertain and to be affected nonlinearly by the link's congestion level. Our numerical experiments demonstrate that the PRC formulation is able to identify routing strategies that can substantially outperform a strategy that is based on a deterministic approach, both with respect to congestion level that are achieved on average and in sets of unfavorable conditions: i.e. up to 5.7% improvement of the expected value or up to 13.4% of the 95th percentile. Moreover, we demonstrate that the PRC formulation outperforms RO formulations that rely on naive linearization schemes, such as a using first-order Taylor series approximation.

It is worth noting that in this article our discussion focused on emphasizing the contribution of the PRC formulation in the context of classical robust optimization theory which assumes a model of the form of problem (1). However, it is our firm opinion that in practice there are many problems for which knowledge of a cost function naturally takes the shape of two bounds for $h_i(x_i)$, i.e. $\hat{f}_i(x_i) \leq h_i(x_i) \leq f_i^+(x_i)$, without any obvious association to a parametric form for $h_i(x_i)$ whose parameters would be uncertain. While we mentioned in a few occasion that the hypothesis made by the PRC formulation makes it especially well adapted for such situations (see for instance Remark 1), there are many interesting research questions that remain open. We mentioned for instance some issues regarding the calibration of Γ when individual historical cost measurements of the form $\{(x_i^j, h_i(x_i^j))\}$ are used. Yet, when one can only rely on historical measurements, there is even the issue of choosing the proper expressions for \hat{f}_i and f_i^+ based on the data. Alternatively, bounds on the terms of the total cost function might also naturally arise in contexts where there are conflicting theories, or experts opinions, that could be used to predict the cost of a decision.⁵ Indeed, in this case, one can then consider the nominal cost function to be the average cost predicted by different theories, and the upper bound function to return, for any value of the decision, the largest cost predicted by any of the theories. It then becomes interesting to clarify what might be the role of \mathcal{U}_α in this context and how one might calibrate Γ in $\mathcal{U}_\alpha(\Gamma)$. Due to both the limited time on our hands and space consideration in this article, we choose, somehow with regret, to leave these questions as directions for future research.

Endnotes

1. Although a specific definition will be proposed in Section 3, the deviation measure \mathcal{D}_i should generally be a positive one such that only $h_i(\cdot, \hat{\xi})$ satisfies $\mathcal{D}(f_i) = 0$ in order to recuperate the nominal problem when $\alpha = 0$.

2. Assuming that $1 \leq x_{ij} < c_{ij} - \hat{b}_{ij} - \bar{b}_{ij}$, that $a_{ij} \geq 0$, and that $0 \leq b_{ij} \leq \hat{b}_{ij} + \bar{b}_{ij}$, one can confirm the convexity of $h_{ij}(\cdot)$ in a_{ij} and b_{ij} by studying the second derivatives:

$$\begin{aligned} \frac{d^2 h_{ij}}{da_{ij}^2} &= \frac{\ln(1+x_{ij})^2(1+x_{ij})^{1+a_{ij}}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c_{ij} - b_{ij} - x_{ij})} \geq 0 \\ \frac{d^2 h_{ij}}{db_{ij}^2} &= \frac{2(1+x_{ij})^{1+a_{ij}} - 2}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c_{ij} - b_{ij} - x_{ij})^3} \geq 0 \\ \frac{d^2 h_{ij}}{da_{ij} db_{ij}} &= \frac{\ln(1+x_{ij})(1+x_{ij})^{1+a_{ij}}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}} (c_{ij} - b_{ij} - x_{ij})^2}. \end{aligned}$$

So that the determinant of the Hessian is necessarily non-negative, i.e. $(d^2 h_{ij}/da_{ij}^2) \cdot (d^2 h_{ij}/db_{ij}^2) - (d^2 h_{ij}/da_{ij} db_{ij})^2 \geq 0$.

3. This was done assuming that the link capacity was 45 Mbps and that packets have an average size of 500 Bytes, i.e., 4000 bits. Since 250 packets of 4000 bits are contained in 1 Mbit, the packet delay values of Figure 1 are multiplied by 250 to obtain the average delay per Mbit.

4. Points extracted from the realistic plot provided in Eijl (2002) and reported in Figures 1 and 5(a).

5. One can take for example theories that make different hypotheses about the stochastic arrival and processing time of packets in a router in order to make conclusions about a router's transmission delay function.

