Chapter 14

Methods for verifying booked capacities

Benjamin Hiller, Christine Hayn, Holger Heitsch,
René Henrion, Hernan Leövey, Andris Möller,
Werner Römisch

Abstract *We formalize the problem to verify the legal requirement that transport situations arising from booked capacity rights shall be technically feasible. In particular, we propose a stochastic version of the problem of verifying booked capacities together with two heuristic solution methods. These methods have been designed as decision-support tools for real-world usage by transmission system operators (TSOs). Our approach is based on combining a stochastical model with an adversarial model to an overall model for the transport situations requested by the transport customers. The first method is based on sampling to capture the stochastic information, whereas the second method uses multivariate quantiles for that purpose. Both methods generate a set of nominations that are checked for technical feasibility to arrive at an overall conclusion.*

As described in Chapter 3, gas transmission system operators (TSOs) sell capacity rights to transport customers. Booking, i.e., buying, capacity rights entitles a transport customer to inject gas at entry points and/or withdraw gas at exit points of the gas network. In particular, TSOs are obliged to offer as much capacity as possible as freely allocable capacity (FAC), which enables transport customers to use entry and exit capacities independently (see Section 3.2.2 for details). However, a TSO may only sell capacity rights for which it can guarantee that each "likely and realistic" [GasNZV 2005, §9] load flow complying with the capacity rights booked by all transport customers can technically be realized. Thus a TSO needs a way to check this requirement.

This chapter presents methods that (heuristically) reduce this problem to checking a (potentially large) set of possible load flows for technical feasibility. In particular, we discuss how a suitable set of load flows may be obtained and how conclusions can be drawn from the corresponding feasibility tests. Methods for performing these feasibility tests are addressed in great detail in Chapters 6–10. The methods presented are, due to the high complexity of the real-world problem, heuristical in several aspects which are discussed in more detail in the final section of this chapter.

14.1 ▪ Motivation and outline of the approach

In order to formalize the problem of verifying that the load flows corresponding to a set of booked capacity contracts are technically feasible, we introduce some terminology (for types of capacity contracts, see Sections 3.2 and 3.3.2). Recall from Section 3.2.1 that using a gas transmission network is a two-step process: First, one has to book, i.e., buy, capacity contracts from the TSO. The day before the actual transmission is going to take place, one has to nominate the amount of gas that will be injected or withdrawn, according to the limitations of the capacity contract.

We define a *booking* B to be the set of all capacity contracts booked with the network operator at a certain point in time. We will see in Section 14.4 how such capacity contracts can be modeled in detail; for now it is sufficient that they prescribe certain conditions on load flows in the gas network. Let V_+ and V_- denote the set of entry and exit points of the gas network, respectively, and denote by $V_\pm$ the union of these two sets. We represent load flows in the gas network by load flow vectors. A *load flow vector* is a vector $P = (P_u)_{u \in V_\pm}$ that specifies, for each entry and exit, a load flow. Throughout this chapter, we will assume that load flows are specified in terms of power. As explained in Section 5.3.3, the load flows specified in power may be converted to ones specified as mass flows for checking their technical feasibility. We call a load flow vector *booking-compliant*, if it satisfies all conditions that are related to the capacity contracts in a booking B.

A *nomination* is a load flow vector $P = (P_u)_{u \in V_\pm}$ that is balanced, i.e., satisfies the condition

$$\sum_{u \in V_+} P_u = \sum_{u \in V_-} P_u.$$

Finally, we call a nomination *technically feasible*, if the gas network can be controlled such that the gas flow specified by the nomination is realized.

It is important to observe that this concept of nomination is an idealization of the process of nominating in two ways (see Section 3.2.1): First, load flows in a real gas network do not have to be balanced at any point in time, but only for longer balancing periods, for instance 24 hours. Second, strictly speaking load flows are only nominated at a subset of the points, e.g., at storages or the virtual trading point (VTP).

We introduce the following sets of load flow vectors:

$$\begin{aligned} \mathscr{B}_= &= \text{set of booking-compliant nominations}, \\ \mathscr{T} &= \text{set of technically feasible nominations}. \end{aligned}$$

Then, the task of the gas network operator can be formalized by asking whether the inclusion

$$\mathscr{B}_= \subseteq \mathscr{T} \tag{14.1}$$

holds, i.e., (14.1) means that any booking-compliant nomination should be technically feasible. This is the *deterministic* version of the *verification of booked capacities*.

Checking this inclusion in a mathematically exact sense seems to be hopeless in practice, since even checking a single booking-compliant nomination for technical feasibility requires solving a non-convex Mixed-Integer Nonlinear Program (MINLP), whose combinatorial part is already NP-hard (Szabó 2012). Under some additional assumptions, it would be sufficient to check the relation $P \in \mathscr{T}$ for a finite number of nominations P. For instance, given a polyhedral structure of the set $\mathscr{B}_=$ and assuming convexity of the set $\mathscr{T}$ (which does not hold true in general), the verification of (14.1) can be done by checking $P \in \mathscr{T}$ for the finitely many vertices of the polytope $\mathscr{B}_=$: If $P_1 \in \mathscr{T}$ and $P_2 \in \mathscr{T}$, so is the line $[P_1, P_2]$

by convexity of $\mathscr{T}$. Thus the fact that all vertices of $\mathscr{B}_=$ are in $\mathscr{T}$ implies that $\mathscr{B}_= \subseteq \mathscr{T}$. However, even for a moderate number of entry and exit points in the network, the number of vertices becomes astronomical. For instance, the polytope arising by intersecting the d-dimensional unit hypercube with one hyperplane given by the equation $\mathbb{1}^T x = k$, $k \in \mathbb{Z}$, has at least $\binom{d}{k}$ vertices if $2 \leq k \leq d-1$. In the simplest case, $\mathscr{B}_=$ might be given by a box $[\ell, u] \subseteq \mathbb{R}^d$ for $\ell, u \in \mathbb{R}^d$ and a balancing equation $\mathbb{1}^T P = c$ with $\|\ell\|_1 < c < \|u\|_1$, which is affinely equivalent to the polytope just mentioned. Thus one cannot benefit from finiteness in a computationally relevant sense.

Given the impossibility of checking (14.1) exactly, the question arises of how to find a finite subset of testing nominations $\{P^1, \dots, P^N\} \subset \mathscr{B}_=$ such that the relations $P^i \in \mathscr{T}$, for $i = 1, \dots, N$, provide a reliable substitute for the verification of (14.1). Of course, sampling of $\mathscr{B}_=$ should take into account historical information about nomination behavior. Moreover, as explained in Section 3.3.1, this is also required by current regulation rules. Depending on the considered gas network, the load flow at a subset of the points may be modeled stochastically. For instance, this is the case for exits belonging to public utilities where gas is usually used for heating and the load flow thus depends on the ambient temperature, which may be modeled stochastically as explained in Chapter 13. It is therefore justified to use probability distributions for the load flow estimated on the basis of historical data to predict future load flow patterns at these points with a certain probability. We collect the points for which this is reasonable in the set V_{stat} and call them *statistical points*. We assume in the following that a stochastical model for the load flows of the statistical points is available. More precisely, we assume there is a random vector $\xi : \Omega \to \mathbb{R}^{V_{\text{stat}}}$ defined on a suitable probability space $(\Omega, \mathscr{A}, \mathbb{P})$, e.g., derived by the data analysis from Chapter 13. We will call the random vector ξ of the loads at the statistical points *random load vector*. An element of $\mathbb{R}^{V_{\text{stat}}}$, and thus in particular a realization of ξ, is called *statistical load scenario*.

