On convex lower-level black-box constraints in bilevel optimization with an application to gas market models with chance constraints

Holger Heitsch[1], René Henrion[1], Thomas Kleinert[2,3], Martin Schmidt[4]

submitted: April 12, 2021

[1] Weierstrass Institute
    Mohrenstr. 39
    10117 Berlin
    Germany
    E-Mail: holger.heitsch@wias-berlin.de
            rene.henrion@wias-berlin.de

[2] Friedrich-Alexander-Universität
    Erlangen-Nürnberg
    Discrete Optimization
    Cauerstr. 11
    91058 Erlangen
    Germany
    E-Mail: thomas.kleinert@fau.de

[3] Energie Campus Nürnberg
    Fürther Str. 250
    90429 Nürnberg
    Germany

[4] Trier University
    Department of Mathematics
    Universitätsring 15
    54296 Trier
    Germany
    E-Mail: martin.schmidt@uni-trier.de

On convex lower-level black-box constraints in bilevel optimization with an application to gas market models with chance constraints

Holger Heitsch, René Henrion, Thomas Kleinert, Martin Schmidt

Abstract

Bilevel optimization is an increasingly important tool to model hierarchical decision making. However, the ability of modeling such settings makes bilevel problems hard to solve in theory and practice. In this paper, we add on the general difficulty of this class of problems by further incorporating convex black-box constraints in the lower level. For this setup, we develop a cutting-plane algorithm that computes approximate bilevel-feasible points. We apply this method to a bilevel model of the European gas market in which we use a joint chance constraint to model uncertain loads. Since the chance constraint is not available in closed form, this fits into the black-box setting studied before. For the applied model, we use further problem-specific insights to derive bounds on the objective value of the bilevel problem. By doing so, we are able to show that we solve the application problem to approximate global optimality. In our numerical case study we are thus able to evaluate the welfare sensitivity in dependence of the achieved safety level of uncertain load coverage.

1   Introduction

Bilevel optimization has become more and more important during the last years and decades. The reason is that this class of problems enables the modeler to describe hierarchical decision processes, which is often of great importance in practical applications. On the other hand, this ability also makes bilevel problems hard to solve—both in theory and practice. For instance, it is well known that even the easiest class of bilevel problems, i.e., models with linear lower and upper level, are strongly NP-hard [22] and even checking the local optimality of a given point is NP-hard [50]. Nevertheless, starting in the 1980's, researchers developed increasingly powerful algorithms to tackle different challenging classes of bilevel problems; see, e.g., [8, 9, 28] for an overview of the field. The most recent trend is to consider more and more challenging bilevel models by also incorporating (mixed-)integer aspects [11, 15, 16, 39, 53], continuous nonconvexities [37, 38], or uncertainty modeling [6, 7, 42, 55]. The resulting bilevel models usually are harder to solve if these complicating aspects appear in the lower level of the bilevel problem. For instance, a convex lower-level problem (that also satisfies a reasonable constraint qualification) can be replaced by its necessary and sufficient optimality conditions, leading to an equivalent single-level reformulation of the bilevel problem. This is of course not possible anymore if the lower-level problem does not possess compact optimality certificates as it is the case for mixed-integer or other nonconvex problems.

One main field of application for bilevel optimization are energy markets, where one often is faced with a hierarchical structure of decision making; see, e.g., the recent survey [52] of bilevel optimization applied in the energy sector. In this paper, we consider a simplified version of the European entry-exit gas market [13, 14], which can be modeled as a bilevel problem. The main goal of this market design is

to decouple the trade and transport of natural gas. This is also reflected in the bilevel structure since the transport-related aspects are incorporated in the upper level, in which the transmission system operator (TSO) acts, while anticipating the market outcome that is determined in the lower-level problem. We refer to [21] for the derivation of this bilevel model and to [5, 46] for some further studies that tackle this bilevel model of the European gas market. All papers mentioned so far, however, make the assumption that all data of the model is deterministic. Obviously, this is not the case in practice. In particular, the future energy consumption is not known and it is rather standard to tackle such aspects using techniques from stochastic optimization.

Stochastic bilevel problems or, more generally, stochastic mathematical programs with equilibrium constraints (SMPECs), have been introduced in [40] and further investigated theoretically, e.g., in [47]. They have found numerous applications in specific electricity market models; see, e.g., [24]. In these models, the typical chronology of first- and second-stage decisions x, y accompanied by the observation of the random parameter ξ is $x \curvearrowright \xi \curvearrowright y$. In this case, x is a here-and-now decision (before randomness is revealed) and y is a wait-and-see decision (reacting on randomness). This implies, in particular, that y can be chosen (as a function of x and ξ) such that the lower-level constraints are satisfied almost surely with respect to randomness for any x. However, in the gas market model we are considering in this paper, the chronology is $x \curvearrowright y \curvearrowright \xi$, which means that both first- and second-stage decisions are of type here-and-now. In this case, the satisfaction of random lower-level constraints cannot be guaranteed almost surely in general, but only with a certain probability. This leads in a natural way to the use of probabilistic or chance constraints. The research on this topic is rather young. Exemplarily, we refer to the theoretical paper [26] devoted to a chance-constrained bilevel model in the chronology $x \curvearrowright \xi \curvearrowright y$ we mentioned above and refer to the two papers [43, 54] dealing with the numerical solution of chance constraints as parts of bilevel problems. These latter articles make simplifying assumptions by restricting randomness to discrete measures or imposing individual rather than joint chance constraints; see also our critical discussion in the case study of Section 3.4. The main contribution of this paper is the consideration of *joint* chance constraints with *continuous* multivariate random distribution on the *lower* level of a bilevel problem.

We thus extend the deterministic setting to incorporate uncertain loads in the lower, i.e., the market, level of the bilevel problem. In the literature, gas buying firms are usually modeled using inverse market demand functions to model price elasticity. While this is a reasonable model in the deterministic setting, the bought quantities do not solely follow an economic rationale but also need to cover mostly random fluctuations in the day-ahead. This is not covered in the models in the literature and we capture this aspect by introducing a respective chance constraint in the lower level of the bilevel problem. On the other hand, we make simplifying assumptions such as considering a linear gas flow model as well as that we only consider passive networks to keep the model practically tractable and the presentation streamlined.

This extension of the model adds a significant mathematical challenge to the overall bilevel problem. Although the chance constraint is convex—and, thus, first-order optimality conditions are usable in general—it is not given in a closed form. Instead, only function and gradient evaluations are available and we have to design methods that can cope with such black-box functions. To the best of our knowledge, this problem class has not been considered so far in the literature.

In this setup, our contribution is the following. First, we study bilevel problems with a convex upper and lower level that contain a black-box constraint in the lower level in Section 2. Since a closed form of the latter is not available, one has to resort to iterative approaches that subsequently approximate the lower-level's feasible set, which is possible by outer approximations since function evaluations and first-order information is available. Thus, the lower-level's feasible set can only be approximated up

to a certain prescribed tolerance. We discuss in detail that this makes it challenging (if possible at all) to ensure global optimality of the bilevel problem. Nevertheless we can show that, if our algorithm terminates, it terminates with a point that is approximate feasible for the lower level. Second, we present a chance-constrained extension of a simplified bilevel model of the European entry-exit gas market and also discuss the economic interpretation of the added chance constraint; cf. Section 3. For this application, we can further derive provably lower bounds by adjusting our method, which allows us to assess the quality of the obtained solutions. It turns out, that we are always very close to a globally optimal solution of the chance-constrained bilevel problem. Third and finally, we present numerical results for the chance-constrained bilevel problem of the European gas market on an academic instance and discuss the effect of uncertainty on the results.

2 Bilevel problems with convex lower-level black-box constraints

2.1 Problem formulation

We consider convex bilevel problems of the form

$$\min_{x,y} \quad F(x,y) \tag{1a}$$

$$\text{s.t.} \quad G(x,y) \leq 0, \tag{1b}$$

$$x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y}, \tag{1c}$$

$$y \in S(x), \tag{1d}$$

in which the solution set of the convex lower-level problem is given by

$$S(x) = \arg\min_{y} \left\{ f(x,y) \colon g(x,y) \leq 0, \ b(y) \leq 0, \ y \in \mathbb{R}^{n_y} \right\}. \tag{2a}$$

The variable vector x denotes the n_x upper-level (or leader) decisions and y denotes the n_y lower-level (or follower) decisions. We assume that $F : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$, $G : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^{m_u}$, $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$, $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^{m_\ell}$, and $b : \mathbb{R}^{n_y} \to \mathbb{R}$ are convex and differentiable functions. Further, we suppose that $b(y) \leq 0$ is a black-box constraint for which we make the following assumptions.

Assumption 1. *The black-box function b is convex and for all $(x,y) \in \{(x,y) \colon G(x,y) \leq 0, \ g(x,y) \leq 0\}$,*

1. *we can evaluate the function $b(y)$,*

2. *we can evaluate the gradient $\nabla b(y)$, and*

3. *the gradient is bounded, i.e., $\|\nabla b(y)\| \leq K$ for a fixed $K \in \mathbb{R}$.*

Overall, the upper-level problem (1a)–(1c) is a convex minimization problem and for fixed upper-level variables $x = \bar{x}$, the lower level is a convex minimization problem as well. In the case that the follower problem (2) has multiple solutions, i.e., $S(x)$ is not a singleton, Problem (1) models the so-called optimistic bilevel solution, which realizes the lower-level solution $y \in S(x)$ that minimizes the upper-level objective function F.

In the following, we denote the shared constraint set by

$$\Omega := \{(x, y) \colon G(x, y) \leq 0, \ g(x, y) \leq 0, \ b(y) \leq 0\},$$

its projection onto the decision space of the leader by

$$\Omega_{\mathrm{u}} := \{x \colon \exists y \text{ with } (x, y) \in \Omega\},$$

and the feasible set of the lower-level problem for a fixed leader decision $x = \bar{x}$ by

$$\Omega_\ell(\bar{x}) := \{y \colon g(\bar{x}, y) \leq 0, \ b(y) \leq 0\}.$$

Further, we denote the optimal value function of the lower-level problem by

$$\varphi(x) = \min_y \{f(x, y) \colon g(x, y), \ b(y) \leq 0, \ y \in \mathbb{R}^{n_y}\}.$$

Using the latter, we rewrite Problem (1) as the equivalent single-level problem

$$\min_{x,y} \quad F(x, y) \tag{3a}$$

$$\text{s.t.} \quad G(x, y) \leq 0, \quad g(x, y) \leq 0, \tag{3b}$$

$$b(y) \leq 0, \tag{3c}$$

$$f(x, y) \leq \varphi(x), \tag{3d}$$

$$x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y}. \tag{3e}$$

2.2 Obstacles and pitfalls

Besides being a bilevel problem, the main challenge of Problem (1), and also of Problem (3), is the black-box constraint $b(y) \leq 0$ in the lower-level problem. Typical methods for solving bilevel problems with convex lower levels replace the lower level with its necessary and sufficient optimality conditions in algebraic form. This is, however, not possible if the lower level contains a black-box function—i.e., a function for which an explicit expression is not available—since without such an expression, optimality conditions are also not available in closed form. Since later on, our method needs to employ global optimization solvers and since general-purpose global optimization software usually requires all constraints to be given in algebraic form (e.g., to derive suitable over- and underestimators), the considered back-box setting adds to the hardness of the bilevel problem at hand. As a remedy, we need to resort to iterative approaches, e.g., cutting-plane techniques [27] or outer approximation [12, 17], and the best we can expect is to fulfill the black-box constraint up to an a-priorily specified tolerance. In order to specify the quality of such solutions, we make use of the concepts of δ-feasibility and ε-δ-optimality known from global optimization; see, e.g., [34]. The same concept is also used in a bilevel-specific context in, e.g., [38], where the authors require this notion since they consider a nonconvex optimization problem in the lower level. In this paper, we use the following definition.

Definition 1. *For $\delta = (\delta_G, \delta_g, \delta_b, \delta_f) \in \mathbb{R}_{\geq 0}^{m_{\mathrm{u}}+m_\ell+2}$, a point $(\bar{x}, \bar{y}) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ is called δ-feasible for Problem (1), if $G(\bar{x}, \bar{y}) \leq \delta_G, g(\bar{x}, \bar{y}) \leq \delta_g, b(y) \leq \delta_b$, and $f(x, y) \leq \varphi(x) + \delta_f$ hold. Moreover, for $\varepsilon \geq 0$, a point $(x^*, y^*) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ is called ε-δ-optimal for Problem (1), if it is δ-feasible and if $F(x^*, y^*) \leq F^* + \varepsilon$ holds, with F^* denoting the optimal objective function value of Problem (1).*

We highlight that a δ-feasible point $(\bar{x}, \bar{y})$ of Problem (1) is, in particular, δ_f-(δ_g, δ_b)-optimal for the lower-level problem (2) with fixed $x = \bar{x}$. The point is (δ_g, δ_b)-feasible because $g(\bar{x}, \bar{y}) \leq \delta_g$ as well as $b(y) \leq \delta_b$ hold and the objective function value is bounded by $f(x, y) \leq \varphi(x) + \delta_f$. If the functions f and g do not lead to further difficulties, we can choose $\delta_f = \delta_g = 0$. If we can, in addition, also optimize exactly over F and G, then the best we can hope for in our bilevel setting is to obtain a 0-δ-optimal solution of Problem (1) with $\delta = (0, 0, \delta_b, 0)$.

In practice, however, this turns out to be not so easy. We already mentioned that we can only expect to fulfill the black-box constraint up to a tolerance. Let us consider the following relaxation of the lower level for a given upper-level decision $x = \bar{x}$:

$$\min_{y \in \mathbb{R}^{n_y}} \quad f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \ b(y) \leq \delta_b. \tag{4}$$

We denote the optimal value function of this problem by $\underline{\varphi}(x)$. Since Problem (4) is a relaxation of the original lower-level problem (2), $\underline{\varphi}(x) \leq \varphi(x)$ holds for all feasible $x \in \Omega_u$.

Virtually all solution techniques for bilevel problems with a convex lower-level problem exploit the single-level reformulation (3) or similar variants that express the optimal value function φ more explicitly, e.g., by using the optimality conditions of the convex lower-level problem. If we replace the lower-level problem with its relaxation (4) and derive a single-level reformulation in analogy to Problem (3), then we need to (i) relax Constraint (3c) and (ii) replace Constraint (3d) by $f(x, y) \leq \underline{\varphi}(x)$. If $\underline{\varphi}(x) < \varphi(x)$ holds for any $x \in \Omega_u$, then such a "first-relax-then-reformulate" approach yields a single-level reformulation that is *not* a relaxation of the single-level reformulation (3). Thus, it is also not a relaxation of the original bilevel problem (1). Consequently, it is not clear whether and how ε-δ-optimality can be guaranteed.