Appendix A Proof of Proposition 3.1

This result is obtained by showing that the NP-complete 3-SAT problem can be reduced to verifying whether

$$\max_{\xi \in \mathcal{U}_\xi} \sum_{i=1}^n h_i(x_i, \xi) \geq 0$$

is true or not.

3-SAT Problem Let W be a collection of disjunctive clauses $W = \{w_1, w_2, \dots, w_M\}$ on a finite set of variables $V = \{v_1, v_2, \dots, v_N\}$ such that $|w_m| = 3 \forall i \in \{1, \dots, M\}$. Let each clause be of the form $w = v_i \vee v_j \vee \bar{v}_k$, where \bar{v} is the negation of v . Is there a truth assignment for V that satisfies all the clauses in W ?

Given an instance of the 3SAT Problem, we can construct the following optimization problem

$$\text{maximize}_{\xi} \quad \sum_{m=1}^M h_m(x_m, \xi) - 1 \quad (13a)$$

$$\text{subject to} \quad 0 \leq \xi_k \leq 1, \forall k \in \{1, \dots, N\}, \quad (13b)$$

where $h_m(x_m, \xi) := \max\{\xi_i; \xi_j; 1 - \xi_k\}$ if the m -th clause is $w_m = v_i \vee v_j \vee \bar{v}_k$. It is straightforward to confirm that \mathcal{U}_ξ is a polyhedron here (actually a simple box) and that each $h_m(x_m, \xi)$ is non-concave in ξ . More importantly, we have that the answer to the 3-SAT problem is positive if and only if the optimal value of problem (13) achieves an optimal value greater or equal to zero.

Appendix B Proof of Proposition 3.4

Let the feasible set \mathcal{X} be represented using the following set of convex constraints:

$$g_i(x) \leq 0 \forall i = 1, 2, \dots, M,$$

where each $g_i(x)$ is a convex function of x . Let also $\bar{h}(x) := \max_{\xi \in \mathcal{U}_\xi} \sum_i h_i(x_i, \xi)$ capture the objective function of problem (1). Given that, based on by Condition 1, it is well known that an optimal solution exists for problem (1), it is also well known that it must be possible to pair such an optimal solution x^* with a set of values λ_i^* that satisfies the Karush-Kuhn-Tucker conditions:

$$\begin{aligned} \lambda_i^* &\geq 0 \forall i = 1, 2, \dots, M \\ g_i(x^*) &\leq 0 \forall i = 1, 2, \dots, M \\ \lambda_i^* g_i(x^*) &= 0 \forall i = 1, 2, \dots, M \\ 0 &\in \partial \bar{h}(x^*) + \sum_i \lambda_i^* \partial g_i(x^*), \end{aligned}$$

where $\partial f_i(x)$ and $\partial g_i(x)$ are the sets of subgradients of $\bar{h}(x) := \max_{\xi \in \mathcal{U}_\xi} \sum_i h_i(x_i, \xi)$ and $g_i(x)$ respectively. Given Condition 1, we must have that the $\max_{\xi \in \mathcal{U}_\xi} \sum_i h_i(x_i^*, \xi) = \sum_i h_i(x_i^*, \xi^*)$ for some $\xi^* \in \mathcal{U}_\xi$. Since the maximum is achieved and given conditions 2 and 3, the set $\partial \bar{h}(x)$ must therefore be the convex hull of all $v(\bar{\xi}) := [h'_1(x_1^*, \bar{\xi}), h'_2(x_2^*, \bar{\xi}), \dots, h'_n(x_n^*, \bar{\xi})]^T$ vectors such that $\bar{\xi} \in \mathcal{U}_\xi$ achieves the supremum. Let $\nabla \bar{h}(x^*) \in \partial \bar{h}(x^*)$ be the right convex combination of these $v(\bar{\xi})$'s that together with the M -tuple $(\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_M(x^*))$, with each $\nabla g_i(x^*) \in \partial g_i(x^*)$, satisfies the following equality

$$\nabla \bar{h}(x^*) = - \sum_{i=1}^M \lambda_i \nabla g_i(x^*).$$

Since for any $\bar{\xi} \in \mathcal{U}_\xi$, Condition 4 ensures that the derivatives $h'_i(x_i^*, \bar{\xi})$ are in the interval $[\hat{f}'_i(x_i^*), f_i^{+'}(x_i^*)]$, the fact the i -th term of $\nabla \bar{h}(x^*)$, referred below as $\nabla_i \bar{h}(x^*)$, is a convex combination of these derivatives

ensures that if we construct some

$$\bar{\alpha}_i := \frac{\nabla_i \bar{h}(x_i^*) - \hat{f}_i'(x_i^*)}{f_i^{+'}(x_i^*) - \hat{f}_i'(x_i^*)}$$

we have that $0 \leq \bar{\alpha}_i \leq 1$ for all i .