However, there are also points for which a stochastic model is not appropriate. For instance, the behavior of entries and storages is mainly market-driven, hence difficult to model in a stochastic way. This lopsided constellation suggests to consider a substitute for inclusion (14.1) which takes into account that a stochastic model for the statistical points is used. To be more precise, let π denote the projection of a load flow vector P onto the statistical points, i.e., $\pi(P) = (P_u)_{u \in V_{\text{stat}}}$. Moreover, let $P_{\text{stat}} \in \mathbb{R}^{V_{\text{stat}}}$ be any load flow vector at the statistical points. Then, (14.1) is equivalent to the partitioned inclusion

$$\pi^{-1}(P_{\text{stat}}) \cap \mathscr{B}_= \subseteq \mathscr{T} \quad \text{for all } P_{\text{stat}} \in \mathbb{R}^{V_{\text{stat}}}, \tag{14.2}$$

stating that all booking-compliant nominations extending any P_{stat} are technically feasible. This requirement, however, is much too restrictive and unrealistic, because the given inclusion always bears the risk to be violated by some extreme but very unlikely load flow vectors P. At this place, the exploitation of stochastic information makes sense. Taking into account the stochastic character of the random load vector ξ, we relax the strict "for all" relation (14.2) by a probabilistic condition of the form

$$\beta := \mathbb{P}(\pi^{-1}(\xi) \cap \mathscr{B}_= \subseteq \mathscr{T}) \geq \alpha, \tag{14.3}$$

requiring that (14.2) holds with a specified probability $\alpha \in (0,1)$ only. This is the *stochastic version of verification of booked capacities* on which we will focus for the remainder of this chapter. We call the probability β the *validity probability* of the booking B. With the terminology introduced above, Inequality (14.3) expresses the condition that with at least probability α, every booking-compliant nomination extending the random load vector ξ will be technically feasible. Of course, the choice of which points to treat as statistical points and the particular choice of a concrete probability level $\alpha \in (0,1)$ (typically close to one) has to be agreed upon, based on experience, governmental rules, or common sense.

In order to explain the proposed methods to numerically check condition (14.3), we introduce two functions $\Gamma \colon \mathbb{R}^{V}_{\text{stat}} \to \mathscr{P}(\mathbb{R}^{V_\pm})$ and $\nu_{\mathscr{T}} \colon \mathscr{P}(\mathbb{R}^{V_\pm}) \to \{0,1\}$, where $\mathscr{P}(\mathbb{R}^{V_\pm})$ denotes the power set of $\mathbb{R}^{V_\pm}$. The function Γ maps a statistical load scenario s to its set of booking-compliant extending nominations, i.e., we have $s \mapsto \pi^{-1}(s) \cap \mathscr{B}_=$, whereas $\nu_{\mathscr{T}}$ indicates whether *all* nominations of a set of nominations are technically feasible. Moreover, we define $\Upsilon_B := \nu_{\mathscr{T}} \circ \Gamma$. With this notation, we can rewrite condition (14.3) as

$$\beta = \mathbb{P}(\Upsilon_B(\xi) = 1) \geq \alpha. \tag{14.4}$$

Assume for the moment that we could evaluate $\Upsilon_B(s)$ for a given statistical load scenario s. One natural way to check condition (14.4) is to compute the probability on the left hand side. We can rewrite this probability as

$$\mathbb{P}(\Upsilon_B(\xi) = 1) = 1 \cdot \mathbb{P}(\Upsilon_B(\xi) = 1) + 0 \cdot \mathbb{P}(\Upsilon_B(\xi) = 0) = \mathbb{E}_{\mathbb{P}}[\Upsilon_B] = \int_\Omega \Upsilon_B \, \mathrm{d}\mathbb{P}.$$

Thus the computation of the probability can be seen as the numerical estimation of high-dimensional integrals. In Section 14.2 we use techniques like Quasi-Monte Carlo methods and scenario reduction to construct a discrete set of statistical load scenarios $\{s^1, \ldots, s^N\}$ with associated probabilities $p_1, \ldots, p_N$. This set is the basis of an estimator for the above integral which evaluates Υ_B once for each s^i, $1 \leq i \leq N$.

A conceptually different approach described in Section 14.3 relies on so-called multivariate quantiles to check condition (14.4) directly. The basic idea is to construct a special statistical load scenario $\hat{s}$ that satisfies

$$\mathbb{P}(\xi \leq \hat{s}) \geq \alpha,$$

i.e., that dominates ξ with high probability. Assuming a certain monotonicity condition, $\Upsilon_B(\hat{s}) = 1$ then implies $\mathbb{P}(\Upsilon_B(\xi) = 1) \geq \alpha$.

It remains to discuss how to evaluate $\Upsilon_B(s)$ for a given statistical load scenario s. Observe that this task can in principle be achieved by a procedure that either generates a booking-compliant nomination extending s that is *not* technically feasible (certifying $\Upsilon_B(s) = 0$) or establishes that no such nomination exists (certifying $\Upsilon_B(s) = 1$). Given the complexity of checking technical feasibility, it is unlikely to develop a practically efficient method for this purpose. Therefore, we propose an adversarial heuristic that tries to construct a small set of extending nominations that are challenging for the gas network, i.e., which are likely to be technically infeasible if the validity probability β of the booking is too low. This adversarial heuristic is based on a model for the set $\mathscr{B}_=$ of booking-compliant nominations, presented in Section 14.4. The heuristic itself is based on sampling again and takes into account expert knowledge from network planners. It is explained in Section 14.5.

Section 14.6 puts all the pieces together and describes the two resulting overall methods, one based on sampling, the other using multivariate quantiles to capture the stochasticity of the load flows at the statistical points. From a high-level viewpoint, both methods work as follows:

1. Generate a set of statistical load scenarios that carry the stochastic information and provide load flows at the statistical points.
2. Use the adversarial heuristic to compute a set of booking-compliant nominations extending each statistical load scenario.
3. Check these nominations for technical feasibility and draw an overall conclusion from these outcomes.

Apart from the adversarial heuristic, there are more heuristic aspects involved in these methods. These are discussed in Section 14.8, which also concludes the chapter.

14.2 ▪ Sampling statistical load scenarios for verifying booked capacities

As described in the previous section, the load flow of a substantial part of the points of the considered gas network can be modeled stochastically. For these statistical points, historical data is available and a carefully selected probability distribution model $\mathbb{P}$ for load flow can be calibrated. We will refer to the dimension d as the number of considered statistical points. As mentioned before, the considered set of statistical points is large, with values of dimension d that can range into several hundreds.

For a fixed booking B, one would like to know the validity probability β under the load flow distribution $\mathbb{P}$. The completion and validation (in the sense of checking the technical feasibility) of a statistical load scenario s^i is a process that can be interpreted as a (measurable) function $\Upsilon_B \colon \mathbb{R}^{V_{\text{stat}}} \to \mathbb{R}$. In the simplest case, Υ_B assigns to every generated statistical load scenario s^i, $1 \le i \le N$, the value of "1" in case of technical validity under the given booking B, or "0" otherwise. In a more general case, we can consider a validity function $\Upsilon_B^* \colon \mathbb{R}^{V_{\text{stat}}} \to [0,1]$, where now the validation process can return values between 0 and 1. In this generalized setting, the task is defined to be the estimation of $\mathbb{E}_{\mathbb{P}}\left[\Upsilon_B^*\right]$ in order to take a decision of accepting or rejecting a booking, which is a high-dimensional integration problem. In our case, and as it is usual in many practical high-dimensional problems in simulation, the statistical points have been modeled with distributions over $\Omega = \mathbb{R}^d$ that allow us to consider a bijective transformation $\Phi \colon \mathbb{R}^d \to (0,1)^d$ changing the original problem into an integration problem over the unit cube. The integration problem takes now the form

$$\mathbb{E}_{\mathbb{P}}[\Upsilon_B^*] = \int_{\mathbb{R}^d} \Upsilon_B^* \,\mathrm{d}\mathbb{P} = \int_{[0,1]^d} \Upsilon_B^*(\Phi^{-1}(x))\,\mathrm{d}x. \tag{14.5}$$

The latter equality is valid since an arbitrary extension of Φ^{-1} to the boundary of $[0,1]^d$ can be carried out, because the zero-measure boundary set does not influence the value of the resulting integral.

In the following we describe high-dimensional integration methods for approximating the desired expectation in (14.5) starting with Monte Carlo methods, moving to Quasi-Monte Carlo and finally to a hybrid method, namely, randomized Quasi-Monte Carlo. The choice of the transformation Φ is of essential importance for the problem of sampling in high-dimensions. It usually has a strong influence in what is called the *effective dimension* of the problem. We will discuss this issue in Section 14.2.1, where we also argue why the class of randomized Quasi-Monte Carlo methods is preferable for our application.