This observation is closely related to the missing independence of irrelevant constraints (IIC) property that is known from single-level optimization. In [35], it is shown that adding an inequality to the lower-level problem that is valid for the bilevel optimal solution may result in a better bilevel solution that is not feasible without the added valid inequality. We demonstrate this effect using a simple linear bilevel problem.

Example 1. *The linear bilevel problem*

$$\min_{x, y \in \mathbb{R}} \quad x \quad \text{s.t.} \quad y \geq \frac{1}{2}x + 1, \ x \geq 0, \ y \in \arg\min_{\bar{y} \in \mathbb{R}} \left\{ \bar{y} : \bar{y} \geq 2x - 2, \ \bar{y} \geq 0 \right\},$$

possesses the optimal solution $(2, 2)$ *with objective function value* 2. *The inequality* $y \geq 1$ *is valid but inactive for the optimal solution* $(2, 2)$. *Adding this constraint to the lower level yields the problem*

$$\min_{x, y \in \mathbb{R}} \quad x \quad \text{s.t.} \quad y \geq \frac{1}{2}x + 1, \ x \geq 0, \ y \in \arg\min_{\bar{y} \in \mathbb{R}} \{ \bar{y} : \bar{y} \geq 2x - 2, \ \bar{y} \geq 1 \},$$

with optimal solution $(0, 1)$ *and objective value* 0.

For our application, Example 1 particularly suggests that stopping a cutting-plane approach once the black-box constraint is fulfilled up to a δ_b-tolerance may not yield a globally optimal solution. According to [35], a cutting-plane approach is only applicable, if the solution of the so-called high point relaxation—the relaxation of the bilevel problem that disregards lower-level optimality—is also a solution of the original bilevel problem. This however means that we do not have a "real" bilevel problem in the first place. In [10], it is shown that a globally optimal solution of a bilevel problem remains locally optimal when adding valid inequalities to the lower level in case of inner semicontinuity of the solution set

mapping $S(x)$. Thus, under this restrictive assumption, a cutting-plane approach may at least yield locally optimal bilevel solutions.

In any case, determining an ε-optimality certificate, i.e., finding an ε such that $F(x, y) \leq F^* + \varepsilon$ holds for (x, y) derived by a cutting-plane algorithm is, in our opinion, an extremely challenging and up to now open task. Thus, the best we can hope for when applying a native cutting-plane approach is to arrive at a δ-feasible point with $\delta = (0, 0, \delta_b, 0)$. In other words, the best we can hope for is an approximate heuristic for the bilevel problem with lower-level black-box constraints.

2.3 A "first-relax-then-reformulate"-approach to compute δ-feasible points

In this section, we formalize a cutting-plane approach that proceeds in a "first-relax-then-reformulate" way. We will show that this approach can compute δ-feasible points of Problem (1). After that, we briefly discuss why an intuitive extension, a "first-reformulate-then-relax" approach, fails.

Since the black-box constraint $b(y) \leq 0$ is convex, we can construct a sequence of linear outer approximations $(E^r, e^r)_{r \in \mathbb{N}}$ of the black-box constraint $b(y) \leq 0$ with the property

$$\{y \in \mathbb{R}^{n_y} : b(y) \leq 0\} \subseteq \{y \in \mathbb{R}^{n_y} : E^{r+1}y \leq e^{r+1}\} \subseteq \{y \in \mathbb{R}^{n_y} : E^r y \leq e^r\}. \tag{5}$$

Such descriptions can be obtained, e.g., by a classic cutting-plane approach [27] that exploits the fact that the first-order Taylor approximation of b is a global underestimator of b.

We note that for a given upper-level solution $\bar{x} \in \Omega_{\mathrm{u}}$ and $r \in \mathbb{N}$, the adapted lower-level problem

$$\min_{y \in \mathbb{R}^{n_y}} \quad f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \; E^r y \leq e^r, \tag{6}$$

is a convex relaxation of the original lower-level problem (2). If we denote the optimal value function of Problem (6) by $\underline{\varphi}^r(x)$, then the following lemma follows directly from (5).

Proposition 1. *For every $r \in \mathbb{N}$ and every upper-level decision $x \in \Omega_{\mathrm{u}}$, it holds*

$$\underline{\varphi}^r(x) \leq \underline{\varphi}^{r+1}(x) \leq \varphi(x).$$

We now follow a "first-relax-then-reformulate" approach by replacing the original lower-level problem (2) with its relaxation (6). This yields the following modified variant of the single-level reformulation (3):

$$\min_{x, y} \quad F(x, y) \tag{7a}$$

$$\text{s.t.} \quad G(x, y) \leq 0, \quad g(x, y) \leq 0, \tag{7b}$$

$$E^r y \leq e^r, \tag{7c}$$

$$f(x, y) \leq \underline{\varphi}^r(x) \tag{7d}$$

$$x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y}. \tag{7e}$$

For the practical tractability of Problem (7), we need the following assumption.

Assumption 2. *For every upper-level decision $\bar{x} \in \Omega_{\mathrm{u}}$ and every $r \in \mathbb{N}$, Slater's constraint qualification holds for the relaxed lower-level problem (6).*

With Assumption 2, we can solve Problem (7) to global optimality, e.g., by replacing the optimal-value-function constraint (7d), which models the optimality of the relaxed lower level (6), and the lower-level

---

Algorithm 1 "First-Relax-Then-Reformulate".

1: Choose $\delta_b > 0$, set $r = 0$, $s = 0$, $\chi = \infty$, $E^0 = [0 \ldots 0] \in \mathbb{R}^{1 \times n_y}$, $e^0 = 0 \in \mathbb{R}$.

2: **while** $\chi > \delta_b$ or $s > 0$ **do**

3:     Construct E^{r+1} and e^{r+1} such that (5) is fulfilled.

4:     **if** Problem (7) is feasible **then**

5:         Solve Problem (7) to obtain (x^{r+1}, y^{r+1}) and set $s = 0$.

6:     **else if** Problem (8) is feasible **then**

7:         Solve Problem (8) to obtain (x^{r+1}, y^{r+1}, s).

8:     **else**

9:         Return "Problem (1) is infeasible.".

10:    **end if**

11:    Set $r \leftarrow r + 1$ and $\chi = b(y^r)$.

12: **end while**

13: Return $(\bar{x}, \bar{y}) = (x^r, y^r)$.

---

primal feasibility with the necessary and sufficient KKT conditions of Problem (6). Note that such an explicit expression is not available for the Constraint (3d), because it would require explicit knowledge on the black-box function b.

Following the discussion of the last section, we re-iterate that, although we relax the black-box constraint, we do not obtain a relaxation of Problem (3) due to the tightened constraint (7d); see Proposition 1. Still, we can exploit Problem (7) to compute a δ-feasible point of Problem (1) as summarized in Algorithm 1. After the initialization (Step 1), we enter a while-loop in which we first check, among others, for approximate feasibility w.r.t. the chance constraint. Inside the loop, we first update the linear outer approximation in Step 3. This can be achieved, e.g., by using the first-order Taylor approximation of the black-box function at the last iterate, which is possible due to Assumption 1. Then, we solve Problem (7) to obtain a point (x^{p+1}, y^{p+1}). In case Problem (7) is infeasible, we resort to the following feasibility problem:

$$\min_{x,y,s} \quad s \tag{8a}$$

$$\text{s.t.} \quad G(x, y) \leq 0, \quad g(x, y) \leq 0, \tag{8b}$$

$$E^r y \leq e^r, \tag{8c}$$

$$f(x, y) \leq \underline{\varphi}^r(x) + s, \tag{8d}$$

$$x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y}. \tag{8e}$$

We argued before that Problem (7) is not a relaxation of the single-level reformulation (3), because the optimality condition of the lower level is tightened. We thus relax the optimality of the lower level to compute a new iterate (x^{r+1}, y^{r+1}, s). Since Problem (7) is infeasible, $s > 0$ must hold.

Note also that Problem (8) must be feasible if Problem (3) is feasible. Consequently, if Problem (8) is infeasible, we can state that Problem (1) is infeasible. Otherwise, if Problem (7) or Problem (8) is feasible, we update χ and r.

Finally, note that Algorithm 1 only terminates if Problem (7) is feasible. This is ensured by the check on s in Step 2.

Theorem 1. *If Algorithm 1 terminates, $(\bar{x}, \bar{y})$ is $(0, 0, \delta_b, 0)$-feasible for Problem* (1).

Proof. The point $(\bar{x}, \bar{y})$ fulfills the constraints $G(\bar{x}, \bar{y}) \leq 0$ and $g(\bar{x}, \bar{y}) \leq 0$. Further, we have $b(\bar{y}) \leq \delta_b$. Finally, $\bar{y}$ is an optimal solution for the relaxed lower-level problem (6) for fixed $\bar{x}$, i.e.,

$f(\bar{x}, \bar{y}) \leq \underline{\varphi}^r(\bar{x})$. Due to Lemma 1, we thus know that $f(\bar{x}, \bar{y}) \leq \varphi(\bar{x})$ holds. In summary, $(\bar{x}, \bar{y})$ is a $(0, 0, \delta_b, 0)$-feasible point of Problem (1). □

We point out that Theorem 1 makes no statement regarding the convergence of Algorithm 1. In fact, convergence cannot be guaranteed. Consider, for example, the case in which Problem (7) is infeasible and for the solution (x^{r+1}, y^{r+1}, s) of the feasibility problem (8) it holds $b(y^{r+1}) \leq \delta_b$. Then, updating the linear outer approximation using the respective first-order Taylor approximation of b at y^{r+1} does not exclude the point (x^{r+1}, y^{r+1}). In the new iteration, the algorithm then delivers either the same solution (x^{r+1}, y^{r+1}, s) again—or we are lucky and the missing IIC property helps to find a new iterate: Adding the first-order Taylor approximation may change $\bar{\varphi}^r(x)$ such that either Problem (7) might become feasible or the solution of Problem (8) differs from the previous iteration. In any way, Algorithm 1 may not terminate if such a lucky situation does not appear. However, we demonstrate in Section 3 that it works very well in practice.

We also note that Algorithm 1 requires to solve a bilevel problem in every iteration. In order to reduce the computational burden, we can initialize the algorithm with a tight initial linear outer approximation $E^0 y \leq e^0$, obtained, e.g., by an initial cut sampling phase.

Before we conclude this section, we briefly discuss a "first-reformulate-then-relax" approach. We already highlighted in the last section that there is not much hope to derive algorithms that yield more than a δ-feasible point of Problem (1). We also discussed that for the "first-relax-then-reformulate" approach, the main problem is that Problem (7) is not a relaxation of Problem (1). However, Problem (8) might lead to the intuition that relaxing the optimality of the follower problem appropriately resolves this issue. Consequently, we may apply a "first-reformulate-then-relax" approach as follows.

Remark 1. *We may first reformulate Problem* (1) *to Problem* (3) *and then relax the black-box constraint* (3c) *and the optimality of the lower level* (3d)*. We recap that the latter is necessary, because we cannot describe $\varphi(x)$ explicitly. We therefore need to find functions $\bar{\varphi}^r$ with*

$$\varphi(x) \leq \bar{\varphi}^{r+1}(x) \leq \bar{\varphi}^r(x)$$

such that the problem

$$\min_{x,y} \quad F(x, y) \tag{9a}$$

$$\text{s.t.} \quad G(x, y) \leq 0, \quad g(x, y) \leq 0, \tag{9b}$$

$$E^r y \leq e^r, \tag{9c}$$

$$f(x, y) \leq \bar{\varphi}^r(x) \tag{9d}$$

$$x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y} \tag{9e}$$

is a relaxation of Problem (1) *for every $r \in \mathbb{N}$. This would allow to compute 0-δ-optimal points of Problem* (1)*. It is, however, not clear how to derive suitable functions $\bar{\varphi}^r$. One strategy might be to use a sequence $(H^r, h^r)_{r \in \mathbb{N}}$ of linear inner approximations of the lower-level problem such that*

$$\{y \in \mathbb{R}^{n_y} \colon H^r y \leq h^r\} \subseteq \{y \in \mathbb{R}^{n_y} \colon H^{r+1} y \leq h^{r+1}\} \subseteq \{y \in \mathbb{R}^{n_y} \colon b(y) \leq 0\}$$

holds. By replacing the black-box constraint with the linear inner description, we would obtain the tightened lower-level problem

$$\bar{\varphi}^r(x) := \min_y \{f(x, y) \colon g(x, y) \leq 0, \ H^r y \leq h^r\}. \tag{10}$$

In order to use Problem (9) *algorithmically, we need an explicit description of* $\bar{\varphi}(x)$. *Therefore, we may use the KKT conditions or strong duality of Problem* (10) *and reformulate Problem* (9) *accordingly. This, in turn, requires Problem* (10) *to be feasible for every upper-level decision* $x \in \Omega_{\mathrm{u}}$, *which would be fulfilled, e.g., for a sufficiently tight linear inner description* (H^r, h^r) *of* $b(y) \leq 0$ *throughout its entire domain. This is not practicable from a computational point of view because such a linearization over the entire domain, in general, requires a significant amount of inequalities such that the resulting problem is very large. In addition, if such a sufficiently tight linear description over the entire domain would be available, we could directly use it to replace the lower level by its KKT conditions. Hence, no tailored algorithm would be necessary in the first place.*

Remark 2. *Let us finally comment on some possible extensions of the setting considered in this section.*

1. *We can easily add integer variables to the upper level of the original bilevel problem. The main rationale of the presented algorithm stays the same. The only difference is that a mixed-integer problem—instead of continuous one—needs to be solved in every iteration of the algorithm.*

2. *The same applies to more general, and maybe nonconvex, functions* F *and* G *in the upper level. The main rationale of the algorithm does not change if nonconvex functions are part of the upper level but the problem to be solved in each iteration of Algorithm 1 is harder. Moreover, one would have to consider an additional* ϵ-δ *notion for the global optimization task of the upper level as well.*

3. *Finally, we presented the models and the algorithm in this section for a scalar black-box function* b *in the lower level. However, besides technical adaptions, nothing changes if* b *is an entire vector of black-box functions.*

3 A European gas market model with chance constraints

In this section, we use the algorithmic approach developed in the previous section to solve a simplified model of the European entry-exit gas market with uncertain load modeling. We therefore take the model from [21] and extend it by a chance constraint to model uncertain loads. This results in a bilevel problem with a black-box constraint in the lower level; see Section 3.2. Afterward, we tailor the proposed solution approach to the specific setting of the considered gas market model and give some more details on the handling of the chance constraint, which is a black-box constraint as discussed in the previous section. Finally, we provide some algorithmic insights and discuss the obtained results from an applied perspective in Section 3.4.

In a nutshell, the European entry-exit gas market is organized as follows:

1. The transmission system operator (TSO) announces so-called technical capacities and booking price floors for every entry and exit node of the network to maximize the overall welfare.