We finalize the proof by showing that if one chooses \mathcal{U}_α to be such that $\bar{\alpha}$ is the unique optimizer of $\max_{\alpha \in \mathcal{U}_\alpha} \sum_i \alpha_i f_i^+(x_i^*) + (1 - \alpha) \hat{f}_i(x_i^*)$, then the optimal solution of problem (3) is necessarily x^* . Note that this condition is easily satisfied with $\mathcal{U}_\alpha := \{\alpha \in [0, 1]^n \mid 0 \leq \alpha_i \leq \bar{\alpha}_i\}$ when $f_i^+(x_i) > \hat{f}_i(x_i)$ for all $x \in \mathcal{X}$ but can also be satisfied in many other ways.

We now verify that $(x^*, \lambda_1^*, \dots, \lambda_m^*)$ satisfies the KKT conditions of problem (3) when $\bar{\alpha}$ is the unique optimizer of $\max_{\alpha \in \mathcal{U}_\alpha} \sum_i \alpha_i f_i^+(x_i^*) + (1 - \alpha) \hat{f}_i(x_i^*)$. In fact, only the last of the KKT condition must be verified. Since $\bar{\alpha}$ is the unique optimizer, the condition reduces to

$$0 \in \begin{bmatrix} \bar{\alpha}_1 \hat{f}_1(x_1^*) + (1 - \bar{\alpha}_1) f_1^{+'}(x_1^*) \\ \bar{\alpha}_2 \hat{f}_2(x_2^*) + (1 - \bar{\alpha}_2) f_2^{+'}(x_2^*) \\ \vdots \\ \bar{\alpha}_n \hat{f}_n(x_n^*) + (1 - \bar{\alpha}_n) f_n^{+'}(x_n^*) \end{bmatrix} + \sum_i \lambda_i^* \partial g_i(x^*),$$

which is equivalent to

$$0 \in \nabla \bar{h}(x^*) + \sum_i \lambda_i^* \partial g_i(x^*).$$

The latter is known to be satisfied using the previously constructed M -tuple $(\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_M(x^*))$.

Given that Condition 2 implies that the functions $f_i^+(x_i)$ and $\hat{f}_i(x_i)$ are strictly convex in x_i , it is therefore the case that the objective function of problem (3) is strictly convex and that its optimal solution is unique. This allows us to confirm that x^* is necessarily the only solution to the KKT conditions and will therefore necessarily be returned when solving problem (3). This concludes our proof. \square

Appendix C Conjugate functions needed for Corollary 3.1

When the uncertainty set for α takes the form of $\mathcal{U}_\alpha(\Gamma)$, one can simply use problem (6) with $B = \begin{bmatrix} 1 \\ -1 \end{bmatrix} [1 \ 1 \ \dots \ 1]$ and $b = [\Gamma \ -\Gamma]^T$. Note also that problem (6) reduces to

$$\begin{aligned} & \underset{x \in \mathcal{X}, y, z, u^+, u^-}{\text{minimize}} && \sum_i y_i + \Gamma(u^+ - u^-) \\ & \text{subject to} && y_i \geq \hat{f}_i(x) \\ & && y_i + z_i \geq f_i^+(x) \\ & && u^+ - u^- = z_i, \forall i \\ & && u^+ \geq 0, u^- \geq 0, \end{aligned}$$

where $u^+ \in \mathbb{R}$ and $u^- \in \mathbb{R}$. One can then simply use $u \in \mathbb{R}$ to replace $u^+ - u^-$ and each z_i thus making the problem reduce to problem (6).