In the classical Monte-Carlo (MC) approach (see Niederreiter 1992) one tries to estimate (14.5) by generating statistical load scenarios pseudo-randomly. Starting with a finite sequence of independent identically distributed (i.i.d.) samples $S_N = \{s^1, s^2, \dots, s^N\}$, where the points s^i, $1 \le i \le N$, are uniformly distributed in $[0,1]^d$, the average of a given target function f

$$Q_N(f) := \frac{1}{N}\sum_{i=1}^{N} f(s^i),$$

is taken as an approximation of a desired integral $I_d(f) = \int_{[0,1]^d} f(x)\,\mathrm{d}x$. The resulting estimator $Q_N(f)$ is unbiased, and the error can be approximated via the central limit theorem, assuming that f is square-integrable. The variance of the estimator $Q_N(f)$ is given by

$$\mathrm{Var}[Q_N(f)] = \frac{\sigma^2(f)}{N} = \frac{1}{N}\left(\int_{[0,1]^d} f^2(x)\,\mathrm{d}x - \left(\int_{[0,1]^d} f(x)\,\mathrm{d}x\right)^2\right).$$

The resulting integration error associated to the MC approach is then of order $O(N^{-\frac{1}{2}})$. The quality of the MC samples relies on the selected pseudo-random number generators of uniform samples in $[0,1]^d$. Good accessible generators to this end are for example the *Mersenne Twister* from Matsumoto and Nishimura (1998) and *MRG32k3a* from L'Ecuyer (1999). MC is in general a very reliable tool in high-dimensional integration, but the order of convergence is in fact poor. Since validation of nominations in real-world gas network can be a very time consuming procedure, the search for good tools improving the accuracy of estimation, or reducing the amount of samples needed to reach a desired accuracy, is essential.

In contrast to MC methods, *Quasi-Monte Carlo (QMC) methods* are deterministic methods based on sequences of points that are more regularly distributed than the pseudo-random points from MC (see L'Ecuyer and Lemieux 2005; Novak and Woźniakowski 2010; Dick and Pillichshammer 2010; Kuo, Schwab, and Sloan 2011). Using QMC, one can expect in many practical situations with high dimensional integrands an error convergence of order $O(N^{-1})$, if the integrands are sufficiently smooth. Typical examples of QMC are modern shifted lattice rules and low-discrepancy sequences. To define what we mean by "regularly distributed", we now introduce the classical notion of discrepancy (see Niederreiter 1992) of a finite sequence of points S_N in $[0,1)^d$.

Definition 14.1. *Let $S = \{s^1, \dots, s^N\}$ be an arbitrary set of points in $[0,1)^d$. The* discrepancy *of S w.r.t. to an interval $[0,a) \subseteq [0,1)^d$ is measured by the function*

$$\mathrm{disc}(S,a) = \sum_{j=1}^{N} \chi_{[0,a)}(s^j) - \prod_{i=1}^{d} a_i,$$

where $\chi_{[0,a)}(\cdot)$ is the characteristic function of $[0,a)$. Let $D = \{1,\dots,d\}$ and define $\hat{x}_I$ for any $x \in [0,1]^d$ and $I \subseteq D$ by

$$\hat{x}_I = \begin{cases} x_i & i \in I, \\ 1 & \textit{otherwise}. \end{cases}$$

Then

$$\mathbb{D}_r(S) = \Big(\sum_{\emptyset \neq I \subseteq D} \int_{[0,1]^{|I|}} |\mathrm{disc}(S,\hat{x}_I)|^r \,\mathrm{d}x_I\Big)^{\frac{1}{r}}$$

is called L_r-discrepancy of the point set S, $r \in [1,\infty]$, with the obvious modification for $r = \infty$. The L_∞-discrepancy is also called star discrepancy and denoted by $\mathbb{D}^\star(S)$.

The star discrepancy gives a measure of the worst difference, for a given finite point set $S = \{s^1, \dots, s^N\}$, between the uniform distribution and the sampled distribution in $[0,1)^d$ attributed to the set S. In the context of QMC, a sequence of points in $[0,1)^d$ is called a low-discrepancy sequence if $\mathbb{D}^\star(S) = O(N^{-1}(\log(N))^d)$ for all truncations of the sequence to its first N terms.

The usual way to analyze QMC as a deterministic method is to choose a linear normed space F of functions on $[0,1)^d$ with norm $||\cdot||$ and an associated discrepancy $\mathbb{D}(S_N)$ for the point sequence S_N. Then, the deterministic integration error can be estimated by

$$|Q_N(f) - I_d(f)| \leq \mathbb{D}(S_N)||f||$$

for all functions $f \in F$. Such estimates are called Koksma-Hlawka type inequalities due to the classical Koksma-Hlawka inequality (see Niederreiter 1992), where $\mathbb{D}(S_N)$ is taken to be the star discrepancy of the point sequence S_N and $||f||$ is the variation in the sense of Hardy and Krause of f.

In modern QMC error analysis, one often considers weighted reproducing kernel Hilbert spaces (RKHS) as function spaces (see Kuo, Schwab, and Sloan 2011). In this context one obtains an error bound in the above form, where $\mathbb{D}(S_N)$ represents a weighted L_2-discrepancy. If the considered weights satisfy some particular decay conditions, describing a decay of importance of the variables or group of variables, then the discrepancy $\mathbb{D}(S_N)$ can be reduced at a rate $O(N^{-1+\delta})$, $\delta \in (0, \frac{1}{2}]$, with a constant δ independent of the dimension d, in a tractable way with specially constructed shifted lattice rules and low-discrepancy sequences (see Kuo, Schwab, and Sloan 2011).

In practice, randomly shifted lattice rules and low-discrepancy sequences are both competitive techniques of QMC. Our choice for generation of statistical load scenarios using *digital sequences*, namely, Sobol' sequences, is a special case of low-discrepancy sequences that are included in the category of (t, m, d)-nets and (t, d)-sequences (Dick and Pillichshammer 2010). In some sense, shifted lattice rules are adaptive in the way that they allow using information of the target integrand to fix the generating vector of the lattice, if the given integrand is smooth enough (Griewank et al. 2013). On the other hand, digital sequences focus more on other features as the discrepancy, thus they are constructed independently of the integrand at hand. In our case for the validation of statistical load scenarios, we do not know explicitly how the integrand looks like. We do know that in practice the validity function can be taken of the form $\Upsilon_B : \mathbb{R}^{V_\pm} \to \{0, 1\}$, presenting discontinuity jumps.

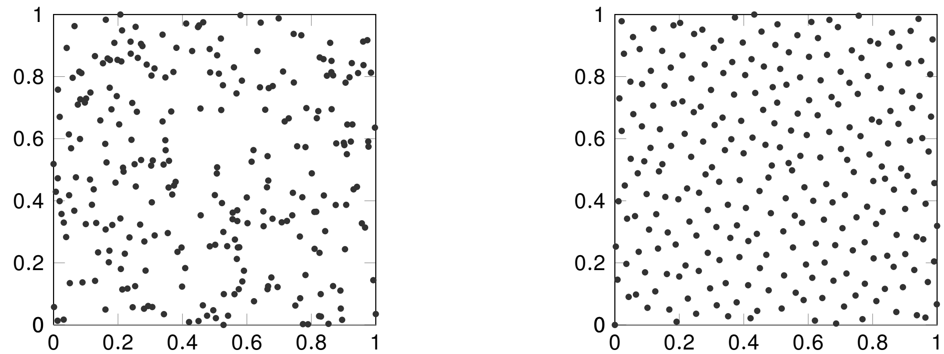
There are some practical advantages in retaining the probabilistic scheme of the sampling, while using these nice deterministic constructions called digital sequences. Therefore we have focused on hybrid methods permitting us to combine the best features of MC and QMC together. Randomization is an important tool in high-dimensional integration if we want to estimate the error of our approximation $Q_N(f)$ to the desired integral. One goal is to randomize the deterministic point set S_N generated by QMC in a way that the randomized points in the set $\tilde{S}_N$ have the uniform distribution over $[0, 1)^d$. Thus the resulting estimator $Q_N(f)$ preserves unbiasedness. The second goal is to preserve the better equidistribution properties of the deterministic construction. The simplest form of randomization applied to digital sequences seems to be the technique called *digital b-ary shifting*, see (L'Ecuyer and Lemieux 2005, Section 5.2) and the references therein.

We choose $b = 2$, i.e., we use random digital binary shifting to obtain our randomized QMC method that works as follows. To obtain a final point $\tilde{u} \in [0, 1]^d$, we generate a point $u \in [0, 1]^d$ from the underlying Sobol' sequence and a pseudo-random vector Δ uniformly distributed in $[0, 1)^d$. We then consider the binary expansions of each component of both u and Δ that are given by

$$\Delta_j = \sum_{l=1}^{\infty} \delta_{jl} 2^{-l} \quad \text{and} \quad u_j = \sum_{l=1}^{\infty} u_{jl} 2^{-l} \quad \text{for all } j = 1, \ldots, d.$$

The random digital binary shifted Sobol point $\tilde{u}$ is then computed by

$$\tilde{u}_j = \sum_{l=1}^{\infty} (u_{jl} + \delta_{jl}) 2^{-l} \quad \text{for all } j = 1, \ldots, d,$$



(a) $N = 256$ Monte Carlo Mersenne-Twister samples for $d = 500$, projection $(8, 9)$

(b) $N = 256$ RQMC random digital binary shifted Sobol' points for $d = 500$, projection $(8, 9)$

Figure 14.1: Comparison of Monte Carlo Mersenne-Twister and RQMC samples

where the addition is modulo 2. A comparison of MC samples generated by the Mersenne-Twister pseudo-random generator and random digital binary shifted Sobol' points is shown in Figure 14.1.