2. Gas selling and buying firms located at entry and exit nodes, respectively, afterward sign capacity-right contracts (called bookings), which are limited by the technical capacities. The prices of these bookings are determined by the booking price floors set by the TSO plus a scarcity-based markup price.

3. The gas selling and buying firms then nominate entry and exit quantities on a day-ahead market. These nominations need to be balanced and are bounded above by the bookings. No other technical or physical restrictions are considered in this level.

    4 The realized nominations are cost-optimally transported through the network by the TSO.

This setting is adequately captured by the four-level model developed in [21]. A peculiarity of the European gas market is the so-called robustness constraint. It requires that the TSO announces only those technical capacities that guarantee that *every* possibly occurring balanced nomination (i.e., entry and exit quantities match) that is restricted by the given set of bookings can indeed be transported through the network. This means, every such nomination needs to be physically and technically feasible. This leads to unnecessarily restrictive technical capacities, which themselves result in welfare losses (cf., e.g., [5]) on the one hand and to mathematically extremely challenging models (cf. [32, 33, 46, 49]) on the other hand. From a mathematical point of view, the robustness constraint can be seen as an adjustable robust constraint; see, e.g., [4, 56]. As shown in [5], an economically and algorithmically more reasonable approach is to require that only the actually realized nominations need to be transportable through the network. This is the setup we consider in what follows and we will come back to this aspect in the next section as well.

All papers published so far that tackle this four-level setup consider the deterministic case, i.e., exit nominations are purely driven by economic mechanisms that are modeled using inverse market demand functions to capture price elasticity. In what follows, we analyze the effects of uncertain demand. Before we do so in Section 3.2, we first recap the deterministic model from the literature in the next section.

3.1 A deterministic bilevel model

In Section 3 of [21] it is shown that the special structure of the four-level gas market model can be exploited to equivalently re-state it as a bilevel problem. In particular, the original second- and third-level problem, which model the profit- or surplus-maximizing bookings and nominations of the gas selling and buying firms, can be merged into a single level; see Theorem 5 in [21]. Moreover, the original fourth-level problem (i.e., the transportation level) can be merged into the original first level, in which the technical capacities and booking price floors are determined; see Theorem 7 in [21]. Before we state the resulting bilevel problem, we introduce some notation that is in line with [5, 21]. The gas network is modeled as a directed and connected graph $G = (V, A)$ with nodes V and arcs A. The node set is split into the set of entry nodes $V_+ \subseteq V$ at which gas is supplied, the set of exit nodes $V_- \subseteq V$ at which gas is discharged, and the set of inner nodes $V_0 \subseteq V$ without gas supply or withdrawal. Thus, $V = V_+ \cup V_- \cup V_0$ holds. The model allows for multiple gas selling or gas buying firms $i \in \mathcal{P}_u$ for $u \in V_+$ or $u \in V_-$, respectively. We denote the set of entry players by $\mathcal{P}_+ := \cup_{u \in V_+} \mathcal{P}_u$ and the set of exit players by $\mathcal{P}_- := \cup_{u \in V_-} \mathcal{P}_u$. We consider multiple time periods $t \in T$ with $|T| < \infty$ of gas trading and transport and model the rational behavior of gas buying firms $i \in \mathcal{P}_-$ based on linear inverse market demand functions $P_{i,t}(q) = a_{i,t} - b_{i,t}q$ with $a_{i,t} > 0, b_{i,t} > 0$; see [36] for the economic background. Further, we characterize gas selling firms using pairwise distinct variable costs of production $c_i^{\mathrm{var}} > 0$, $i \in \mathcal{P}_+$. When abstracting from the above mentioned robustness constraint, the bilevel reformulation of the four-level model from [21] is given by

$$\max_{q^{\mathrm{TC}}, \underline{\pi}^{\mathrm{book}}, \pi, q} \quad \varphi^{\mathrm{u}}(q^{\mathrm{nom}}, q) = \sum_{t \in T} \left(\sum_{i \in \mathcal{P}_-} \int_0^{q_{i,t}^{\mathrm{nom}}} P_{i,t}(s)\,\mathrm{d}s - \sum_{i \in \mathcal{P}_+} c_i^{\mathrm{var}} q_{i,t}^{\mathrm{nom}} \right) \tag{11a}$$

$$- \sum_{t \in T} \sum_{a \in A} c^{\mathrm{trans}}(q_{a,t})$$

$$\text{s.t.} \quad 0 \le q_u^{\mathrm{TC}}, \ 0 \le \underline{\pi}_u^{\mathrm{book}} \quad \text{for all } u \in V_+ \cup V_-, \tag{11b}$$

$$\sum_{u \in V_+ \cup V_-} \sum_{i \in \mathcal{P}_u} \pi_u^{\text{book}} q_i^{\text{book}} = \sum_{t \in T} \sum_{a \in A} c^{\text{trans}}(q_{a,t}), \tag{11c}$$

$$(\pi, q) \in \mathcal{F}(q^{\text{nom}}), \tag{11d}$$

$$(q^{\text{book}}, q^{\text{nom}}) \in \arg\max (12), \tag{11e}$$

where the lower level reads

$$\max_{q^{\text{book}}, q^{\text{nom}}} \quad \sum_{t \in T} \left(\sum_{i \in \mathcal{P}_-} \int_0^{q_{i,t}^{\text{nom}}} P_{i,t}(s)\, \mathrm{d}s - \sum_{i \in \mathcal{P}_+} c_i^{\text{var}} q_{i,t}^{\text{nom}} \right) \tag{12a}$$

$$- \sum_{u \in V_+ \cup V_-} \sum_{i \in \mathcal{P}_u} \underline{\pi}_u^{\text{book}} q_i^{\text{book}}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{P}_u} q_i^{\text{book}} \le q_u^{\text{TC}} \quad \text{for all } u \in V_+ \cup V_-, \tag{12b}$$

$$0 \le q_{i,t}^{\text{nom}} \le q_i^{\text{book}} \quad \text{for all } i \in \mathcal{P}_- \cup \mathcal{P}_+,\ t \in T, \tag{12c}$$

$$\sum_{i \in \mathcal{P}_-} q_{i,t}^{\text{nom}} - \sum_{i \in \mathcal{P}_+} q_{i,t}^{\text{nom}} = 0 \quad \text{for all } t \in T. \tag{12d}$$

In the upper level (11), the TSO specifies technical capacities q_u^{TC} and booking price floors $\underline{\pi}_u^{\text{book}}$ for every entry and exit node $u \in V_+ \cup V_-$. The TSO uses the booking price floors to cover the transportation costs; see the bilinear constraint (11c). This constraint involves the bookings q^{book}, which are, together with the nominations q^{nom}, outcome of the lower-level problem; see Constraint (11e). The actual nominations are transported in a cost-optimal way through the network. To be more specific, the squared nodal pressures π and mass flows q on the arcs resulting from the nomination need to fulfill the technical network limitations denoted by $\mathcal{F}(q^{\text{nom}})$ such that the transportation costs c^{trans} are minimized. The generic notation in Constraint (11d) allows to use various transportation models. In this paper, we use the modeling and notation that is used in [5]. For an arc $a = (u, v)$, $q_{a,t} > 0$ denotes the mass flow in the direction of the arc, i.e., from u to v, and $q_{a,t} < 0$ denotes flow in the opposite direction. The mass flow has to satisfy given bounds of the pipes:

$$-\infty \le q_a^- \le q_{a,t} \le q_a^+ \le \infty \quad \text{for all } a \in A,\ t \in T.$$

In addition, we model mass balance at every node of the network using the constraints

$$\sum_{a \in \delta_u^{\text{out}}} q_{a,t} - \sum_{a \in \delta_u^{\text{in}}} q_{a,t} = \sum_{i \in \mathcal{P}_u} q_{i,t}^{\text{nom}} \qquad \text{for all } u \in V_+,\ t \in T,$$

$$\sum_{a \in \delta_u^{\text{out}}} q_{a,t} - \sum_{a \in \delta_u^{\text{in}}} q_{a,t} = -\sum_{i \in \mathcal{P}_u} q_{i,t}^{\text{nom}} \qquad \text{for all } u \in V_-,\ t \in T,$$

$$\sum_{a \in \delta_u^{\text{out}}} q_{a,t} - \sum_{a \in \delta_u^{\text{in}}} q_{a,t} = 0 \qquad \text{for all } u \in V_0,\ t \in T,$$

where δ_u^{out} and δ_u^{in} represent outgoing and incoming edges at node u, respectively. Moreover, the pressure loss law

$$\pi_{u,t} - \pi_{v,t} = \Phi_a(q_{a,t}) \quad \text{for all } a = (u, v) \in A,\ t \in T, \tag{13}$$

couples the mass flows $q_{a,t}$ on each arc with the squared pressures $\pi_{u,t}, \pi_{v,t}$ at the incident nodes. The squared pressures are bounded by the squared pressure bounds

$$0 < \pi_u^- \le \pi_{u,t} \le \pi_u^+ \le \infty \quad \text{for all } u \in V,\ t \in T.$$

The function Φ_a in (13) denotes the pressure loss function for arc $a \in A$. In this paper, we make the following assumption, which is also used in [5].

Assumption 3. *The pressure loss function Φ_a is linear for all $a \in A$.*

We discuss this simplification of the physics model at the end of this subsection.

The transportation costs c^{trans} generally arise from compensating pressure losses via controllable elements. However, in this paper, we only consider passive networks without controllable elements like compressors or control valves. We also discuss the simplification of considering passive networks at the end of this subsection in more detail. To mimic transportation costs in our passive network model, we penalize squared pressure losses on all arcs of the network, i.e.,

$$c^{\mathrm{trans}}(q_{a,t}) = c_t^{\mathrm{trans}}|\pi_{u,t} - \pi_{v,t}|, \tag{14}$$

where $c_t^{\mathrm{trans}} > 0$ is a given parameter and $a = (u, v) \in A$, $t \in T$. Note that the squared pressures $\pi_{u,t}$, $\pi_{v,t}$ are coupled to the mass flow $q_{a,t}$ via Constraint (13).

In the lower-level problem (12), gas buyers and sellers choose their booking and nomination quantities to maximize their individual profit, i.e., each player solves individual maximization problems. It is shown in [21] that under the assumption of perfect competition, this is equivalent to the welfare maximizing problem (12), which consists of the merged second and third level of the original four-level model. We briefly shed some more light on the derivation of Problem (12). In the second level of the original four-level model, each player i chooses the bookings q_i^{book} to maximize the anticipated outcome that is realized after trade in the third level. Under the assumption of perfect competition, the individual second-level problems can be equivalently reformulated as the aggregated second-level problem

$$\max_{q^{\mathrm{book}}} \quad \text{(12a)} \quad \text{s.t.} \quad \text{(12b)} \quad \text{and} \quad q_i^{\mathrm{book}} \geq 0 \text{ for all } i \in \mathcal{P}_+ \cup \mathcal{P}_-, \tag{15}$$

in which the bookings of the players are bounded by the technical capacities that are set by the TSO in the upper level; see Constraint (12b). The nominations $q_{i,t}^{\mathrm{nom}}$ that appear in the objective function (12a) result from the solution of the third level, where the actual interaction of gas buyers and sellers takes place.

In this third level, all players choose nominations restricted by their bookings (determined in the second level) to maximize their individual profit or surplus with respect to equilibrium market prices π_t^{nom}. Under the assumption of perfect competition, all players act as price takers and, consequently, every gas seller $i \in \mathcal{P}_+$ maximizes the profit in every time period $t \in T$ as follows:

$$\max_{q_{i,t}^{\mathrm{nom}}} \quad (\pi_t^{\mathrm{nom}} - c_i^{\mathrm{var}})q_{i,t}^{\mathrm{nom}} \quad \text{s.t.} \quad 0 \leq q_{i,t}^{\mathrm{nom}} \leq q_i^{\mathrm{book}}. \tag{16}$$

Similarly, every gas buyer $i \in \mathcal{P}_-$ maximizes its surplus in every time period $t \in T$:

$$\max_{q_{i,t}^{\mathrm{nom}}} \quad \int_0^{q_i^{\mathrm{nom}}} P_{i,t}(s)\,\mathrm{d}s - \pi_t^{\mathrm{nom}}q_{i,t}^{\mathrm{nom}} \quad \text{s.t.} \quad 0 \leq q_{i,t}^{\mathrm{nom}} \leq q_i^{\mathrm{book}}. \tag{17}$$

The endogenously resulting equilibrium market price π_t^{nom} ensures that the shared market-clearing condition (12d) is satisfied in every time period $t \in T$. It is shown in [21] that the nomination level can be modeled as a mixed nonlinear complementarity problem (MNCP) by stacking the first-order optimality conditions of all Problems (16) and (17) and by further adding the market-clearing constraint (12d). The market price π_t^{nom} is then exactly the dual variable of the market-clearing condition (12d). In Theorem 3

of [21] it is shown that the resulting MNCP can be recast as an equivalent welfare maximization problem. Further, in Theorem 5 of [21], it is shown that when integrating this problem into the second-level problem (15), one obtains Problem (12).

Overall, the upper-level objective maximizes total welfare (11a), which consists of exit player utilities minus the costs of the entry players and the transportation costs of the TSO. Since the inverse demand functions $P_{i,t}$ are linear, the exit player utilities are convex-quadratic. In addition, the nonsmooth absolute values that appear in the transportation costs can be linearized with additional binary and continuous variables and constraints, exactly like it is done in Section 4 in [5]. After this linearization, we optimize a convex-quadratic objective function over linear and bilinear constraints in the upper level. For fixed upper-level decisions, the lower level is a continuous and concave-quadratic maximization problem.

Remark 3. *We extend this setting by integrating uncertain loads into the lower level* (12) *in the next section. However, this means adding a further complicating aspect to an already challenging model. In this light, we also want to discuss the simplifications that we made in the modeling of the deterministic model. First, we simplified the transportation model by imposing Assumption 3, i.e., by assuming a linear pressure loss function. Using various other transportation models, e.g., the Weymouth equation [19, 51] is possible but results in additional nonlinear constraints, leading to further nonconvexities in the upper level. Second, we abstract from controllable elements like compressors and instead used the proxy* (14) *to account for transportation costs that are mostly driven by these controllable elements. Allowing for active instead of passive networks requires additional integer variables in the upper level. Thus, we would obtain a nonconvex mixed-integer nonlinear problem in the upper level. Third and finally, we also simplified the market mechanism by abstracting from the robustness constraint. This constraint states that every balanced nomination that fulfills the technical capacities must be guaranteed to be transportable by the TSO. In [5], a characterization of feasible bookings from [32] is used to obtain a tractable reformulation of the robustness constraint. This characterization is only valid for passive networks with linear transportation models. It is further reported that adding the robustness constraint results in a significant computational burden.*

Overall, we abstract from all these aspects because the primary focus of this paper is on the handling of black-box constraints in the lower level; see also the next section in which we extend the deterministic lower level by chance constraints. As discussed in Section 2.2, a setting with black-box constraints in a bilevel lower-level problem is very challenging. Dropping any of the mentioned simplifications further complicates this already demanding setting even more.