Appendix D Verifying Conditions 1 to 4 for Problem (11)

Indeed, while conditions 1 and 2 can be directly verified, the last two can be confirmed by studying all the relevant derivatives. First, since

$$\frac{dh_{ij}(x_{ij}, a_{ij}, b_{ij})}{dx_{ij}} = \frac{(1 + a_{ij})(1 + x_{ij})^{a_{ij}}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})} + \frac{(1 + x_{ij})^{1+a_{ij}} - 1}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})^2}$$

$$\frac{d^2h_{ij}(x_{ij}, a_{ij}, b_{ij})}{dx_{ij}^2} = \frac{a_{ij}(1 + a_{ij})(1 + x_{ij})^{a_{ij}-1}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})} + \frac{2(1 + a_{ij})(1 + x_{ij})^{a_{ij}}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})^2} + \frac{2(1 + x_{ij})^{1+a_{ij}} - 2}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})^3},$$

we therefore have that \hat{f}_{ij}^- and f_{ij}^+ are both convex since in both cases \hat{a}_{ij} and $\hat{a}_{ij} + \bar{a}_{ij}$ are greater than zero and since each x_{ij} belongs to the interval $[0, c_{ij} - \hat{b}_{ij} - \bar{b}_{ij}]$. Finally, Condition 4 is also satisfied with $\mathcal{U}_{ab}(\Gamma)$ since we have that

$$\frac{d^2h_{ij}(x_{ij}, a_{ij}, b_{ij})}{dx_{ij} da_{ij}} = \frac{(1 + (1 + a_{ij}) \ln(1 + x_{ij}))(1 + x_{ij})^{1+a_{ij}}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})} + \frac{\ln(1 + x_{ij})(1 + x_{ij})^{1+a_{ij}}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})^2} \geq 0$$

$$\frac{d^2h_{ij}(x_{ij}, a_{ij}, b_{ij})}{dx_{ij} db_{ij}} = \frac{a_{ij}(1 + a_{ij})(1 + x_{ij})^{a_{ij}-1}}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})^2} + \frac{2(1 + x_{ij})^{1+a_{ij}} - 2}{\sum_{\mathbf{d} \in \mathbf{D}} \rho^{\mathbf{d}}(c_{ij} - b_{ij} - x_{ij})^3} \geq 0,$$

for similar reasons.

References

- Addis, B., A. Capone, G. Carello, L.G. Gianoli, B. Sanso. 2013. A robust optimization approach for energy-aware routing in mpls networks. IEEE 2013 International Conference on Computing, Networking and Communications (ICNC). 567–572.
- Addis, B., A. Capone, G. Carello, L.G. Gianoli, B. Sansò. 2014. On the energy cost of robustness and resiliency in IP networks. Computer Networks 75, Part A(0) 239–259.
- Altın, A., E. Amaldi, P. Belotti, M. Pınar. 2007. Provisioning virtual private networks under traffic uncertainty. Networks 49(1) 100–115.
- Altın, A., B. Fortz, H. Ümit. 2012. Oblivious ospf routing with weight optimization under polyhedral demand uncertainty. Networks 60(2) 132–139.
- Ardestani-Jaafari, A., E. Delage. 2014. The value of flexibility in robust location-transportation problem. Les Cahiers du GERAD G-2014-83, GERAD, HEC Montréal.
- Ben-Ameur, W., H. Kerivin. 2005. Routing of uncertain traffic demands. Optimization and Engineering 6(3) 283–313.
- Ben-Tal, A., D. den Hertog. 2011. Immunizing conic quadratic optimization problems against implementation errors. Tech. rep., CentER Discussion Paper Series No. 2012-053.
- Ben-Tal, A., D. den Hertog, J-P Vial. 2012. Deriving robust counterparts of nonlinear uncertain inequalities. Tech. rep., CentER Discussion Paper Series No. 2012-053.
- Ben-Tal, A., L. El Ghaoui, H. Lebret. 1998. Robust semidefinite programming. Handbook on Semidefinite Programming 27.
- Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. Robust Optimization. Princeton University Press.
- Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. Mathematics of Operations Research 23(4) 769–805.
- Ben-Tal, A., A. Nemirovski. 1999. Robust solutions of uncertain linear programs. Operations research letters 25(1) 1–13.
- Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. Mathematical Programming 88(3) 411–424.
- Ben-Tal, A., A. Nemirovski. 2002. Robust optimization—methodology and applications. Mathematical Programming 92(3) 453–480.
- Ben-Tal, A., A. Nemirovski. 2003. On approximate robust counterparts of uncertain semidefinite and conic quadratic programs. System Modeling and Optimization XX 130 1–22.
- Bertsimas, D., D.B. Brown, C. Caramanis. 2011. Theory and applications of robust optimization. SIAM Review 53(3) 464–501.
- Bertsimas, D., O. Nohadani, K.M. Teo. 2010. Robust optimization for unconstrained simulation-based problems. Operations Research 58(1) 161–178.
- Bertsimas, D., M. Sim. 2003. Robust discrete optimization and network flows. Mathematical Programming 98(1) 49–71.

- Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* 52(1) 35–53.
- Calafiore, G., M.C. Campi. 2005. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming* 102(1) 25–46.
- Coudert, D., A.M. Koster, T.K. Phan, M. Tieves. 2013. Robust redundancy elimination for energy-aware routing. 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing. GreenCom, IEEE, 179–186.
- Dantzig, G.B. 1955. Linear programming under uncertainty. *Management Science* 1(3-4) 197–206.
- Eijl, C.A. Van. 2002. Capacity planning for carrier-scale ip networks. *BT Technology Journal* 20(3) 116–123.
- El Ghaoui, L., H. Lebret. 1997. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications* 18(4) 1035–1064.
- Fortz, B., M. Thorup. 2004. Increasing internet capacity using local search. *Computational Optimization and Applications* 29 13–48.
- Gabrel, V., C. Murat, A. Thiele. 2014. Recent advances in robust optimization: An overview. *European Journal of Operational Research* 235(3) 471–483.
- Gendreau, M., B. Sansò, D.A. Stanford. 1996. Optimizing routing in Packet-switched networks with Non-Poisson offered traffic. *Telecommunication Systems-Modeling, Analysis, Design and Management* 5(2) 323–340.
- Grötschel, M., L. Lovász, A. Schrijver. 1981. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica* 1 169–197.
- Hijazi, H., P. Bonami, A. Ouorou. 2013. Robust delay-constrained routing in telecommunications. *Annals of Operations Research* 206(1) 163–181.
- Hsiung, K.L., S.J. Kim, S. Boyd. 2008. Tractable approximate robust geometric programming. *Optimization and Engineering* 9(2) 95–118.
- Li, M., S.A. Gabriel, Y. Shim, S. Azarm. 2011. Interval uncertainty-based robust optimization for convex and non-convex quadratic programs with applications in network infrastructure planning. *Networks and Spatial Economics* 11(1) 159–191.
- Mattia, S. 2013. The robust network loading problem with dynamic routing. *Computational Optimization and Applications* 54(3) 619–643.
- Mulyana, E., U. Killat. 2005. Optimizing IP networks for uncertain demands using outbound traffic constraints. *Proc. of 2nd INOC*. 695–701.
- Orlowski, S., R. Wessälly, M. Pióro, A. Tomaszewski. 2010. Sndlib 1.0 survivable network design library. *Networks* 55(3) 276–286.
- Ouorou, A. 2011. Affine decision rules for tractable approximations to robust capacity planning in telecommunications. *Network Optimization. Lecture Notes in Computer Science* 6701 277–282.
- Ouorou, A. 2013. Tractable approximations to a robust capacity assignment model in telecommunications under demand uncertainty. *Computers & Operations Research* 40(1) 318–327.
- Plante, M., B. Sansò. 2002. A typology for multi-technology, multi-service broadband network synthesis. *Telecommunication Systems* 19(1) 39–73.
- Poss, M., C. Raack. 2013. Affine recourse for the robust network design problem: Between static and dynamic routing. *Networks* 61(2) 180–198.
- Restrepo, J.C.C., C.G. Gruber, C.M. Machuca. 2009. Energy profile aware routing. *IEEE International Conference on Communications Workshops*. 1–5.
- Sansò, B., H. Mellah. 2009. On reliability, performance and Internet power consumption. *DRCN09, Seventh International Workshop of the Design of Reliable Communication Networks*. 259–264.
- Soyster, A.L. 1973. Technical noteconvex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research* 21(5) 1154–1157.
- Takeda, A., S. Taguchi, T. Tanaka. 2010. A relaxation algorithm with a probabilistic guarantee for robust deviation optimization. *Computational Optimization and Applications* 47(1) 1–31.
- Zeng, B., L. Zhao. 2013. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters* 41(5) 457–461.