14.2.1 ▪ Justification of using randomized QMC methods

A partial explanation to the success of QMC against MC can be given by considering the ANOVA *(Analysis of Variance) decomposition* of the functions at hand. If the integrands (maybe after a proper transformation) have the property that few ANOVA terms corresponding to the interaction of few variables accumulate most of the variance of the integrand, and if these important ANOVA terms exhibit enough smoothness, then we can expect that QMC will perform better than MC for integration.

Using ANOVA we decompose a function into a sum of simpler functions (see Sobol' 2001). Let $D = \{1, \ldots, d\}$. For any subset $I \subseteq D$, let $|I|$ denote its cardinality and $D \setminus I$ be its complementary set in D. Let $x_I = (x_j)_{j \in I}$ be the $|I|$-dimensional vector containing the coordinates of x with indices in I. Now assume that f is a square integrable function. Then we can write f as the sum of 2^d ANOVA terms:

$$f(x) = \sum_{I \subseteq D} f^I(x),$$

where the ANOVA terms $f^I(x)$ are defined recursively by

$$f^I(x) = \int_{[0,1]^{d-|I|}} f(x_I, x_{D \setminus I}) \, \mathrm{d}x_{D \setminus I} - \sum_{I' \subsetneq I} f^{I'}(x),$$

and $f^{\emptyset} = I_d(f)$. The sum of the right-hand side is over strict subsets $I' \neq I$, and we use the convention $\int_{[0,1]^0} f(x) \, \mathrm{d}x_{\emptyset} = f(x)$. Note that the ANOVA decomposition is L_2-orthogonal.

In many practical applications, one encounters functions for which the total variance is concentrated in a small portion of its ANOVA terms. The notion of effective dimension of a function was first introduced by Caflisch, Morokoff, and Owen (1997) to describe the contribution of a group of variables to the total variance.

Definition 14.2. *A function f is said to have* effective dimension d_t *in the truncation sense with proportion p, for $0 < p < 1$, if d_t is the smallest integer that satisfies*

$$\sum_{I \subseteq \{1,\ldots,d_t\}} \sigma_I^2(f) \geq p\sigma^2(f),$$

where $\sigma_I^2(f)$ denotes the variance of the ANOVA *term $f^I(x)$.*

It is known that the lower order ANOVA terms of an integrand can exhibit substantially more smoothness than the integrand itself, even if the integrand presents discontinuity jumps (Griebel, Kuo, and Sloan 2010; Griebel, Kuo, and Sloan 2013; Heitsch, Leövey, and Römisch 2012). Therefore effective dimension reduction techniques based on suitable transformations of the integrand are essential. In Gaussian integration, the particular choice of matrix factorization usually has a strong influence in the effective dimension of the problem and in the performance of QMC. Principal components analysis (PCA) decomposition (Wang and Fang 2003) is usually recommended to be applied if feasible, and this is the method of choice for generating multivariate Gaussian samples for statistical load scenarios. It is well known that PCA can reduce the effective dimension and improve the performance of QMC for many integrands considered in mathematical finance (Glasserman 2004; Wang and Sloan 2005), and the same has been shown recently in two-stage linear stochastic optimization problems (Heitsch, Leövey, and Römisch 2012).

In most examples encountered in practical applications requiring moderate or small sample sizes N, one expects that randomized QMC will work at least as good as MC. Thus, there is usually no loss in replacing MC by particular good randomized versions of QMC. We can expect in many cases even a benefit using QMC if the given integrands can be well approximated by low dimensional smooth functions (Kuo, Schwab, and Sloan 2011), exhibiting in many cases order of convergence close to $O(N^{-1})$.

14.2.2 ▪ Scenario reduction

Scenario reduction may be desirable in many situations when the underlying scenario models already happen to be large scale and the incorporation of a large number of scenarios leads to high computation times. The basic idea of scenario reduction consists in determining a (nearly) best approximation in terms of a suitable probability metric of the underlying discrete probability distribution by a probability measure with smaller support. The metric should be associated to the mathematical model in a canonical way such that the model behaves stable with respect to changes of the probability distribution. Several canonical metrics are discussed by Römisch (2003).

Since the relevant optimization and feasibility problems of this book are mixed-integer nonlinear with stochastic inputs, the so-called discrepancies (see, for example, Römisch 2003) and, in particular, the Kolmogorov distance appear to be suitable probability metrics. We refer to the monograph by Rachev (1991) for a survey of probability metrics and their properties.

Let $\mathbb{P}$ denote a discrete probability distribution on $\mathbb{R}^d$ with statistical load scenarios s^i and probabilities p_i, $i = 1, \ldots, N$, and $\mathbb{P}_J$ a discrete probability distribution with scenarios s^j and probabilities p'_j, $j \notin J$. Hence, the support of $\mathbb{P}_J$ is a subset of the support of $\mathbb{P}$, and J denotes the index set of deleted scenarios from those of $\mathbb{P}$. The *Kolmogorov distance* between two probability distributions is defined as the uniform distance of their (cumulative) distribution functions. In our special case, we have

$$\mathbb{D}_K(\mathbb{P},\mathbb{P}_J) = \sup_{x\in\mathbb{R}^d} \left| \frac{1}{N} \sum_{i=1, s^i \le x}^{N} p_i - \sum_{j\notin J, s^j\le x} p'_j \right|.$$

However, the numerical results by Henrion, Küchler, and Römisch (2009) show that the problem of optimal scenario reduction with respect to the Kolmogorov distance can presently not be solved in reasonable time for higher dimensions d.

Therefore, we employ alternatively the L_1-Wasserstein or *Kantorovich distance* of $\mathbb{P}$ and $\mathbb{P}_J$ given by

$$W_1(\mathbb{P},\mathbb{P}_J) = \min \left\{ \sum_{i=1}^{N} \sum_{j\notin J} \eta_{ij} \left\| s^i - s^j \right\|_1 \,\middle|\, \eta_{ij} \ge 0, \sum_{i=1}^{N} \eta_{ij} = p'_j, \sum_{j\notin J} \eta_{ij} = p_i \right\}.$$

Clearly, computing $W_1(\mathbb{P},\mathbb{P}_J)$ means solving a linear program of dimension nN. The problem of optimal scenario reduction then consists in determining the best approximation of $\mathbb{P}$ by a distribution $\mathbb{P}_J$ with $J \subset \{1,\dots,N\}$ and $|J| = N-n$ for given $n < N$, i.e., it may be written as

$$\min\left\{ W_1(\mathbb{P},\mathbb{P}_J) \,\middle|\, J \subset \{1,\dots,N\},\ |J| = N-n,\ p'_j \ge 0 \text{ for } j\notin J,\ \sum_{j\notin J} p'_j = 1 \right\}$$

or

$$\min\left\{ D_J \,\middle|\, J \subset \{1,\dots,N\}, |J| = N-n \right\}, \tag{14.6}$$

where D_J denotes the minimum of $W_1(\mathbb{P},\mathbb{P}_J)$ with respect to $p'_j \ge 0$, $j \notin J$, and $\sum_{j\notin J} p'_j = 1$. The minimum may be computed as (see Dupačová, Gröwe-Kuska, and Römisch 2003):

$$D_J = \sum_{j\in J} p_j \min_{i\notin J} \left\| s^i - s^j \right\|,$$

and the corresponding optimal weights p'_j, $j \notin J$, are given by the (optimal) redistribution rule

$$p'_j = p_j + \sum_{i\in I(j)} p_i \quad \text{for all } j \notin J,$$

where

$$I(j) = \arg\min_{i\in J} \left\| s^i - s^j \right\|.$$

The combinatorial optimization problem (14.6) is called n-median problem in the literature and is known to be NP-hard (Kariv and Hakimi 1979).