3.2 A probabilistic extension

We now extend the deterministic setting of the last section to capture demand uncertainties. In reality, exit players $i \in \mathcal{P}_-$ nominate quantities $q_{i,t}^{\mathrm{nom}}$ without exactly knowing the actual load $\xi_{i,t}$ that they need to cover, say one day ahead. We assume that the load vector $\xi = (\xi_{i,t})_{i \in \mathcal{P}_-, t \in T}$ has a log-concave cumulative distribution function. Thanks to Prékopa's Theorem [44, Theorem 4.2.1.], this is in particular true if the distribution has a log-concave density. It is well-known that many prominent multivariate distributions do have a log-concave distribution, e.g., (for any choice of parameters) Gaussian distributions and the uniform distribution on convex and compact sets, as well as (for a restricted range of parameters) Dirichlet, Gamma, Wishart, and log-normal distributions. In the following, we will assume that ξ obeys a multivariate Gaussian distribution, i.e., $\xi \sim \mathcal{N}(m, \Sigma)$. We further assume that the TSO is interested in a fail-safe network and imposes a fee μ on the exit players $i \in \mathcal{P}_-$ to ensure that the realized loads are covered up to a specified safety level $p \in [0, 1]$.

We model this by the following joint (over all times and exit players) probabilistic constraint

$$\mathbb{P}\left(\xi_{i,t} \leq q_{i,t}^{\mathrm{nom}} \text{ for all } i \in \mathcal{P}_-,\ t \in T\right) \geq p, \tag{18}$$

which we add to the lower-level problem (12). The left-hand side of (18) is nothing but the cumulative distribution function of ξ evaluated at the vector q_-^{nom} of nominations at all exits and times. The mentioned log-concavity of this Gaussian distribution function implies that the log-transformed probabilistic load coverage constraint

$$h(q_-^{\mathrm{nom}}) := \log p - \log \mathbb{P}\left(\xi_{i,t} \leq q_{i,t}^{\mathrm{nom}} \text{ for all } i \in \mathcal{P}_-,\ t \in T\right) \leq 0 \tag{19}$$

is convex. Consequently, adding Constraint (19) to the lower-level problem (12) yields the extended lower-level problem

$$\max_{q^{\mathrm{book}}, q^{\mathrm{nom}}} \quad (12a) \quad \text{s.t.} \quad (12b)\text{–}(12d),\ (19), \tag{20}$$

which is still convex.

We argued above that Constraint (18) models a situation, in which the TSO forces the exit players to fulfill a given safety level with respect to the actual loads ξ by imposing a sufficiently large fee μ. Thus, the maximization problem (17), which each gas buyer $i \in \mathcal{P}_-$ solves in the third level, changes to

$$\max_{q_{i,t}^{\mathrm{nom}}} \quad \int_0^{q_i^{\mathrm{nom}}} P_{i,t}(s)\, \mathrm{d}s - \pi_t^{\mathrm{nom}} q_{i,t}^{\mathrm{nom}} - \mu h(q_-^{\mathrm{nom}}) \quad \text{s.t.} \quad 0 \leq q_{i,t}^{\mathrm{nom}} \leq q_i^{\mathrm{book}}. \tag{21}$$

One can think of the fee μ as the "price of the stochasticity". Similarly to the deterministic case, we can again use the first-order optimality conditions to model the nomination level as an MNCP. In addition to the shared market-clearing condition (12d), we also need to add the further "clearing condition"

$$0 \leq \mu \perp -h(q_-^{\mathrm{nom}}) \geq 0 \tag{22}$$

to the MNCP, which couples the fee μ with the log-transformed chance constraint (19). The dual variable of the latter then corresponds to the fee μ. As in the deterministic case, we can recast the resulting MNCP as an equivalent maximization problem (given that the latter satisfies Slater's constraint qualification). When integrating this problem into the second-level problem, we obtain Problem (20).

In summary, we obtain the following probabilistic bilevel problem

$$\max \quad \varphi^{\mathrm{u}}(q^{\mathrm{nom}}, q) \quad \text{s.t.} \quad (11b)\text{–}(11d),\ (q^{\mathrm{book}}, q^{\mathrm{nom}}) \in \arg\max (20). \tag{23}$$

3.3 A tailored solution approach

The probabilistic bilevel model (23) is a problem in which the lower level is convex. However, we do not have a closed-form expression of the log-transformed chance constraint (19), which is part of the lower-level constraint set. In order to circumvent this difficulty, we tailor the solution approach for bilevel problems with black-box constrained lower levels from Section 2.3 to the specific setting of Problem (23). We therefore first give some more details on the log-transformed chance constraint (19).

The probability function h in (19) plays the role of the black-box constraint function b in (2). We are going to check next the satisfaction of Assumption 1 for h. By (19), we have that

$$h(q_-^{\mathrm{nom}}) = \log p - \log \Phi(q_-^{\mathrm{nom}}), \tag{24}$$

where Φ is a multivariate Gaussian distribution function. The basic requirement in Assumption 1 that h is convex, has already been verified above using the log-concavity of Φ. Proceeding with the three items of Assumption 1, the first two items claim the possibility to evaluate h along with its gradient at any given argument y. By (24), this amounts to to evaluate Φ and $\nabla\Phi$ at any given nomination vector q_-^{nom}. As a multivariate Gaussian distribution function, Φ is given as a multivariate integral. Therefore, one has no access to a closed formula but is constrained to sufficiently precise approximations. An efficient way to get fairly precise values of Φ has been described and implemented by Genz [20]. Moreover, it is well-known (see, e.g., [3, Lemma 1]) that the partial derivatives of Φ can be represented as

$$\frac{\partial\Phi}{\partial u_i}(u) = \frac{1}{\sqrt{2\pi\Sigma_{ii}}} \exp\left[-\frac{1}{2\Sigma_{ii}}(u_i - m_i)^2\right] \cdot \tilde{\Phi}(\tilde{u}), \tag{25}$$

where m_i, Σ_{ii} are the corresponding components of the parameters m, Σ of the given Gaussian distribution and $\tilde{\Phi}$ is another Gaussian distribution function evaluated in one dimension less. Here, the parameters of $\tilde{\Phi}$ can be explicitly derived from the parameters m, Σ of Φ and, similarly, the argument $\tilde{u}$ is an explicit function of the original argument u. Hence, one and the same code (as the one by Genz) may serve to provide values and gradients of the cumulative distribution function at any argument. This provides items (i) and (ii) of Assumption 1. Another possibility to do so is the application of the spheric-radial decomposition of Gaussian random vectors, which applies not just to probabilities of random inequality systems with random left-hand side as in (18) but to arbitrary structures, preferably convex in the random vector. Values and gradients of the probability function can be determined simultaneously by evaluating certain spherical integrals. This approach has been intensively analyzed theoretically (see, e.g., [1, 2]) and successfully applied in practice [23].

It remains to justify item (iii) of Assumption 1. By (19), we have that

$$h(q_-^{\mathrm{nom}}) = \log p - \log\Phi(q_-^{\mathrm{nom}}),$$

where Φ is a multivariate Gaussian distribution function. As such it is continuously differentiable and strictly positive, implying that

$$\nabla h(q_-^{\mathrm{nom}}) = \frac{1}{\Phi(q_-^{\mathrm{nom}})}\nabla\Phi(q_-^{\mathrm{nom}}). \tag{26}$$

Since Assumption 1 refers in particular to all x, y satisfying the general constraint $g(x, y) \leq 0$, this implies in the concrete problem (23) the constraint $q_-^{\mathrm{nom}} \geq 0$; see (12c). From the fact that distribution functions are non-decreasing with respect to the partial order of the space they are defined on, we infer that

$$\Phi(q_-^{\mathrm{nom}}) \geq \Phi(0) > 0 \quad \text{for all } q_-^{\mathrm{nom}} \geq 0.$$

Combining this with (25) and (26), one arrives at the desired boundedness property

$$\|\nabla h(q_-^{\mathrm{nom}})\|_\infty \leq \frac{1}{\Phi(0) \cdot \Sigma^{\mathrm{min}}} \quad \text{for all } q_-^{\mathrm{nom}} \geq 0,$$

where $\Sigma^{\mathrm{min}} := \min_i \Sigma_{ii}$ and where we used that $\tilde{\Phi}$ as a distribution function satisfies the relation $\tilde{\Phi} \leq 1$.

Overall, the probabilistic bilevel problem (23) matches the setting of the general bilevel problem (1), the Assumptions 1 and 2, as well as its extensions as discussed in Remark 2 such that we obtain δ-feasible points by applying Algorithm 1. The main steps in every iteration of Algorithm 1 are

    1 to derive a tighter linear outer approximation (E^r, e^r) and

2 to solve Problem (7) to global optimality.

In our setting, we can achieve the former by a classic cutting-plane approach [27]. In iteration r, we use the point $q_-^r = (q_-^{\mathrm{nom}})^r$, which is part of the solution provided by Algorithm 1 in iteration $r-1$, as the linearization point. By exploiting that for convex functions, the first-order Taylor approximation is a global underestimator, we derive the cut

$$h(q_-^r) + \nabla_{q_-^{\mathrm{nom}}}(q_-^r)^\top (q_-^{\mathrm{nom}} - q_-^r) \le 0. \tag{27}$$

Thus, in iteration r, we construct E^{r+1} by adding the row $\nabla_{q_-^{\mathrm{nom}}}(q_-^r)^\top q^{\mathrm{nom}}$ to E^r and construct e^{r+1} by adding the entry $\nabla_{q_-^{\mathrm{nom}}}(q_-^r)^\top q_-^r - h(q_-^r)$ to e^r. Consequently, we obtain the relaxation

$$\max_{q^{\mathrm{book}}, q^{\mathrm{nom}}} \quad \sum_{t \in T} \left(\sum_{i \in \mathcal{P}_-} \int_0^{q_{i,t}^{\mathrm{nom}}} P_{i,t}(s)\, \mathrm{d}s - \sum_{i \in \mathcal{P}_+} c_i^{\mathrm{var}} q_{i,t}^{\mathrm{nom}} \right) \tag{28a}$$
$$- \sum_{u \in V_+ \cup V_-} \sum_{i \in \mathcal{P}_u} \underline{\pi}_u^{\mathrm{book}} q_i^{\mathrm{book}}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{P}_u} q_i^{\mathrm{book}} \le q_u^{\mathrm{TC}}, \quad u \in V_+ \cup V_-, \qquad\qquad [\pi_u^{\mathrm{book}}] \tag{28b}$$

$$0 \le q_{i,t}^{\mathrm{nom}} \le q_i^{\mathrm{book}}, \quad i \in \mathcal{P}_+ \cup \mathcal{P}_-,\, t \in T, \qquad [\gamma_{i,t}^{\pm}] \tag{28c}$$

$$\sum_{i \in \mathcal{P}_-} q_{i,t}^{\mathrm{nom}} - \sum_{i \in \mathcal{P}_+} q_{i,t}^{\mathrm{nom}} = 0, \quad t \in T, \qquad [\pi_t^{\mathrm{nom}}] \tag{28d}$$

$$h(q_-^j) + \nabla_{q_-^{\mathrm{nom}}} h(q_-^j)^\top (q^{\mathrm{nom}} - q_-^j) \le 0, \quad j = 1, \dots, r, \qquad [\mu_j] \tag{28e}$$

of the lower-level problem (20) by replacing Constraint (19) with the Constraints in (28e).

As already discussed for the general bilevel problem (1), if Assumption 2 holds, then Task (ii) from above can be achieved by replacing lower-level optimality, i.e., Constraint (7d), with the KKT conditions of the relaxed lower-level problem (6). In this way, we obtain an algorithmically tractable reformulation of Problem (7). For our application, the KKT conditions of the relaxed lower-level problem (28) consist of primal feasibility (28b)–(28e), the stationarity conditions

$$\sum_{t \in T} \gamma_{i,t}^+ = \underline{\pi}_u^{\mathrm{book}} + \pi_u^{\mathrm{book}}, \quad i \in \mathcal{P}_u,\, u \in V_+ \cup V_-, \tag{29a}$$

$$\pi_t^{\mathrm{nom}} - \gamma_{i,t}^- + \gamma_{i,t}^+ + \sum_{j=1}^r \nabla_{q_{i,t}^{\mathrm{nom}}} h(q_-^j) \mu_j = a_{i,t} + b_{i,t} q_{i,t}^{\mathrm{nom}}, \quad i \in \mathcal{P}_-,\, t \in T, \tag{29b}$$

$$\pi_t^{\mathrm{nom}} + \gamma_{i,t}^- - \gamma_{i,t}^+ = c_i^{\mathrm{var}}, \quad i \in \mathcal{P}_+,\, t \in T, \tag{29c}$$

the complementarity conditions

$$\pi_u^{\mathrm{book}} \left(q_u^{\mathrm{TC}} - \sum_{i \in \mathcal{P}_u} q_i^{\mathrm{book}} \right) = 0, \quad u \in V_+ \cup V_-, \tag{30a}$$

$$\gamma_{i,t}^- q_{i,t}^{\mathrm{nom}} = 0, \quad i \in \mathcal{P}_+ \cup \mathcal{P}_-,\, t \in T, \tag{30b}$$

$$\gamma_{i,t}^+ \left(q_i^{\mathrm{book}} - q_{i,t}^{\mathrm{nom}} \right) = 0, \quad i \in \mathcal{P}_+ \cup \mathcal{P}_-,\, t \in T, \tag{30c}$$

$$\mu_j \left(\nabla_{q_-^{\mathrm{nom}}} h(q_-^j)^\top (q_-^j - q_-^{\mathrm{nom}}) - h(q_-^j) \right) = 0, \quad j = 1, \dots, r, \tag{30d}$$

and the nonnegativity conditions

$$\pi_u^{\text{book}} \geq 0, \quad u \in V_+ \cup V_-, \tag{31a}$$

$$\gamma_{i,t}^-, \gamma_{i,t}^+ \geq 0, \quad i \in \mathcal{P}_+ \cup \mathcal{P}_-, \ t \in T, \tag{31b}$$

$$\mu_j \geq 0, \quad j = 1, \ldots, r. \tag{31c}$$

Altogether, in every iteration of Algorithm 1 applied to Problem (23), we have to solve the problem

$$\max \quad \varphi^{\text{u}}(q^{\text{nom}}, q) \quad \text{s.t.} \quad \text{(11b)–(11d), (28b)–(28e), (29)–(31).} \tag{32}$$

This problem contains several difficulties. First, although the primal constraints (28b)–(28e) of the relaxed lower-level problem are linear, the KKT complementarity constraints (29) are nonconvex. Fortunately, they can be expressed as a set of mixed-integer linear constraints by introducing additional binary variables and sufficiently large big-M values; see [18]. For a linear constraint $a^\top x \leq b$ and its dual variable $\lambda \geq 0$, the complementarity condition $\lambda(b - a^\top x) = 0$ can be linearized as follows:

$$b - a^\top x \leq M_p u, \quad \lambda \leq M_d(1 - u), \quad u \in \{0, 1\}. \tag{33}$$

This linearization is only correct, if the values M_p and M_d are large enough. On the other hand, too large values are also not desirable because the may cause numerical instabilities. In practice, these values are often derived heuristically, which may result in suboptimal solutions; see [41]. In fact, finding correct big-Ms may in general be as hard as solving the original bilevel problem [29]. Sometimes, however, problem-specific knowledge can be used to obtain correct big-M values; see, e.g., [30, 46, 48]. For the deterministic model that we discussed in Section 3.1, valid big-M values are derived from economic relationships in [5]. In the probabilistic setting, economic relationships are disturbed. Still, we can derive valid big-Ms in a similar way as it is done in [5] by exploiting technical limits of the network. For the only Lagrangian multiplier that does not appear in the deterministic setting, we derive a valid upper bound in Appendix B.

In addition, we are facing the bilinear upper-level constraint (11c). Hence, after having obtained the reformulation (33) of the complementarity conditions (30), we need to solve a mixed-integer maximization problem with a concave-quadratic objective function over linear and bilinear constraints in every iteration of Algorithm 1. Such problems can be solved by modern global solvers such as Gurobi, which tackle bilinear constraints by spatial branching [25]. We highlight that by design of spatial branching, a solution of Problem (32) fulfills Constraint (11c) only up to a pre-specified solver tolerance $\delta_G > 0$. Consequently, the solution we obtain from Algorithm 1 is a $(\delta_G, 0, \delta_b, 0)$-feasible point of Problem (23); see also Remark 2.

Finally, we show how we can exploit structural properties of the specific model at hand to extend Algorithm 1. The extended variant is capable of computing a tight upper bound on the optimal objective function value of Problem (23) such that we get an ex-post guarantee regarding the optimality gap of the δ-feasible solution obtained by Algorithm 1. To this end, we compare the KKT conditions (29)–(31) of the relaxed lower-level problem (28) with the KKT conditions of the original probabilistic lower level (20). The latter are given by primal feasibility (12b)–(12d) and (19), KKT stationarity (29a), (29c), and

$$\pi_t^{\text{nom}} - \gamma_{i,t}^- + \gamma_{i,t}^+ + \nabla_{q_{i,t}^{\text{nom}}} h(q_-^{\text{nom}})\mu = a_{i,t} + b_{i,t}q_{i,t}^{\text{nom}}, \quad i \in \mathcal{P}_-, \ t \in T, \tag{34}$$

KKT complementarity (30a)–(30c) and (22), as well as KKT nonnegativity conditions (31). We notice two differences when comparing the two sets of KKT conditions. First, instead of the single complementarity condition (22), we obtain the set of complementarity conditions (30d). Second, we note that in the stationarity condition (29b) we approximate the product $\nabla_{q_{i,t}^{\text{nom}}} h(q_-^{\text{nom}})\mu$ that appears in (34) by the

---

Algorithm 2 Bounding of Problem (23).

1: Choose $\delta_b > 0$ and set $r = 0$, $\chi = \infty$.
2: **while** $\chi > \delta_b$ **do**
3:     Solve Problem (35).
4:     **if** Problem (35) is infeasible **then**
5:         Return "Problem (23) is infeasible".
6:     **end if**
7:     Extract q_-^{r+1} and the optimal objective value $\bar{\varphi}^u$ from the solution.
8:     Set $\chi = h(q_-^{r+1})$ and $r \leftarrow r + 1$.
9: **end while**
10: Return $\bar{\varphi}^u$.

---

sum $\sum_{j=1}^{r} \nabla_{q_{i,t}^{\text{nom}}} h(q_-^j) \mu_j$. As a consequence, the single-level problem (32) that we solve in every iteration r of Algorithm 1 (applied to Problem (23)) is neither a relaxation of Problem (23) nor is Problem (32) in iteration r a relaxation of Problem (32) in iteration $r + 1$. This is in line with the previous discussions in Section 2.2 and 2.3. However, by omitting the Constraints (29b), (30d), and (31c) from Problem (32), we obtain the desired relaxation.

Lemma 1. *For every $r \in \mathbb{N}$, the problem*

$$
\begin{aligned}
\max \quad & \varphi^u(q^{\text{nom}}, q) \\
\text{s.t.} \quad & \text{Upper-level feasibility: (11b)–(11d),} \\
& \text{Relaxed lower-level feasibility: (28b)–(28e),} \\
& \text{Parts of relaxed lower-level stationarity: (29a), (29c),} \\
& \text{Parts of relaxed lower-level complementarity: (30a)–(30c),} \\
& \text{Parts of relaxed lower-level nonnegativity: (31a), (31b)}
\end{aligned}
\tag{35}
$$

is a relaxation of the probabilistic bilevel problem (23). In addition, Problem (35) in iteration r is a relaxation of Problem (35) in iteration $r + 1$.

The lemma follows directly from the construction of the problems. We note that Problem (35) relates to relaxing the optimal response of the exit players $i \in \mathcal{P}_-$. For general bilevel problems, relaxing the optimality of the lower-level problem may result in weak bounds on the optimal objective value of the bilevel problem. In our application, however, we observe that the objective functions (11a) and (12a) of the two levels are rather aligned. In fact, the only difference is that in the upper level, we account for the transportation costs $\sum_{t \in T} \sum_{a \in A} c^{\text{trans}}(q_{a,t})$ of the TSO and in the lower level, we account for the booking costs $\sum_{u \in V_+ \cup V_-} \sum_{i \in \mathcal{P}_u} \pi_u^{\text{book}} q_i^{\text{book}}$ of the players. These terms are forced to be equal by the upper-level constraint (11c). Thus, our working hypothesis is that relaxing the optimality of exit players as it is done in Problem (35) still yields a good approximation of the optimal response of the exit players. We thus expect that the bounds obtained by Problem (35) are rather tight. We solve Problem (35) iteratively as described in Algorithm 2. Essentially, this directly applies Kelley's cutting-plane approach [27] to the single level problem (35). Since the hypothesis is that the obtained bound is tight, we also expect that the cuts obtained in Algorithm 2 are useful for the original problem.

In practice, we thus proceed as follows. We invoke Algorithm 2 until it converges. We store the bound $\bar{\varphi}^u$ and initialize (E^0, e^0) in Algorithm 1 with all inequalities (28e) that we generated in Algorithm 2. This corresponds to extending Problem (35) to Problem (32), which we solve in every iteration of Algorithm 1.

If Algorithm 1 then terminates with a δ-feasible point of Problem (23), we compute the gap

$$\frac{|\bar{\varphi}^{\mathrm{u}} - \varphi^{\mathrm{u}}(\bar{q}^{\mathrm{nom}}, \bar{q})|}{|\varphi^{\mathrm{u}}(\bar{q}^{\mathrm{nom}}, \bar{q})|},$$

which gives an ex-post optimality certificate. We analyze the quality of this gap in more detail in the next section.

3.4 Numerical study

In this section, we illustrate the proposed approach using the example of a specific instance of the probabilistic gas market model introduced in Section 3.2. We first give details on this instance, before we evaluate the approach from a computational point of view. Finally, we interpret the results that we obtain for various probability levels of the chance constraint.

Physical and economic setup

For our analysis, we consider a network with eleven nodes as shown in Figure 1. Three of the nodes are entries at which gas is injected to the network, three others are exits at which gas is withdrawn from the network, and the remaining nodes are so-called inner nodes. The nodes are connected by eleven arcs and the network contains one cycle. The length L_a of each pipe is specified in Figure 1. In addition, we
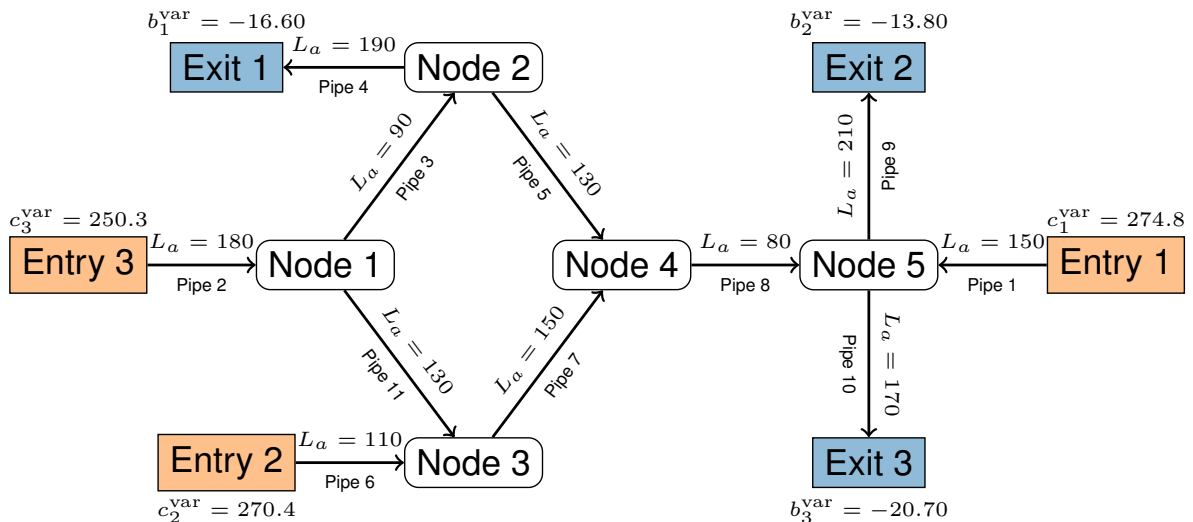

Figure 1: Gas network considered in the numerical study with variable costs c_i^{var} (in $\mathrm{EUR}/(1000\,\mathrm{Nm}^3/\mathrm{h})$) of entries, slopes b_i (in $\mathrm{EUR}/(1000\,\mathrm{Nm}^3/\mathrm{h})^2$) of exits, and length L_a (in km) of pipes.

assume an equal diameter $D_a = 0.5\,\mathrm{m}$, roughness $k_a = 0.1\,\mathrm{mm}$, and capacities $q_a^{\pm} = \pm 266\,\mathrm{kg\,s}^{-1}$ for all pipes $a \in A$. With respect to the flow model, we follow [5]. This means, we consider stationary gas flow, i.e., we abstract from temporal dependencies and only consider horizontal pipes. In this setting, we can use the pressure loss function

$$\Phi(q_a) = \frac{\lambda_a c^2 L_a}{(0.25\pi)^2 D_a^5}|q_a|q_a, \tag{36}$$

see, e.g., [19], in which the constant $c = 340 \, \mathrm{m \, s^{-1}}$ denotes the speed of sound in natural gas. The friction coefficient λ_a can be approximated by the formula of Nikuradse; see [19]:

$$\lambda_a = \left(2 \log_{10} \left(\frac{D_a}{k_a} \right) + 1.138 \right)^{-2} .$$

Equation (36) is a suitable simplification of gas flow physics but is still nonlinear. In order to arrive at a linear approximation, see Assumption 3 and Remark 3, we replace $|q_a|$ by a mean flow $q_a^{\mathrm{mean}} = 100 \, \mathrm{kg \, s^{-1}}$. Again, exactly as done in [5], we set $c_t^{\mathrm{trans}} = 1$.

For all inner nodes u, we have lower and upper pressure bounds of 15 bar and 140 bar. At all entries we have a lower pressure bound of 20 bar. Further, at Entry 1 and Entry 2 we have an upper pressure bound of 80 bar, and at Entry 3 we have 160 bar. At Exit 1, Exit 2, and Exit 3, we have lower pressure bounds of 20 bar, 10 bar, and 5 bar, and all upper pressure bounds are set to 120 bar. We note that the pressure bounds are rather large such that the network is not very restrictive. The reason behind this choice is that we want to analyze the effect of uncertain loads. To this end, we analyze various choices of $p \in [0, 1]$ in the chance constraint (18). Hedging against the uncertainty increases nominated quantities, especially for p close to 1. Assuming large pressure bounds prevents technical infeasibilities that may otherwise result from large nominations.

In our example, we consider one player at each of the entry and exit nodes and $|T| = 12$ time steps, which refer to the months of a year. We thus set the initial willingness to pay, which corresponds to the intercepts $a_{i,t}$, higher in winter months than in summer months. For a fixed time step $\bar{t}$, fluctuations of the intercepts $a_{i,\bar{t}}$ across the players $i \in \mathcal{P}_-$ are rather small. We specify the exact choices for $a_{i,t}$ in Table 2 in the appendix. In contrast to the intercepts, we assume that the price elasticity, i.e., the slopes $b_{i,t}$ of the exit players, are constant over time, and specify $b_i = b_{i,t}$ in Figure 1. Exit 2 has a rather elastic demand, while the demand of Exit 3 is the least elastic. For the gas sellers, we set the variable costs to be constant over time as specified in Figure 1. Entries 1 and 2 have similar variable costs, while Entry 3 has a significantly lower variable cost.

The vector ξ of loads at the exits is random and we assume that it follows a multivariate Gaussian distribution $\xi \sim \mathcal{N}(m, \Sigma)$ with mean vector m and covariance matrix Σ. In reality, also other types of distributions (e.g., log-normal, uniform, discrete) and mixtures thereof may be relevant [31, Chapter 13]. When specifying the parameters of this distribution, we do not rely on statistical data analysis of a concrete real-life network. We rather follow the idea that the mean vector of loads represents the equilibrium exit nominations for the deterministic bilevel model from Section 3.1, thus modeling an elastic demand, which changes only slowly over time. Hence, our mean load is—among other parameters—a consequence of the inverse demand functions $P_{i,t}$ as specified in Table 2. For real-world problems, one would rather solve an inverse problem and calibrate the inverse demand functions in a way to derive a specific mean vector as a solution of the deterministic bilevel problem. We specify our actual choices of m and Σ in the appendix.

The load vector ξ itself can be understood as a random inelastic deviation from its mean, which is more relevant on the shorter time scale of a day-ahead market. This scattering around the mean is defined by the covariance matrix. In order to fix the covariance matrix, it is sufficient to define standard deviations of the single time- and node-components as well as correlations between these components. Orientation for both features can be obtained from general gas load data as analyzed in [31]. As a consequence, we built up our covariance matrix in a fairly heuristic way (zero correlation between exits at different times, common positive value of correlation between exits at fixed time, common relative standard deviation between exits, but constant over time). Clearly, the dimension of the random vector (or its multivariate distribution) is the same as that of the vector of exit nominations, namely 36 (3 exits combined with 12 time

steps). All data required to reproduce our experiments are publicly available at `https://github.com/m-schmidt-math-opt/bilevel-w-chance-constr-lower-level`.