There are several approaches for the computational solution of n-median problems. First, we mention the reformulation of (14.6) as mixed-integer linear program and the possibility of applying standard software (e.g., Cplex). For example, if $y_i \in \{0,1\}$, $i = 1,\dots,N$, denotes the decision variable whether s^i is deleted ($y_i = 0$) or not ($y_i = 1$), Problem (14.6) allows the reformulation

$$
\begin{aligned}
\min \quad & \sum_{i,j=1}^{N} p_j x_{ij} \left\| s^i - s^j \right\|_1 \\
\text{s.t.} \quad & \sum_{i=1}^{N} y_i = n, \\
& \sum_{1\le j\le N,\, j\ne i} x_{ij} + y_i = 1 && \text{for all } 1 \le i \le N, \\
& x_{ij} \le y_i && \text{for all } 1 \le i,j \le N, \\
& x_{ij} \in [0,1] && \text{for all } 1 \le i,j \le N, \\
& y_i \in \{0,1\} && \text{for all } 1 \le i \le N,
\end{aligned}
$$

as mixed-integer linear program with N^2 continuous and N binary variables. Indeed, since $y_i = 1$ for $i \notin J$, and, hence, $x_{ij} = 0$ for all $i \in J$ and $j \notin J$, one obtains

$$
\frac{1}{n} \sum_{i,j=1}^{N} p_j x_{ij} \left\| s^i - s^j \right\|_1 = \frac{1}{n} \sum_{i\notin J} \sum_{j\in J} p_j \left\| s^i - s^j \right\|_1 \ge \sum_{j\in J} p_j \min_{i\notin J} \left\| s^i - s^j \right\|_1 = D_J,
$$

and the lower bound is attained for

$$
x_{ij} = \frac{\min_{i\notin J} \|s^i - s^j\|_1}{n\,\|s^i - s^j\|_1} \quad \text{for all } i \notin J, j \in J.
$$

Alternatively, we mention the method based on column generation and on branch-cut-and-price algorithm in (Avella, Sassano, and Vasil'ev 2007), which is suitable for large-scale models, approximation methods based on semidefinite programming (Peng and Wei 2007), and a hybrid heuristic (Resende and Werneck 2004) including randomized greedy heuristics.

A forward greedy heuristic was first studied by Cornuéjols, Fisher, and Nemhauser (1977) among other exact and approximate methods. The computational experience in (Heitsch and Römisch 2003; Heitsch and Römisch 2007) suggests that simple forward and backward greedy heuristics with final optimal redistribution lead to good results in many situations. We thus employ such heuristics to solve the n-median problem. Figure 14.2 provides an illustrative example of the scenario reduction approach applied to a temperature depending gas flow at some typical exit point.

Finally, we mention that the scenario reduction approach is extended by Heitsch and Römisch (2007) to Kantorovich-Rubinstein type metrics (e.g., Fortet-Mourier metrics) and by Henrion, Küchler, and Römisch (2008) to rectangular and polyhedral discrepancies.

14.3 ▪ Generating quantiles for verifying booked capacities

In the previous section, scenario-based approximations of a given probability measure have been discussed. We now want to complement this approach by another possibility of characterizing multivariate distributions, namely the generation of quantiles. To this aim, let ξ be a d-dimensional random vector. To know the distribution of ξ means to know the probabilities $\mathbb{P}(\xi \in A)$ for all Borel-measurable subsets $A \subseteq \mathbb{R}^d$. Fortunately, all these probabilities of possibly complicated sets can be recovered from probabilities $\mathbb{P}(\xi \in z + \mathbb{R}^d_{\le 0}) = \mathbb{P}(\xi \le z)$ of relatively simple sets, so-called cells, which are the negative

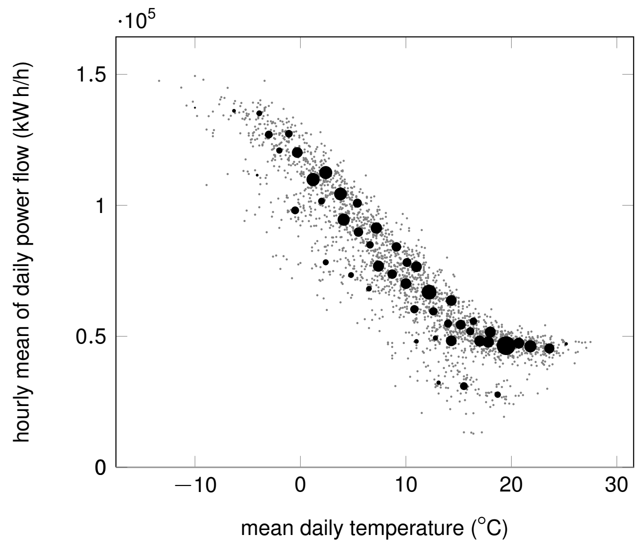


Figure 14.2: Illustration of optimal scenario reduction from $N = 2340$ temperature depending gas load scenarios with identical probability $\frac{1}{N}$ to $n = 50$. The new probabilities after redistribution are proportional to the diameters of the points representing the remaining scenarios.

orthants attached to arbitrary points $z \in \mathbb{R}^d$. The cumulative distribution function associated with ξ is defined as

$$F_\xi(z) := \mathbb{P}(\xi \le z).$$

Therefore, F_ξ carries the whole information about the distribution of ξ. In the one-dimensional case, $d = 1$, one defines a *(univariate) p-quantile*, $p \in [0, 1]$, of the random variable ξ as the quantity

$$q_p(\xi) := \inf\{t \mid F_\xi(t) \ge p\},$$

which can be understood as the inverse of the distribution function. The benefit of univariate quantiles consists in the equivalence

$$F_\xi(z) \ge p \iff z \ge q_p(\xi), \tag{14.7}$$

which allows to transform a probabilistic inequality into an explicit inequality. Note that univariate quantiles are easily calculated for all prominent distributions by standard software.

A multidimensional analogue of this concept is the so-called *multivariate p-quantile* (Prékopa 2012). It is defined for a d-dimensional random vector ξ as the set

$$Q_p(\xi) := \{z \in \mathbb{R}^d \mid F_\xi(z) \ge p \text{ and } F_\xi(y) \ge p,\, y \le z \text{ imply } y = z\}, \tag{14.8}$$

which is easily seen to reduce in the one-dimensional case, $d = 1$, to the classical quantile $q_p(\xi)$. If ξ has a density which is positive everywhere (as the Gaussian one), then the p-quantile is just the p-level set of the distribution function F_ξ, i.e.,

$$Q_p(\xi) = \{z \mid F_\xi(z) = p\}.$$

In contrast to the univariate case, multivariate quantiles are no longer singletons but sets (typically curved hypersurfaces in $\mathbb{R}^d$). The generalization of (14.7) now reads as follows:

$$F_\xi(z) \geq p \iff \exists q \in Q_p(\xi)\colon z \geq q. \tag{14.9}$$

Let us illustrate the use of multivariate quantiles in the context of the probabilistic inequality (14.3) which is central in the present chapter. We make the following *monotonicity assumption* for the feasibility of statistical load scenarios $P_{\text{stat}}, P'_{\text{stat}} \in \mathbb{R}^{V_\pm}$ (for notation see Section 14.1):

$$P'_{\text{stat}} \leq P_{\text{stat}} \text{ and } \Upsilon_B(P_{\text{stat}}) = 1 \implies \Upsilon_B(P'_{\text{stat}}) = 1, \tag{14.10}$$

i.e., if a statistical load scenario is technically feasible in the sense of (14.2), then the same should hold true for all statistical load scenarios which are at most as large in each component. We recall that technical feasibility of P_{stat} means that all booking-compliant nominations P extending P_{stat} are technically feasible. Under the simplifying assumption (14.10), one immediately infers that the probabilistic inequality (14.3) is satisfied whenever we find a quantile $q \in Q_p(\xi)$ with $\Upsilon_B(q) = 1$. Indeed, this last relation along with (14.9) and (14.10) implies that

$$\mathbb{P}(\Upsilon_B(\xi) = 1) \geq \mathbb{P}(\Upsilon_B(\xi) = 1, \xi \leq q) = \mathbb{P}(\xi \leq q) = F_\xi(q) \geq p.$$

The generation of such a quantile can be realized by employing codes for evaluating multivariate distribution functions, see, e.g., (Genz and Bretz 2009) for the examples of Gaussian and t-distributions.

Since there exists a continuous set of multivariate quantiles, the technical infeasibility of one of them does not exclude the technical feasibility of a different one. Hence, we can generate another quantile if necessary. With respect to the probabilistic inequality (14.3), this quantile-based approach has the advantage of possibly requiring only a few statistical load scenarios to be validated in contrast to a typically large number of statistical load scenarios for the sampling method of the preceding section. On the other hand, it relies on the monotonicity assumption (14.10) which is not strictly satisfied in reality and can distort the true value of the probability to be determined in (14.3).

14.4 ▪ Modeling capacity contracts

As discussed in Chapter 3, there is a sophisticated regulatory framework governing the use of gas transmission networks. Consequently, there are many different kinds of capacity contracts (see Table 3.1) and related conditions that are relevant for capacity planning in gas transmission networks. In this section we describe the typical data constituting a capacity contract and introduce a mathematical model for them. We note that apart from the capacity contracts described here, there are further types of contracts relevant for planning and operating a gas network. In particular, there are contractual limits for the pressures at entries and exits, which we will briefly discuss at the end of Section 14.5.

The model presented below not only covers capacity products that are sold to transport customers, but allows to incorporate agreements between TSOs regarding the interconnection capacities between different networks or market areas. In principle, these agreements correspond to transmission capacities but they feature more complex conditions like alternative capacities that may not be used at the same time. To capture these conditions, we introduce binary variables and thus arrive at a Mixed-Integer Linear Program (MILP) instead of an Linear Program (LP) model, as suggested in the introduction. The implications of this choice for the overall method are discussed in Section 14.8. We will briefly elaborate on the modeling power of our capacity contract model at the end of this section.

A *capacity contract* defines the capacity rights of a gas transport customer, i.e., the minimum and maximum amount of gas to supply or withdraw, including additional terms and conditions. Recall that we call the set of all capacity contracts booked at a certain point in time a booking, denoted B, and that we call a nomination booking-compliant, if it satisfies all conditions that are related to the capacity contracts. In our model, a capacity contract $\mathfrak{c}$ consists of one or more *capacity positions*, which is either a *freely allocable capacity (FAC) position* or a *restrictively allocable capacity (RAC) position*. The difference between FAC positions and RAC positions is in the balancing requirement: The total nomination of all entry FAC positions together has to match the total nomination of all exit FAC position (see the discussion of market areas in Section 3.2.2). In contrast, the nominations on entry and exit RAC positions have to match *within the same contract*, which usually limits nominations to a few entry and exit points. We denote by $\mathfrak{C}^{\mathrm{FAC}}$ and $\mathfrak{C}^{\mathrm{RAC}}$ the sets of contracts containing FAC positions and RAC positions, respectively. Moreover, let C^{FAC} be the set of all FAC positions and C^{RAC} the set of all RAC positions.

Each capacity position c defines the capacity rights at a set of points $V(c)$ in the gas network. We require that all points in $V(c)$ are of the same type, i.e., either all entries or all exits. In most cases, $V(c)$ is just a single point, i.e., a single entry or exit. Putting more points in $V(c)$ allows to realize zoning (see Section 3.2.2), i.e., defining a common capacity for a group of points that have to be treated as a single (virtual) point w.r.t. these capacity rights. A capacity position is usually valid for a certain time interval and possibly a restricted temperature range, since often the amount of capacity required depends on the temperature as defined in the contract. We implicitly assume that all capacity positions in B are valid for a common temperature T and date and thus define a booking situation that may occur at a single day.

The capacity of a capacity position c is given by the parameters $\underline{x}^c$ and $\overline{x}^c$, $0 \le \underline{x}^c \le \overline{x}^c$, which give the minimum and maximum power, respectively, that may be nominated on this capacity position. To model capacity contracts that specify alternative capacities, there is, for each capacity position c, a set $C^{\mathrm{cap}}_{\neg}(c)$ of capacity positions that are "incompatible" with c, i.e., it is not allowed to nominate on c and another capacity position from $C^{\mathrm{cap}}_{\neg}(c)$ at the same time. $C^{\mathrm{cap}}_{\neg}(c)$ may contain capacity positions from any contract, not just the one to which c belongs.

We now describe a mixed-integer linear model for the set of booking-compliant nominations for a given booking B. Of course, the first condition is that the load flow vector $(P_u)_{u\in V_\pm}$ has to be a nomination, i.e., balanced:

$$\sum_{u\in V_+} P_u = \sum_{u\in V_-} P_u. \tag{14.11}$$

For each capacity position c and each point $u \in V(c)$, we introduce a variable $P_u^c \ge 0$ for the power nominated on c. Denoting by $C^{\mathrm{cap}}(u)$ the set of capacity positions c at point u, the total power nominated at u is then given by

$$P_u = \sum_{c\in C^{\mathrm{cap}}(u)} P_u^c \quad \text{for all } u \in V. \tag{14.12}$$

We set $P_u = 0$ if there is no capacity position for point u. Moreover, we have a binary variable $x^c \in \{0,1\}$ for each capacity position c indicating whether this capacity position is used ($x^c = 1$) or not ($x^c = 0$).

For each capacity position $c \in C^{\text{FAC}} \cup C^{\text{RAC}}$, the model for booking-compliant nominations has to ensure that the load flows nominated at the points $V(c)$ are within the booked capacity limits $\underline{x}^c$ and $\overline{x}^c$, and that there is no load flow if the capacity position is not used. This is achieved by the constraints

$$\underline{x}^c x^c \leq \sum_{u \in V(c)} P_u^c \leq \overline{x}^c x^c \quad \text{for all } c \in C^{\text{FAC}} \cup C^{\text{RAC}}. \tag{14.13}$$

In the common case that $V(c) = \{u\}$, we have that P_u^c is in the interval $[\underline{x}^c, \overline{x}^c]$ if the capacity position c is used. In the case of zoning, i.e., if $|V(c)| > 1$, power may be nominated at any point in $V(c)$ as long as the sum is within the given capacity limits. Moreover, we have to ensure that capacity position $c \in C^{\text{FAC}} \cup C^{\text{RAC}}$ and each incompatible capacity position $\bar{c} \in C_{\neg}^{\text{cap}}(c)$ are not used simultaneously:

$$x^c + x^{\bar{c}} \leq 1 \quad \text{for all } c \in C^{\text{FAC}} \cup C^{\text{RAC}},\ \bar{c} \in C_{\neg}^{\text{cap}}(c). \tag{14.14}$$

In contrast to FAC positions, RAC positions within contract $\mathfrak{c}$ are not independent of each other, but need to be balanced. Let $\mathfrak{C}^{\text{RAC}} \subseteq B$ be the subset of contracts with restrictively allocable capacity (RAC) positions and consider a contract $\mathfrak{c} \in \mathfrak{C}^{\text{RAC}}$, denoting by $C^{\text{RAC}}(\mathfrak{c})$ the set of all RAC positions in contract $\mathfrak{c}$. Balancing is then ensured by the condition

$$\sum_{c \in C^{\text{RAC}}(\mathfrak{c})} \sum_{u \in V(c) \cap V_+} P_u^c = \sum_{c \in C^{\text{RAC}}(\mathfrak{c})} \sum_{u \in V(c) \cap V_-} P_u^c \quad \text{for all } \mathfrak{c} \in \mathfrak{C}^{\text{RAC}}. \tag{14.15}$$

This model is already quite powerful. It allows to model standard FAC and RAC products (see Table 3.1), which may be either defined for single points or entire entry and exit zones. Moreover, conditional versions of these products are also covered, as long as the conditions refer to temperature only, which is the usual case. This property is due to the fact that we verify booked capacities assuming a small temperature range (see Section 14.6 for details). We can thus evaluate temperature-dependent conditions and use the variant of the capacity contract that applies. The same reasoning applies to capacity contracts that are valid for a limited period only. It is also possible to model more complex conditions arising from interconnection agreements between different TSOs. These are typically based on the condition that certain hybrid points are used as an entry or as an exit point. Using the exclusion mechanism, one can ensure that either the capacities applying for entry usage or the ones for exit usage are used for nomination. So far, the model assumes that all capacities are firm capacities. It is possible to include interruptible capacities as well. However, deciding whether or not nominations should be interrupted has to be done during the nomination validation step since this is an operational measure. As the nomination validation methods presented in this book do not yet support this, we focus on firm capacities only.

14.5 ▪ An adversarial heuristic for generating booking-compliant nominations from statistical load scenarios

Assume that we obtained a statistical load scenario $s = (s_u)_{u \in V_{\text{stat}}}$ and we now want to check whether all extending booking-compliant nominations of s are technically feasible (i.e., evaluate $\Upsilon_B(s)$ in the notation of Section 14.1). We propose the following adversarial heuristic to construct a small set of extending nominations that are challenging for the gas network, i.e., which are likely to be a subset of the technically infeasible extending nominations of s. This approach is similar in spirit to the one explained in Section 4.2.