Computational Analysis

In this section we briefly evaluate the tailored solution approach presented in Section 3.3 from a computational perspective. We implemented the approach in Python 3.7 and used Gurobi 9.1.1 to solve all involved optimization problems. All computations have been carried out on a compute cluster; see [45] for details on the installed hardware.

In order to have a benchmark, we first consider the deterministic model from Section 3.1, which can be solved by global solvers out-of-the-box. This model has 930 variables, thereof 282 binaries, and 1260 constraints. It is solved in $1.17\,\text{s}$ and can thus be considered easy.

We now turn to the probabilistic model for various probability levels p of the chance constraint (18). The model sizes increase slightly compared to the deterministic model, because we obtain a new complementarity condition in every iteration of the algorithm. Since we linearize the complementarity conditions, we particularly obtain an additional binary variable in every iteration.

For our computations, we set $\delta_b = 10^{-3}$, i.e., we terminate Algorithm 1 and 2 if $h(q_-^{\text{nom}}) \le 10^{-3}$ holds. Preliminary results showed that it is very important to equip Algorithm 2 with initial cuts. The reason is that without an initial cut, we obtain a solution from the bounding problem (35), for which it is very likely that $\mathbb{P}\left(\xi_{i,t} \le q_{i,t}^{\text{nom}} \text{ for all } i \in \mathcal{P}_-,\ t \in T\right)$ is close to zero. Consequently, we would linearize the chance constraints at the tails of the Gaussian distribution, where the gradient of the chance constraint is almost zero. For the log-transformed chance constraint (19), this results in partial derivatives going to infinity such that we cannot construct the cuts (27). The situation can be resolved by adding initial valid inequalities that bound the nominations appropriately from below. A straightforward way to achieve this is to exploit the p-quantile Q_p of the standard Gaussian distribution as follows:

$$q_{i,t}^{\text{nom}} \ge m_{i,t} + Q_p \sqrt{\Sigma_{(i,t),(i,t)}} \quad \text{for all } i \in \mathcal{P}_-,\ t \in T. \tag{37}$$

These "quantile cuts" correspond to individual p-level constraints on load coverage for all exit players and for all time steps separately, i.e.,

$$\mathbb{P}\left(\xi_{i,t} \le q_{i,t}^{\text{nom}}\right) \ge p \quad \text{for all } i \in \mathcal{P}_-,\ t \in T, \tag{38}$$

which are clearly valid for the original chance constraint. Another approach is to add a single first-order cut (27) at a point $\tilde{q}_-^{\text{nom}}$ with
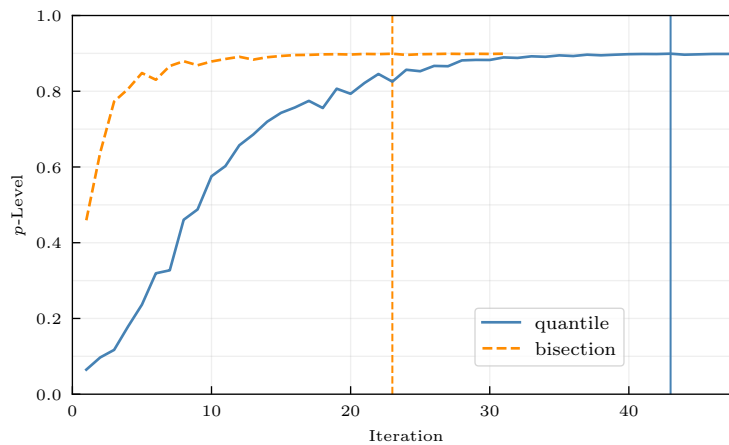
$$p - \delta_p \le \mathbb{P}\left(\xi_{i,t} \le \tilde{q}_-^{\text{nom}} \text{ for all } i \in \mathcal{P}_-,\ t \in T\right) \le p$$

for a given $\delta_p > 0$. We can find such a point, e.g., via bisection. The latter approach turned out to be very effective in our computations such that we apply it by default.

In Table 1, we present the numbers of iterations and runtimes in total and separated for the three phases of our tailored algorithm, i.e., bisection to find an initial cut, the bounding procedure Algorithm 2, and the feasibility procedure Algorithm 1. We observe several aspects in Table 1. First, in general, the larger fraction of the total iterations and runtimes is spent in the bounding procedure, while mostly only a few iterations are needed afterward to compute a δ-feasible point. However, the bounding problems are rather cheap such that we cannot say that the bounding phase requires more runtime than the feasibility phase. Second, we cannot observe a relationship between the p-level and the total runtime,

Table 1: Iterations, runtimes (in s), and optimality gaps (in $\%$) for various choices of the probability level p.

	Bisection	Bounding		δ-Feasibility		Total		
p	Runtime	#Iter.	Runtime	#Iter.	Runtime	#Iter.	Runtime	Gap
0.60	12.13	32	36.80	10	28.97	42	77.9	0.001
0.65	14.15	28	32.00	16	40.71	44	86.86	0.001
0.70	11.13	26	29.70	13	39.70	39	80.53	0.001
0.75	9.04	25	28.55	6	14.19	31	51.78	0.002
0.80	7.98	25	29.06	4	6.26	29	43.3	0.005
0.85	11.08	21	24.01	3	7.41	24	42.5	0.006
0.90	11.05	23	26.34	8	27.52	31	64.91	0.017
0.95	5.96	24	27.99	6	14.14	30	48.09	0.010
0.99	6.96	26	29.76	10	1250.65	36	1287.37	0.187

Figure 2: Evolution of the p-level and the ex-post computed gap over the iterations for the probabilistic model with $p = 0.9$.

i.e., in general a larger p-level does not result in longer runtimes. The long runtime for $p = 0.99$ is caused by a single iteration of Algorithm 1, such that we consider it an outlier. Third, we spent a significant amount of the runtime to find an initial cut via bisection. We already discussed that initial cuts are crucial for the applicability of the approach. However, the quantile cuts (37) can be added for free such that it appears to be a reasonable option to add these cuts. We shed some more light on this by discussing the evolution of the p-level over the iterations of our approach for both cases, i.e., using quantile cuts or running an initial bisection. By p-level, we mean the value of the original chance constraint (18) when it is evaluated at the solution of the respective iteration. Figure 2 demonstrates this comparison exemplary for the instance $p = 0.90$. We see that we start with a much higher p-level when we equip our approach with the bisection phase. This results in a faster convergence to the desired p-level. Ultimately, this reduction in iterations overcompensates the expensive bisection. This justifies to use the bisection instead of the quantile cuts.

Another conclusion from Figure 2 can be drawn with respect to the usefulness of the cuts generated in the bounding phase. The vertical lines in Figure 2 mark the end of the bounding phase. While it takes several iterations in the bounding phase to finally arrive at the desired p-level, we remain close to this p-level when entering the feasibility phase. This behavior can be observed for all tested instances

and means that the cuts generated in the bounding phase are very useful in Algorithm 1, most likely because they linearize the log-transformed chance constraint around the δ-feasible point to which we converge.

We now turn back to Table 1 and discuss the most important observation. Except $p = 0.99$, for which we terminate with a gap of $0.187\,\%$, we always terminate with a negligibly small gap that allows to consider the computed solutions to be approximate optimal. Thus, the computational results allow to use our tailored approach from Section 3.3 in a case study to analyze and interpret the effects of uncertain loads on economic quantities like nominations and total welfare.

Case Study

In the following, we discuss the outcomes of our computations for different constraints on load coverage. The first instance is defined as the solution of the nominal problem without any load coverage constraint. As noted before, this solution corresponds to mean exit nominations. The second instance imposes individual $90\,\%$-constraints on load coverage for all exit players and for all time steps as it is done in Constraint (37), respectively (38). As discussed, the advantage of this quantile-approach relies on the fact that it is as easy to obtain as the nominal solution, where the lower bound is just zero. The drawback is that these solutions—while yielding robust load coverage for each exit player and in each time step separately, may be far from guaranteeing robust load coverage for all exit players and times simultaneously. This observation—which will be supported by our results below—is not surprising because only a very restricted part of the underlying multivariate distribution is taken into account, i.e., no covariances between load components are considered. Our main interest is about the third instance, namely the joint (over all times and exits) probabilistic load coverage, which results in the probabilistic constraint (18). In contrast to the quantile-based solution, robust load coverage can be achieved. As seen in the last section, the robust solution of the chance-constrained problem comes at the cost of an additional computational burden. In the following we evaluate the impact of different probability levels $p \in [0, 1]$ as detailed in Table 1.

Figure 3 shows the time-dependent profiles of nominations at the three exits for the mean vector, the 0.9-quantile constraint and the probabilistic constraints with safety levels from the indicated range. It can be observed how much nominations have to be increased in order to satisfy the desired safety. It can be seen that, while the mean profiles are just scaled/shifted versions of each other, which actually is a consequence of keeping the slopes of the inverse demand functions $P_{i,t}$ in our model time-invariant, the other nomination profiles change their shape over time and exits.

Figure 4 provides a plot of total welfare (11a) and of the fee μ for uniform load coverage occurring in (21) (also interpreted as the "price of stochasticity") as a function of the chosen probability level. We reiterate that in our iterative solution approach, μ is approximated by a series of μ_j. However, we can compute the "real" fee μ ex-post via the KKT stationarity condition (34) of the original probabilistic lower-level problem (20). It can be seen that the loss of welfare is moderate when increasing the safety towards 0.9, whereas it severely decreases when driving safety further towards 1. A similar pattern is observed for the price of uniform load coverage. These diagrams suggest that in the given example, $90\,\%$ safety of load coverage would represent a reasonable compromise between the TSO's interest in nominations covering the random load on the one hand and the welfare losses induced by this safety of load coverage.

The probabilistic effect of exit nominations is illustrated in Figure 5: The black thick curves correspond to nominations according to mean (top), 0.9-quantile solution (middle), and probabilistic constraint for $p = 0.9$ (bottom). In order to visualize load coverage by these nomination profiles, we simulated ten
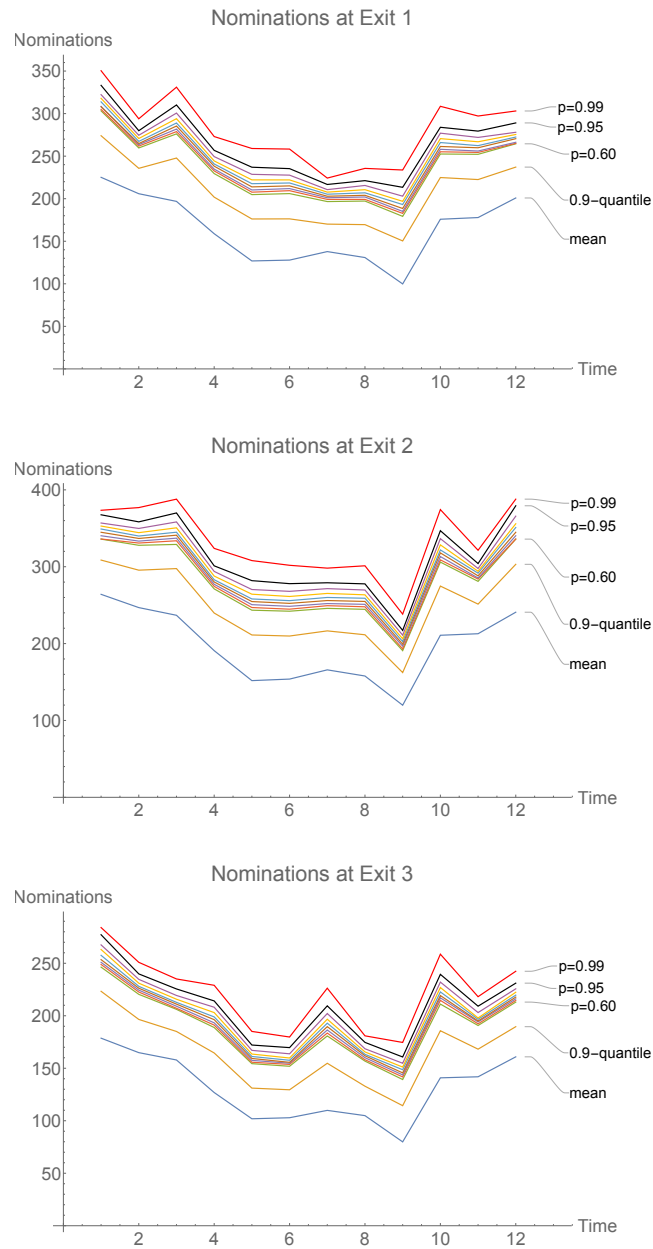
Figure 3: Time-dependent profiles of nominations at the three exits for the mean vector, the 0.9-quantile constraint and the probabilistic constraints with varying safety levels.
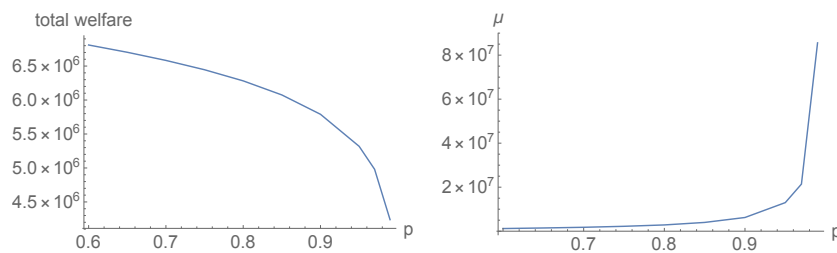

Figure 4: Total welfare and price of load coverage as functions of safety level.
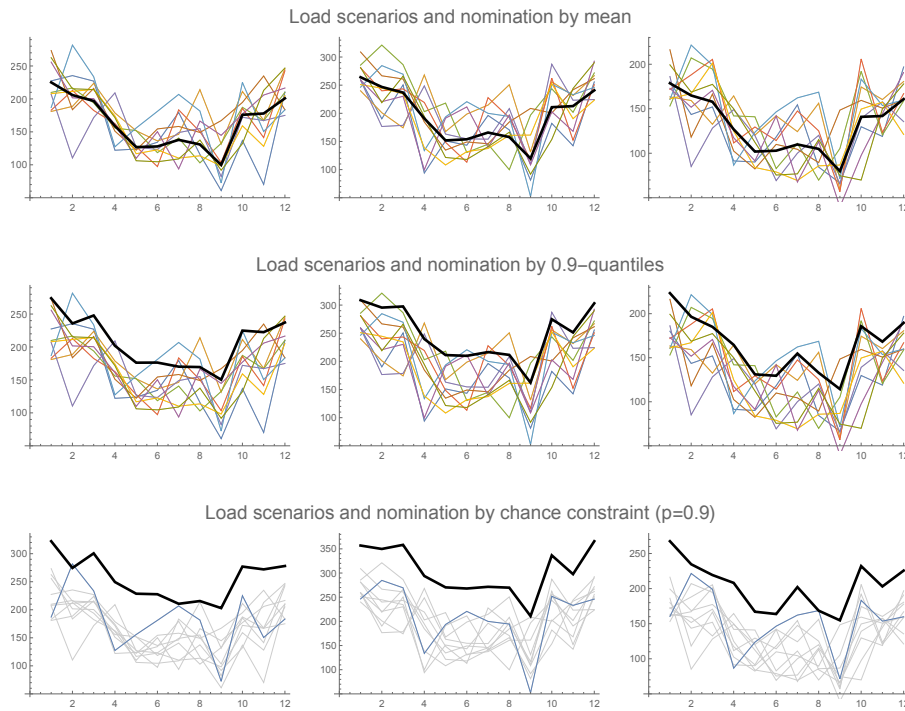
Figure 5: Nomination profiles of Exit 1 (left), Exit 2 (middle) and Exit 3 (right) for mean (top), 0.9-quantile-solution (middle), and probabilistic constraint ($p = 0.9$) (bottom) along with ten simulated load scenarios (colored: violating; gray: feasible).

scenarios for load profiles according to the given multivariate Gaussian distribution. Note that each scenario is related to all exits and all times. Hence, for instance the three thin blue curves in the lower diagrams correspond to one scenario. Nominating according to the mean load not surprisingly yields that at each exit and each time the load of approximately one half of the scenarios is covered by the nomination. On the other hand, each of the ten scenarios is not covered by the nomination at some exit at some time (such scenarios not uniformly covered by the nomination are colored in the diagrams). This means that the TSO has to expect almost surely uncovered load in the network over the considered time horizon. In contrast, the quantile-based exit nomination guarantees at each exit and each time separately a load coverage for all but one (on average) scenarios, which corresponds well to the chosen 90 %-quantile. However, uniform load coverage is as poor as for the mean-based nomination: none of the ten scenarios is uniformly covered for all exits and times (i.e., all curves are colored). Things change when imposing a joint chance constraint: here, nine out of ten scenarios are uniformly covered by the corresponding nomination profile. Only one scenario (colored in blue) remains unsatisfied at Exits 1 and 3 at time 2.

Finally, the nomination counterparts of entries are plotted in Figure 6 for six different safety instances. The arrangement in these diagrams is different from that of the exits in Figure 3: Here, we opposed the three entries within each diagram. Three major observations can be made.

1 The sum of entry nominations increases with increasing safety required.

2 Entries 2 and 3 exhibit an almost constant time profile in all instances.

3 Both entries change their role from inactive to strongly active when increasing the safety level.
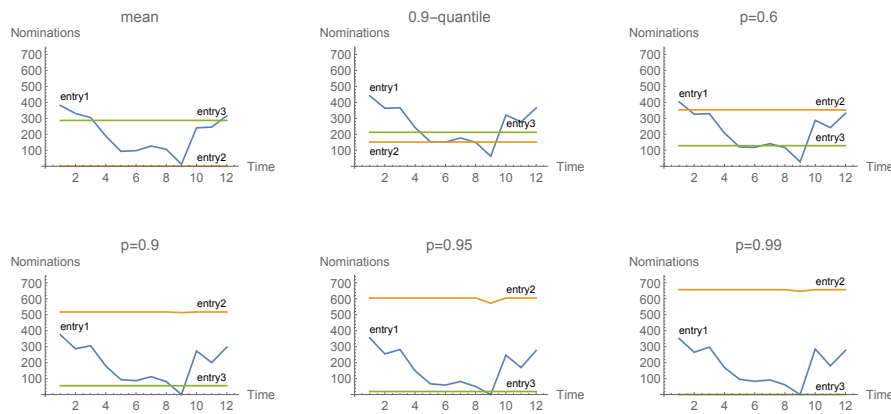
Figure 6: Profiles of entry nominations for selected safety requirements.

The first observation is not surprising since exit nominations increase as well and both entry and exit nominations have to be balanced. The second observation is due to technical limitations in combination with economic reasons. The maximum throughput of the network is limited mainly by pressure bounds. The technical capacities are set to ensure that the pressure bounds are fulfilled. Entry 2 and 3 have lower variable costs compared to Entry 1 such that it economically makes sense to use as much gas as possible from these two entries. Since total demand in most cases exceeds the technical capacities at Entries 2 and 3, we observe an almost constant profile.

Finally, we turn to observation 3, which is the least intuitive. For low safety levels, the cheapest entry, Entry 3, is supplying a lot of gas. With higher safety levels, Entry 3 is substituted by the significantly more expensive Entry 2, until Entry 3 is driven out of the market entirely. One possible explanation for this behavior might be as follows. Whenever Entry 3 supplies large amounts of gas, the pressure loss over Pipe 2 results in a low pressure at Node 1; see Figure 1. When the nominations increase, more gas has to be transported from left to right—in particular, more gas than Entry 3 is able to deliver due to technical limitations. Thus, Entry 2, the second cheapest producer, needs to step in to supply Exit 2 and 3. However, the low pressure at Node 1 forces Entry 2 to "send" gas over Pipe 11 (and thus over Pipe 3 and Pipe 5 to Node 4) to satisfy all pressure loss constraints. This results in very high transportation costs compared to "sending" gas directly via Pipe 7 to Node 4. Consequently, with higher overall nominations, it becomes more attractive for the TSO to shut down Entry 3 entirely, because its cheaper production is overcompensated by higher overall transportation costs. This underlines also the complicated interplay of physics and economics.

4   Conclusion

We considered bilevel optimization problems with convex lower levels that involve black-box constraints, i.e., constraint functions for which no closed form is available. To tackle such problems, we developed a cutting-plane algorithm that can compute approximate bilevel-feasible points. Afterward, we applied this method to a bilevel model of the European gas market in which we model uncertain gas loads via a joint chance constraint in the lower level. Since such a constraint cannot be stated in closed form, this setting fits the setup of a bilevel problem with a lower-level black-box constraint. Using further problem-specific techniques we can also show that the computed approximate feasible points are indeed approximate

Table 2: Intercepts $a_{i,t}$ (in EUR$/(1000\,\mathrm{Nm}^3/\mathrm{h})$) and slopes b_i (in EUR$/(1000\,\mathrm{Nm}^3/\mathrm{h})^2$) of the exit players.

	Exit 1	Exit 2	Exit 3
$a_{i,1}$	4040.60	4039.80	4039.80
$a_{i,2}$	3692.00	3681.00	3687.90
$a_{i,3}$	3542.60	3543.00	3543.00
$a_{i,4}$	2911.80	2908.20	2901.30
$a_{i,5}$	2380.60	2370.00	2383.80
$a_{i,6}$	2397.20	2397.60	2404.50
$a_{i,7}$	2563.20	2563.20	2549.40
$a_{i,8}$	2447.00	2452.80	2445.90
$a_{i,9}$	1932.40	1928.40	1928.40
$a_{i,10}$	3194.00	3184.20	3191.10
$a_{i,11}$	3227.20	3211.80	3211.80
$a_{i,12}$	3609.00	3598.20	3605.10
b_i	-16.60	-13.80	-20.70

optimal as well, which allowed to analyze the sensitivity of welfare outcomes in dependence of the achieved safety level for load coverage.

The considered setting is a rather new sub-field in bilevel optimization. Thus, many aspects can be elaborated on further in the future. Let us sketch two of them. First, it might be interesting to study a bilevel setting in which the lower-level black-box also depends on the leader's decision. In the context of chance constraints, this can be seen as a variant of decision-dependent uncertainty in the lower-level problem. Second, we are sure that other algorithmic approaches to bilevel problems with black-box constraints are possible and every algorithmic improvement will be beneficial for tackling relevant problems from practice.

A Appendix Data

In Table 2 we specify the intercepts $a_{i,t}$ and slopes b_i of the exit players. We also specify the mean vector m of the multivariate Gaussian distribution of the load vector ξ. For better readability, we denote the vector in parts $m_i = (m_{i,t})_{t \in T}$ for all $i \in \mathcal{P}_-$:

$$m_1 = \begin{pmatrix} 225.11 \\ 205.86 \\ 196.86 \\ 158.86 \\ 126.86 \\ 127.86 \\ 137.86 \\ 130.86 \\ 99.86 \\ 175.86 \\ 177.86 \\ 200.86 \end{pmatrix}, \quad m_2 = \begin{pmatrix} 264.08 \\ 246.83 \\ 236.83 \\ 190.83 \\ 151.83 \\ 153.83 \\ 165.83 \\ 157.83 \\ 119.83 \\ 210.83 \\ 212.83 \\ 240.83 \end{pmatrix}, \quad m_3 = \begin{pmatrix} 178.65 \\ 164.88 \\ 157.88 \\ 126.88 \\ 101.88 \\ 102.88 \\ 109.88 \\ 104.88 \\ 79.88 \\ 140.88 \\ 141.88 \\ 160.88 \end{pmatrix}.$$

Finally, we denote the covariance matrix Σ of the loads ξ. For better readability, we only denote the $|T|$-many $|\mathcal{P}_-| \times |\mathcal{P}_-|$ nonzero blocks Σ_t:

$$\Sigma_1 = \begin{bmatrix} 1455.73 & 796.24 & 797.54 \\ 796.24 & 1209.78 & 727.06 \\ 797.54 & 727.06 & 1213.74 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 542.31 & 530.88 & 344.84 \\ 530.88 & 1443.57 & 562.62 \\ 344.84 & 562.62 & 609.11 \end{bmatrix},$$

$$\Sigma_3 = \begin{bmatrix} 1579.69 & 1130.54 & 505.62 \\ 1130.54 & 2247.48 & 603.09 \\ 505.62 & 603.09 & 449.54 \end{bmatrix}, \quad \Sigma_4 = \begin{bmatrix} 1124.77 & 768.18 & 591.80 \\ 768.18 & 1457.33 & 673.63 \\ 591.80 & 673.63 & 864.93 \end{bmatrix},$$

$$\Sigma_5 = \begin{bmatrix} 1480.80 & 1069.88 & 526.31 \\ 1069.88 & 2147.19 & 633.76 \\ 526.31 & 633.76 & 519.61 \end{bmatrix}, \quad \Sigma_6 = \begin{bmatrix} 1429.86 & 991.65 & 471.21 \\ 991.65 & 1910.37 & 544.66 \\ 471.21 & 544.66 & 431.35 \end{bmatrix},$$

$$\Sigma_7 = \begin{bmatrix} 634.58 & 598.06 & 529.14 \\ 598.06 & 1565.69 & 831.16 \\ 529.14 & 831.16 & 1225.62 \end{bmatrix}, \quad \Sigma_8 = \begin{bmatrix} 910.52 & 757.94 & 395.81 \\ 757.94 & 1752.57 & 549.14 \\ 395.81 & 549.14 & 477.95 \end{bmatrix},$$

$$\Sigma_9 = \begin{bmatrix} 1554.29 & 784.02 & 636.37 \\ 784.02 & 1098.56 & 535.00 \\ 636.37 & 535.00 & 723.74 \end{bmatrix}, \quad \Sigma_{10} = \begin{bmatrix} 1463.03 & 1146.00 & 803.26 \\ 1146.00 & 2493.50 & 1048.66 \\ 803.26 & 1048.66 & 1225.05 \end{bmatrix},$$

$$\Sigma_{11} = \begin{bmatrix} 1213.19 & 628.39 & 429.07 \\ 628.39 & 904.13 & 370.41 \\ 429.07 & 370.41 & 421.52 \end{bmatrix}, \quad \Sigma_{12} = \begin{bmatrix} 801.40 & 825.66 & 380.58 \\ 825.66 & 2362.90 & 653.50 \\ 380.58 & 653.50 & 502.05 \end{bmatrix}.$$

B   Appendix Big-M

In this section, we derive an upper bound on the Lagrangian multipliers μ_j as used in Section 3.2. To this end, we consider the stationarity condition (29b), i.e.,

$$\pi_t^{\mathrm{nom}} - \gamma_{i,t}^- + \gamma_{i,t}^+ + \sum_{j=1}^r \nabla_{q_{i,t}^{\mathrm{nom}}} h(q_-^j) \mu_j = a_{i,t} + b_{i,t} q_{i,t}^{\mathrm{nom}}, \quad i \in \mathcal{P}_-,\ t \in T.$$

First, we assume $q_{i,t}^{\mathrm{nom}} > 0$ for all $i \in \mathcal{P}_-$ and $t \in T$, which is justified in the probabilistic setting: According to (37), we can bound the nominations of the exit players from below. With this, we obtain

$$\pi_t^{\mathrm{nom}} + \gamma_{i,t}^+ + \sum_{j=1}^r \nabla_{q_{i,t}^{\mathrm{nom}}} h(q_-^j) \mu_j = a_{i,t} + b_{i,t} q_{i,t}^{\mathrm{nom}}, \quad i \in \mathcal{P}_-,\ t \in T.$$

From now on, we consider an arbitrary but fixed index j. Since the multivariate Gaussian distribution function Φ is strictly positive, it follows from (25) that $\nabla_{q_{i,t}^{\mathrm{nom}}} h(q_-^j) < 0$ holds for all $i \in \mathcal{P}_-, t \in T$, and $j = 1, \ldots, r$, we can solve for μ_j and obtain

$$\mu_j = \frac{a_{i,t} + b_{i,t} q_{i,t}^{\mathrm{nom}} - \pi_t^{\mathrm{nom}} - \gamma_{i,t}^+ - \sum_{k \in \{1,\ldots,r\}\setminus\{j\}} \nabla_{q_{i,t}^{\mathrm{nom}}} h(q_-^k) \mu_k}{\nabla_{q_{i,t}^{\mathrm{nom}}} h(q_-^j)}, \quad i \in \mathcal{P}_-,\ t \in T.$$

Note that all $\nabla_{q_{i,t}^{\mathrm{nom}}} h(q_-^j)$ for all $i \in \mathcal{P}_-, t \in T$, and $j = 1, \ldots, r$ are given (and negative) constants. Hence, to obtain a valid upper bound on μ_j we first have to properly bound the other terms in the

numerator. Since we divided by a negative number, a valid upper for μ_j can be obtained by taking the smallest possible numerator. Due to $b_{i,t} < 0$, a lower bound for

$$a_{i,t} + b_{i,t} q_{i,t}^{\text{nom}}$$

is given by

$$L_1 = a_{i,t} + b_{i,t} \bar{q}_{i,t}^{\text{nom}},$$

where $\bar{q}_{i,t}^{\text{nom}}$ is an upper bound for $q_{i,t}^{\text{nom}}$. Taking the network constraints into account, we can set

$$\bar{q}_{i,t}^{\text{nom}} = \sum_{a \in \delta_u^{\text{in}}} q_a^+ - \sum_{a \in \delta_u^{\text{out}}} q_a^-,$$

where $u \in V$ is the node in the network at which player $i \in \mathcal{P}_-$ is located.