First we need a means to specify which parts of a nomination are derived from the statistical load scenario s and which are provided by the adversarial model. To this end, we distinguish *substitutable capacity positions* and *non-substitutable capacity positions*. The idea is that load flows for the substitutable capacity positions are "substituted" from s, whereas load flows for the non-substitutable capacity positions are determined adversarially. Thus capacities that are assumed to be used in the future as they were used in the past may be modeled using substitutable capacity positions and the remaining capacities as non-substitutable capacity positions.

It is usually not possible to generate the statistical load scenario s such that it necessarily complies with the booked capacity contracts. For instance, the support of the multivariate normal distribution is unbounded. Hence, it may happen that samples exceeding the available capacities or even negative samples are generated. To deal with this issue, we adjust s in a first step such that booking-compliant extending nominations exist, obtaining the *adjusted statistical load scenario* s'. Note that this adjustment in general affects the stochastic properties of s', i.e., s' may no longer be a multivariate quantile with the same p-value or its probability is different from that of s. We therefore choose s' as close as possible to s. Finally, we select some of the nominations extending s'.

We refine the model for booking-compliant nominations from Section 14.4 as follows. For every point $u \in V_{\text{stat}}$, we divide the set of capacity positions $C^{\text{cap}}(u)$ into the set of substitutable capacity positions $C^{\text{cap}}_{\text{s}}(u)$ and the set $C^{\text{cap}}_{\text{ns}}(u)$ of non-substitutable capacity positions. Equation (14.12) is then replaced, for any point $u \in V_{\text{stat}}$, by

$$P_u = \sum_{c \in C^{\text{cap}}_{\text{s}}(u)} P^c_u + \sum_{c \in C^{\text{cap}}_{\text{ns}}(u)} P^c_u, \tag{14.16}$$

i.e., we now distinguish between substitutable and non-substitutable capacity positions. Ideally, assuming that $C^{\text{cap}}_{\text{s}}(u)$ is the set of substitutable capacity positions at a point $u \in V_{\text{stat}}$, we would like to have the power nominated on $C^{\text{cap}}_{\text{s}}(u)$ to be given by s_u, i.e., $\sum_{c \in C^{\text{cap}}_{\text{s}}(u)} P^c_u = s_u$. To cover the potential need for adjustment, we introduce a slack variable $\Delta_u \in \mathbb{R}$ for any point $u \in V_{\text{stat}}$. The capacity constraints (14.13) together with the constraint

$$\sum_{c \in C^{\text{cap}}_{\text{s}}(u)} P^c_u = s_u + \Delta_u \quad \text{for all } u \in V_{\text{stat}}, \tag{14.17}$$

ensure that we obtain booking-compliant values for the substitutable capacity positions $C^{\text{cap}}_{\text{s}}(u)$.

We determine values for the Δ_u-variables whose sum of absolute values is as small as possible, i.e., that change the original sample vector s the least, using the following MILP:

$$\begin{aligned} \min \quad & \sum_{u \in V_{\text{stat}}} \Delta'_u && \qquad (14.18)\\ \text{s.t. } & (14.11), (14.16), (14.13), (14.14), (14.15), (14.17), \\ & \Delta'_u \geq \Delta_u && \text{for all } u \in V_{\pm}, \\ & \Delta'_u \geq -\Delta_u && \text{for all } u \in V_{\pm}, \\ & P_u, P^c_u, \Delta_u \geq 0 && \text{for all } u \in V_{\pm}, \\ & x^c \in \{0,1\} && \text{for all } c \in C^{\text{FAC}} \cup C^{\text{RAC}}. \end{aligned}$$

For each sampled statistical load scenario $s = (s_u)_{u \in V_{\text{stat}}}$, we solve the (easy) MILP (14.18) and use its optimal solution to construct the adjusted statistical load scenario $s' = (s'_u)_{u \in V_{\text{stat}}}$ defined by

$$s'_u := s_u + \Delta_u.$$

Note that we do not restrict the values of the variables at non-statistical points – these are just computed to ensure that there is a feasible extension of the adjusted statistical load scenario.

Having obtained the adjusted statistical load scenario s', we now need to choose booking-compliant extending nominations. The goal is to construct such extending nominations that are likely to be a subset of the technically infeasible extending nominations of s'. Usually, more extreme nominations, i.e., high-load and low-load nominations, but also ones with high regional imbalances, are more critical w.r.t. technical feasibility than less extreme ones. To obtain such nominations, we construct nominations that are extreme points of the set of booking-compliant nominations that extend s', which we just formulated as a MILP. To this end, we choose a random direction in the space of all nominations and compute a solution of the extension MILP that maximizes the value in this direction.

Network planners often have some expert knowledge about the gas network (see Section 4.2.3), for instance about points which are "equivalent" from a gas network point of view in the following sense: Given a total load flow for a set of points V', it does not matter (much) how this total load flow is distributed among the single points in V' (e.g., since the points are rather close). Thus nominations with similar total load flow at a set of equivalent points should be considered similar as well, despite differently distributing the total load flow among the nodes. Moreover, when constructing a challenging nomination it does not make sense to use an entry and an exit that are close to each other at the same time. Hence, either the entry or the exit should be used. To model both aspects, we assume that the points V are partitioned into subsets $V^1, \dots, V^r$ of "equivalent" points. This allows us to focus on the vector $\bar{P}$ defined by

$$\bar{P} := \left(\sum\nolimits_{u \in V^i \cap V_+} P_u - \sum\nolimits_{u \in V^i \cap V_-} P_u\right)_{1 \le i \le r},$$

that captures equivalence of the points by considering the net injections (which may be negative) of each point set V^i. Indeed, we will determine the final nomination P such that the vector $\bar{P}$ is extreme in the sense explained above, i.e., equivalent entries and exits will never be used at the same time if this is not enforced by some capacity positions.

In general we do not know which nominations are likely to be technically infeasible. We therefore try to select a set of nominations that are rather separate. To this end, we choose the direction θ for determining each extending nomination uniformly at random from the r-dimensional hypersphere. Since we generate just a few random directions, we use the same randomized QMC techniques to sample the direction that are also used for sampling statistical load scenarios. Again, the rationale is that QMC methods provide a better approximation to the uniform distribution than pseudo-random MC methods. We use the direction θ to determine a single extending booking-compliant nomination, i.e., an element of $\pi^{-1}(s') \cap \mathscr{B}_=$, that maximizes the value of that nomination in this direction. This is done by solving the following MILP that is similar to (14.18):

$$\max \sum_{1 \le i \le r} \theta_i \left(\sum_{u \in V^i \cap V_+} P_u - \sum_{u \in V^i \cap V_-} P_u \right) \tag{14.19a}$$

$$\text{s.t.} \quad (14.11), (14.16), (14.13), (14.14), (14.15), \tag{14.19b}$$

$$\sum_{c \in C_s^{\text{cap}}(u)} P_u^c = s'_u \quad \text{for all } u \in V_{\text{stat}}, \tag{14.19c}$$

$$P_u, P_u^c, \Delta_u \ge 0 \quad \text{for all } u \in V_\pm, \tag{14.19d}$$

$$x^c \in \{0,1\} \quad \text{for all } c \in C^{\text{FAC}} \cup C^{\text{RAC}}. \tag{14.19e}$$

Note that we use the constraint (14.19c) instead of (14.17), i.e., we require that the feasible solutions are booking-compliant nominations extending s'. Observe further that this construction does *not* provide a uniform sampling of the extreme points of the feasible set of the MILP.

To obtain complete inputs for the nomination validation methods, we provide the pressure limits corresponding to each nomination as well. These pressure limits depend on the actual use of the network, i.e., the load flows. However, once a nomination is determined, it is clear which entries and exits are used and the corresponding pressure limits can be taken into account when checking technical feasibility.

14.6 ▪ Methods to verify booked capacities

We now described all ingredients of the two methods outlined in Section 14.1 to check Inequality (14.3), i.e., to verify that the validity probability β of a given booking B is at least α. In the following, we present our overall approach to this problem. As discussed in Chapter 13, we construct a stochastic model for the load flows at the statistical points for each temperature class in order to deal with the temperature dependency of the load flows.

The two methods we propose differ in how they incorporate the stochastic information:

1. The first method uses sampled statistical load scenarios to represent the stochastic nature of exit loads (see Section 14.2).
2. The second method uses multivariate quantiles to represent the stochastic nature of exit loads (see Section 14.3).

Formally, both methods require convexity of the set $\mathcal{T}$ of technically feasible nominations to be valid. In addition, the quantile-based method is formally only valid if the monotonicity assumption (14.10) is fulfilled. However, it may require validating significantly fewer nominations to certify that the validity probability β is at least α. In contrast, the sampling-based method does not rely on the monotonicity assumption, but requires many statistical load scenarios and extending nominations, resulting in considerable computational effort for validating those nominations.

So far we assumed that checking a nomination for technical feasibility either establishes feasibility or proves infeasibility. In practice, however, it may also happen that there is no conclusive answer, e.g., since the time limit for the computation has been reached before technical feasibility could be decided. In the description of our methods we assume that each check for technical feasibility yields one of the answers "feasible", "infeasible", or "unknown".

In both methods we need to check whether all extending booking-compliant nominations of s are technically feasible (i.e., evaluate $\Upsilon_B(s)$ in the notation of Section 14.1); we do this by considering a set of n booking-compliant nominations extending s generated by the adversarial heuristic explained in Section 14.5). We consider s to be technically feasible (i.e., $\Upsilon_B(s) = 1$), if all n extending nominations of s are "feasible". In the case that at least one extending nomination is "infeasible", s is infeasible as well (i.e., $\Upsilon_B(s) = 0$). The remaining case is that some extending nominations are "feasible" and the rest is "unknown", so the feasibility of s is unknown, too.

14.6.1 ▪ Sampling-based verification of booked capacities

This method estimates the validity probability β of the given booking B using sampled statistical load scenarios. Since the feasibility of a statistical load scenario may be unknown, we actually compute two estimates $\hat{\beta}$ and $\hat{\beta}^*$ for the validity probability β: $\hat{\beta}$ is a pessimistic estimate, assuming that all of the statistical load scenarios whose feasibility is unknown are infeasible. In contrast, $\hat{\beta}^*$ is optimistic since it assumes that all of the statistical load scenarios with unknown feasibility are in fact feasible.

Using sampling to approximately check the probability requirement (14.3), the overall procedure to verify a booking is as follows. Note that we decided to base the final decision for or against validity of the booking on the pessimistic estimate $\hat{\beta}$.

A. For each temperature class perform the following steps:

1. Based on the stochastic model for this temperature class, sample a set of statistical load scenarios $s^1, \dots, s^M$ using the randomized QMC method described in Section 14.2. Each of these statistical load scenarios has probability $1/M$.
2. Use the scenario reduction technique outlined in Section 14.2.2 to find a smaller representative subset $s^1, \dots, s^N$ of the statistical load scenarios ($N \ll M$) together with updated probabilities $p_1, \dots, p_N$.
3. Set $I = I^* = \emptyset$.
4. For each statistical load scenario $s = s^i$, $1 \le i \le N$:
   a) Compute an adjusted statistical load scenario s' for which at least one booking-compliant nomination extending s' exists (see Section 14.5).
   b) Generate a set $P^1, \dots, P^n$ of booking-compliant nominations extending s' (see Section 14.5).
   c) Check the technical feasibility of $P^1, \dots, P^n$. If all of them are "feasible", add i to the index set I. If none of them is "infeasible", add i to the index set I^*.
5. Compute the estimates $\hat{\beta}$ and $\hat{\beta}^*$ for the validity probability β as

$$\hat{\beta} = \sum_{i \in I} p_i, \qquad \hat{\beta}^* = \sum_{i \in I^*} p_i.$$

B. The overall booking is considered feasible if we have $\hat{\beta} \ge \alpha$ for all temperature classes.

14.6.2 ▪ Quantile-based verification of booked capacities

This method uses multivariate quantiles to ensure that the validity probability β is at least α. To be correct, it requires the monotonicity assumption (14.10) to hold. Again, it may happen that the feasibility of a statistical load scenario (now a quantile) is unknown.

The overall procedure for quantile-based verification of the probability requirement (14.3) is as follows. To arrive at an overall conclusion in case the feasibility of all considered quantiles is unknown, we take the pessimistic view again and decide for invalidity then.

A. For each temperature class perform the following steps:

1. Based on the stochastic model for this temperature class, generate a set of multivariate quantiles $q^1, \dots, q^N$.
2. Try each multivariate quantile $q = q^i$, $1 \le i \le N$, in turn, stopping if q^i has been successfully validated (see below):
   a) Compute an *adjusted quantile* q' for which at least one booking-compliant nomination extending q' exists (see Section 14.5).
   b) Generate a set $P^1, \dots, P^n$ of booking-compliant nominations extending q' (see Section 14.5).
   c) Check whether *all* nominations $P^1, \dots, P^n$ are technically feasible. If this is the case, we successfully validated q^i, i.e., q^i certifies (assuming monotonicity)

that the desired probability level α is attained and stop. Otherwise, i.e., if at least one nomination is "infeasible" or "unknown", continue with the next quantile.

B. The overall booking is considered feasible if, for all temperature classes, we found a quantile for which all booking-compliant extensions are technically feasible. It is considered infeasible otherwise.

14.7 ▪ Handling interruptible capacities and flow-adjusting measures

The presented methods for verifying booked capacities may already be applied in special cases where interruptible capacities and operational options like flow commitments need not be considered. For real-world planning, however, both aspects need to be taken into account.

One way to model interruptible capacities is as follows. Recall from Section 15.2 that interruptible capacities are only taken into account since, in general, the amount of booked firm entry and exit capacities does not match. Thus interruptible capacities need to be used on the deficit side to ensure overall balancing. If interruptible capacities are really necessary to match demand and supply, the TSO may, in principle, interrupt any allocation that may not be technically realized. To ensure technical feasibility on the surplus side, it is thus sufficient that there is at least one partial nomination for the interruptible capacities such that the overall nomination is technically feasible. In the NoVa models, we may thus relax the nominated power values corresponding to interruptible capacity to power intervals, adding a new balancing constraint that ensures that the (fixed) demand of the surplus side is met. However, the NoVa approaches then need to deal with non-fixed nominations, which is not possible in the sMINLP and RedNLP models.

Extending the NoVa methods presented in this book to deal with flow commitments and other flow-adjusting measures is mathematically much more challenging. The reason is that flow commitments and similar measures grant a TSO the right to influence the load flow at certain entry or exit points, but the TSO has no control over how flows are relocated among the remaining points in response to changing supply or demand at a few points. Thus NoVa as discussed in this book needs to be interfaced with both, a model for how load flows may be adjusted based on contractual agreements and another adversarial model for the reaction of the transport customers. Straightforward modeling of flow commitments thus leads to a bilevel problem, which is likely to be practically intractable for the scale of problems considered here.

14.8 ▪ Conclusions

The presented approach to the problem of verifying booked capacities has been designed with applicability to real-world gas networks in mind. In order to obtain a tractable method, several aspects are done in a heuristic fashion, which we discuss in the following.

- ▹ The set $\mathscr{T}$ of technically feasible nominations is non-convex, in general. However, many sources of nonconvexity of the stationary problem of validating nominations can be removed when operating the network in a transient regime (see Section 11.1). Thus, this assumption is justified to some extent from a practical point of view. In fact, it is one way to limit the complexity of the problem and is implicitly used in all industrial methods known to us, see Section 4.2.2. Alternatively, it might be possible to deal with the nonconvexity by some kind of space-partitioning method.
- ▹ In general, the adjustment of statistical load scenarios or multivariate quantiles destroys

their stochastic properties. In principle it is possible to use some form of rejection sampling to ensure that the statistical load scenarios without adjustment allows booking-compliant extending nominations. However, this might be rather inefficient.

- ▹ The adversarial heuristic non-uniformly samples among the extreme points corresponding to the mixed-integer hull of the MILP (14.19). In contrast to explicitly given polytopes, it is currently not known how to efficiently sample uniformly from such a mixed-integer set.

Moreover, we see several places in which more mathematical research can lead to improvements of the presented methods:

- ▹ The effectiveness of the adversarial heuristic in finding technically infeasible nominations and the formal justification of ingredients, if only in special cases, should be investigated. For instance, does the proposed method eventually find a technically infeasible nomination if one exists?
- ▹ One goal is to obtain rigorous results establishing the assumed convergence properties of our randomized QMC method for our setting.
- ▹ Existing stability results for scenario reduction techniques establish that the optimal solution w.r.t. the reduced scenario set does not deviate much from that for the original scenario set. These should be adapted to the presented setting.

In any case, the proposed methods are a substantial improvement over the current state-of-the-art in industry. We present preliminary computational results for both methods in Chapter 15. For these computations, technical feasibility of each nomination is checked using the methods from Chapters 6–10. These results indicate that the methods may already be applied to gas networks of industrially relevant size.