Let us now further assume that the overall booking price is bounded above, i.e.,

$$\underline{\pi}_u^{\text{book}} + \pi_u^{\text{book}} \leq L_2$$

for some $L_2 > 0$ and all $i \in \mathcal{P}_u$, $u \in V_+ \cup V_-$. Thus, with $\gamma_{i,t}^+ \geq 0$ and (29a), we can also bound all $\gamma_{i,t}^+$ from above by L_2. By using (29c), we then see that

$$L_3 = \max_{i \in \mathcal{P}_+} \left\{ c_i^{\text{var}} \right\} + L_2$$

is a valid upper bound for the resulting market equilibrium price π_t^{nom}.

The last term,

$$- \sum_{k \in \{1,\ldots,r\} \setminus \{j\}} \nabla_{q_{i,t}^{\text{nom}}} h(q_-^k) \mu_k,$$

in the numerator get as small as possible for $\mu_k = 0$ for all $k \in \{1, \ldots, r\} \setminus \{j\}$ since $\nabla_{q_{i,t}^{\text{nom}}} h(q_-^k) < 0$.

Taking this all together, we see that

$$\mu_j \leq L^j \quad \text{with} \quad L^j = \frac{L_1 - L_2 - L_3}{\nabla_{q_{i,t}^{\text{nom}}} h(q_-^j)}$$

holds. Note that we can, in principle, do this derivation for every $i \in \mathcal{P}_-$ and $t \in T$ to get the tightest possible upper bound L^j.

References

[1] W. van Ackooij and R. Henrion. „Gradient Formulae for Nonlinear Probabilistic Constraints with Gaussian and Gaussian-Like Distributions". In: *SIAM Journal on Optimization* 24.4 (2014), pp. 1864–1889. DOI: `10.1137/130922689`.

[2] W. van Ackooij and R. Henrion. „(Sub-)Gradient Formulae for Probability Functions of Random Inequality Systems under Gaussian Distribution". In: *SIAM/ASA J. Uncertain. Quantification* 5.1 (2017), pp. 63–87. DOI: `10.1137/16M1061308`.

[3] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati. „On probabilistic constraints induced by rectangular sets and multivariate normal distributions". In: *Math. Methods Oper. Res.* 71 (2010), pp. 535–549. DOI: `10.1007/s00186-010-0316-3`.

[4]   A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, Princeton, NJ, 2009. DOI: `10.1515/9781400831050`.

[5]   T. Böttger, V. Grimm, T. Kleinert, and M. Schmidt. „The Cost of Decoupling Trade and Transport in the European Entry-Exit Gas Market". 2020. URL: `http://www.optimization-online.org/DB_HTML/2020/06/7851.html`.

[6]   J. Burtscheidt and M. Claus. „Bilevel Linear Optimization Under Uncertainty". In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, 2020, pp. 485–511. DOI: `10.1007/978-3-030-52119-6_17`.

[7]   J. Burtscheidt, M. Claus, and S. Dempe. „Risk-Averse Models in Bilevel Stochastic Linear Programming". In: *SIAM Journal on Optimization* 30.1 (2020), pp. 377–406. DOI: `10.1137/19M1242240`.

[8]   S. Dempe. *Foundations of Bilevel Programming*. Springer, 2002. DOI: `10.1007/b101970`.

[9]   S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova. *Bilevel Programming Problems*. Springer, 2015. DOI: `10.1007/978-3-662-45827-3`.

[10] S. Dempe and S. Lohse. „Dependence Of Bilevel Programming On Irrelevant Data". 2011.

[11] S. T. DeNegre and T. K. Ralphs. „A branch-and-cut algorithm for integer bilevel linear programs". In: *Operations research and cyber-infrastructure*. Springer, 2009, pp. 65–78. DOI: `10.1007/978-0-387-88843-9_4`.

[12] M. A. Duran and I. E. Grossmann. „An outer-approximation algorithm for a class of mixed-integer nonlinear programs". In: *Mathematical Programming* 36.3 (1986), pp. 307–339. DOI: `10.1007/BF02592064`.

[13] European Parliament and Council of the European Union. *Directive 2009/73/EC of the European Parliament and of the Council concerning common rules for the internal market in natural gas and repealing Directive 2003/55/EC*. 2009.

[14] European Parliament and Council of the European Union. *Regulation No 715/2009 of the European Parliament and of the Council on conditions for access to the natural gas transmission networks and repealing Regulation No 1775/2005*. July 13, 2009.

[15] M. Fischetti, I. Ljubić, M. Monaci, and M. Sinnl. „A New General-Purpose Algorithm for Mixed-Integer Bilevel Linear Programs". In: *Operations Research* 65.6 (2017), pp. 1615–1637. DOI: `10.1287/opre.2017.1650`.

[16] M. Fischetti, I. Ljubić, M. Monaci, and M. Sinnl. „On the use of intersection cuts for bilevel optimization". In: *Mathematical Programming* 172.1-2 (2018), pp. 77–103. DOI: `10.1007/s10107-017-1189-5`.

[17] R. Fletcher and S. Leyffer. „Solving mixed integer nonlinear programs by outer approximation". In: *Mathematical Programming* 66.1 (1994), pp. 327–349. DOI: `10.1007/BF01581153`.

[18] J. Fortuny-Amat and B. McCarl. „A Representation and Economic Interpretation of a Two-Level Programming Problem". In: *The Journal of the Operational Research Society* 32.9 (1981), pp. 783–792. JSTOR: `2581394`.

[19] A. Fügenschuh, B. Geißler, R. Gollmer, A. Morsi, M. E. Pfetsch, J. Rövekamp, M. Schmidt, K. Spreckelsen, and M. C. Steinbach. „Physical and technical fundamentals of gas networks". In: *Evaluating Gas Network Capacities*. Ed. by T. Koch, B. Hiller, M. E. Pfetsch, and L. Schewe. SIAM-MOS series on Optimization. SIAM, 2015. Chap. 2, pp. 17–44. DOI: `10.1137/1.9781611973693.ch2`.

[20] A. Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Heidelberg: Springer-Verlag, 2009. DOI: `10.1007/978-3-642-01689-9`.

[21] V. Grimm, L. Schewe, M. Schmidt, and G. Zöttl. „A Multilevel Model of the European Entry-Exit Gas Market". In: *Mathematical Methods of Operations Research* 89.2 (2019), pp. 223–255. DOI: `10.1007/s00186-018-0647-z`.

[22] P. Hansen, B. Jaumard, and G. Savard. „New branch-and-bound rules for linear bilevel programming". In: *SIAM Journal on scientific and Statistical Computing* 13.5 (1992), pp. 1194–1217. DOI: `10.1137/0913069`.

[23] H. Heitsch. „On probabilistic capacity maximization in a stationary gas network". In: *Optimization* 69.3 (2020), pp. 575–604. DOI: `10.1080/02331934.2019.1625353`.

[24] R. Henrion and W. Römisch. „On M-stationary points for a stochastic equilibrium problem under equilibrium constraints in electricity spot market modeling". In: *Appl. Math.* 52.6 (2007), pp. 473–494. DOI: `10.1007/s10492-007-0028-z`.

[25] R. Horst and H. Tuy. *Global optimization: Deterministic approaches*. Springer Science & Business Media, 2013. DOI: `10.1007/978-3-662-03199-5`.

[26] S. Ivanov. „A Bilevel Stochastic Programming Problem with Random Parameters in the Follower's Objective Function". In: *J. Appl. Ind. Math* 12.4 (2018), pp. 27–45. DOI: `10.1134/S19904789918040063`.

[27] J. E. Kelley Jr. „The Cutting-Plane Method for Solving Convex Programs". In: *Journal of the Society for Industrial and Applied Mathematics* 8.4 (1960), pp. 703–712. DOI: `10.1137/0108053`.

[28] T. Kleinert, M. Labbé, I. Ljubić, and M. Schmidt. *A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization*. Tech. rep. 2021. URL: `http://www.optimization-online.org/DB_HTML/2021/01/8187.html`.

[29] T. Kleinert, M. Labbé, F. Plein, and M. Schmidt. „Technical Note—There's No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization". In: *Operations Research* 68.6 (2020), pp. 1716–1721. DOI: `10.1287/opre.2019.1944`.

[30] T. Kleinert and M. Schmidt. „Global optimization of multilevel electricity market models including network design and graph partitioning". In: *Discrete Optimization* 33 (2019), pp. 43–69. DOI: `10.1016/j.disopt.2019.02.002`.

[31] T. Koch, B. Hiller, M. E. Pfetsch, and L. Schewe. *Evaluating Gas Network Capacities*. Ed. by M. E. Pfetsch, T. Koch, L. Schewe, and B. Hiller. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015. DOI: `10.1137/1.9781611973693`.

[32] M. Labbé, F. Plein, and M. Schmidt. „Bookings in the European Gas Market: Characterisation of Feasibility and Computational Complexity Results". In: *Optimization and Engineering* 21.1 (2020), pp. 305–334. DOI: `10.1007/s11081-019-09447-0`.

[33] M. Labbé, F. Plein, M. Schmidt, and J. Thürauf. „Deciding Feasibility of a Booking in the European Gas Market on a Cycle is in P". In: *Networks* (2021). Online first. DOI: `10.1002/net.22003`.

[34] M. Locatelli and F. Schoen. *Global optimization: theory, algorithms, and applications*. SIAM, 2013. DOI: `10.1137/1.9781611972672`.

[35] C. M. Macal and A. P. Hurter. „Dependence of bilevel mathematical programs on irrelevant constraints". In: *Computers & Operations Research* 24.12 (1997), pp. 1129–1140. DOI: `10.1016/S0305-0548(97)00025-7`.

[36] A. Mas-Colell, M. D. Whinston, J. R. Green, et al. *Microeconomic Theory*. Vol. 1. Oxford university press New York, 1995.

[37] A. Mitsos. „Global solution of nonlinear mixed-integer bilevel programs". In: *Journal of Global Optimization* 47.4 (2010), pp. 557–582. DOI: `10.1007/s10898-009-9479-y`.

[38] A. Mitsos, P. Lemonidis, and P. I. Barton. „Global solution of bilevel programs with a nonconvex inner program". In: *Journal of Global Optimization* 42.4 (2008), pp. 475–513. DOI: `10.1007/s10898-007-9260-z`.

[39] J. T. Moore and J. F. Bard. „The mixed integer linear bilevel programming problem". In: *Operations Research* 38.5 (1990), pp. 911–921. DOI: `10.1287/opre.38.5.911`.

[40] M. Patriksson and L. Wynter. „Stochastic mathematical programs with equilibrium constraints". In: *Operations Research Letters* 25.4 (1999), pp. 159–167. DOI: `https://doi.org/10.1016/S0167-6377(99)00052-8`.

[41] S. Pineda and J. M. Morales. „Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold". In: *IEEE Transactions on Power Systems* (2019). DOI: `10.1109/TPWRS.2019.2892607`.

[42] J. Pita, M. Jain, M. Tambe, F. Ordóñez, and S. Kraus. „Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition". In: *Artificial Intelligence* 174.15 (2010), pp. 1142–1171. DOI: `10.1016/j.artint.2010.07.002`.

[43] S. Pramanik and D. Banerjee. „Chance Constrained Quadratic Bi-level Programming Problem". In: *International Journal of Modern Engineering Research* 2.4 (2012), pp. 2417–2424. DOI: `10.1134/S19904789180400063`.

[44] A. Prékopa. *Stochastic Programming*. Kluwer, Dordrecht, 1995.

[45] Regionales Rechenzentrum Erlangen. *Woodcrest Cluster*. URL: `https://hpc.fau.de/systems-services/systems-documentation-instructions/clusters/woody-cluster/` (visited on 02/15/2021).

[46] L. Schewe, M. Schmidt, and J. Thürauf. *Global Optimization for the Multilevel European Gas Market System with Nonlinear Flow Models on Trees*. Tech. rep. 2020. URL: `http://www.optimization-online.org/DB_HTML/2020/08/7973.html`.

[47] A. Shapiro and H. Xu. „Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation". In: *Optimization* 57.3 (2008), pp. 395–418. DOI: `10.1080/02331930801954177`.

[48] S. Siddiqui and S. A. Gabriel. „An SOS1-Based Approach for Solving MPECs with a Natural Gas Market Application". In: *Networks and Spatial Economics* 13.2 (2013), pp. 205–227. DOI: `10.1007/s11067-012-9178-y`.

[49] J. Thürauf. *Deciding the Feasibility of a Booking in the European Gas Market is coNP-hard*. Tech. rep. 2020. URL: `http://www.optimization-online.org/DB_HTML/2020/05/7803.html`.

[50] L. Vicente, G. Savard, and J. Júdice. „Descent approaches for quadratic bilevel programming". In: *Journal of Optimization Theory and Applications* 81.2 (1994), pp. 379–399.

[51] T. R. Weymouth. „Problems in Natural Gas Engineering". In: *Trans. Amer. Soc. of Mech. Eng.* 34.1349 (1912), pp. 185–231.

[52] S. Wogrin, S. Pineda, and D. A. Tejada-Arango. „Applications of Bilevel Optimization in Energy and Electricity Markets". In: *Bilevel Optimization: Advances and Next Challenges.* Ed. by S. Dempe and A. Zemkoho. Cham: Springer International Publishing, 2020, pp. 139–168. DOI: `10.1007/978-3-030-52119-6_5`.

[53] P. Xu and L. Wang. „An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions". In: *Computers & Operations Research* 41 (2014), pp. 309–318. DOI: `10.1016/j.cor.2013.07.016`.

[54] J. Yang, M. Zhang, B. He, and C. Yang. „Bi-level programming model and hybrid genetic algorithm for flow interception problem with customer choice". In: *Computers & Mathematics with Applications* 57.11 (2009), pp. 1985–1994. DOI: `10.1016/j.camwa.2008.10.035`.

[55] I. Yanikoglu and D. Kuhn. „Decision Rule Bounds for Two-Stage Stochastic Bilevel Programs". In: *SIAM Journal on Optimization* 28.1 (2018), pp. 198–222. DOI: `10.1137/16M1098486`.

[56] İ. Yanıkoğlu, B. L. Gorissen, and D. den Hertog. „A survey of adjustable robust optimization". In: *European Journal of Operational Research* 277.3 (2019), pp. 799–813. DOI: `10.1016/j.ejor.2018.08.031`.