*Stochastic Optimization for Operating Chemical Processes under Uncertainty

René Henrion[1], Pu Li[2], Andris Möller[1], Marc C. Steinbach[3], Moritz Wendt[2], and Günter Wozny[2]

[1] Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS), Berlin, Germany
[2] Institut für Prozess- und Anlagentechnik, Technische Universität Berlin, Germany
[3] Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Germany
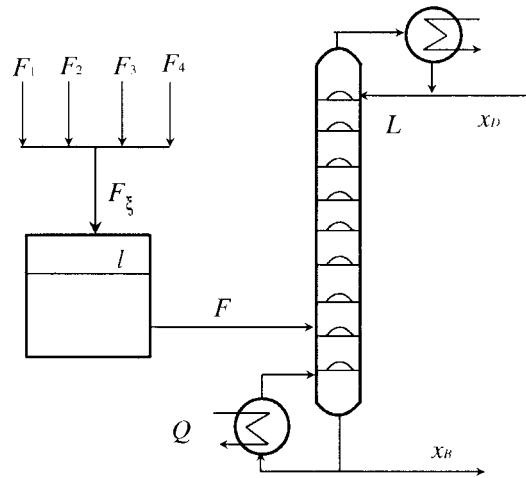
Abstract Mathematical optimization techniques are on their way to becoming a standard tool in chemical process engineering. While such approaches are usually based on deterministic models, uncertainties such as external disturbances play a significant role in many real-life applications. The present article gives an introduction to practical issues of process operation and to basic mathematical concepts required for the explicit treatment of uncertainties by stochastic optimization.

1 OPERATING CHEMICAL PROCESSES

Chemical industry plays an essential role in the daily life of our society. The purpose of a chemical process is to transfer some (cheap) materials into other (desired) materials. Those materials include any sorts of solids, liquids and gas and can be single components or multicomponent mixtures. Common examples of chemical processes are reaction, separation and crystallization processes usually composed of operation units like reactors, distillation columns, heat exchangers and so on. Based on market demands, those processes are designed, set up and put into operation. From the design, the process is expected to be run at a predefined operating point, i.e., with a certain flow rate, temperature, pressure and composition [22].

Distillation is one of the most common separation processes which consumes the largest part of energy in chemical industry. Figure 1 shows an industrial distillation process to separate a mixture of methanol and water to high purity products (methanol composition in the distillate and the bottom should be $x_D \geq 99.5$ mol% and $x_B \leq 0.5$ mol%, respectively). The feed flow F to the column is from outflows of different upstream plants. These streams are first accumulated in a tank (a middle buffer) and then fed to the column. The column is operated at atmospheric pressure. From the design, the diameter of the column, the number of trays, the reboiler duty Q and the reflux flow L will be defined for the given product specifications.

For an existing chemical process, it is important to develop flexible operating policies to improve its profitability and reducing its effect of pollution. The ever-changing market conditions demand a high flexibility for chemical processes under different product specifications and different feedstocks. On the other hand, the increasingly stringent limitations to process emissions (e.g., $x_B \leq 0.5$ mol% in the above example) require suitable new operating conditions satisfying these constraints. Moreover, the properties of processes themselves change during process

Figure 1. An industrial distillation column with a feed tank

operation, e.g., tray efficiencies and fouling of the equipment, which leads to reduction of product quality if the operating point remains unchanged. Therefore, keeping a constant operating point given by the process design is nowadays an out-dated concept. That is to say, optimal and robust operating policies should be searched for and implemented online, corresponding to the real-time process situations.

In the past, heuristic rules were used for improving process operation in chemical industry. However, since most chemical processes behave nonlinear, time-dependent and possess a large number of variables, it was impossible to find the optimal solutions or even feasible solutions by heuristic rules. Therefore, systematic methods including modeling, simulation and optimization have been developed in the last two decades for process operation. These methods are model-based deterministic approaches and have been more and more used in chemical industry [10].

1.1 Process Modeling

Conservation laws are used for modeling chemical processes. A balance space is first chosen, for which model equations will be established by balancing mass, momentum and energy input into and output from the space [3]. Thus variables of a space can be classified into independent and dependent variables. Independent variables are input variables including manipulated variables and disturbance variables. For instance, the reflux flow and the reboiler duty are usually manipulated variables for a distillation column, while the feed flow and composition are disturbance variables. Dependent variables are output variables (usually called state variables) which depend on the input variables. The compositions and temperatures on the trays inside the column are dependent variables. Besides conservation laws, cor-

relation equations based on physical and chemical principles are used to describe relations between state variables. These principles include vapor-liquid equilibrium if two phases exist in the space, reaction kinetics if a reaction takes place and fluid dynamics for describing the hydraulics influenced by the structure of the equipment.

Let us consider modeling a general tray of a distillation column, as shown in Figure 2, where i and j are the indexes of components ($i = 1, NK$) and trays (from the condenser to the reboiler), respectively. The dependent variables on each tray are the vapor and liquid compositions $y_{i,j}, x_{i,j}$, vapor and liquid flow rate V_j, L_j, liquid molar holdup M_j, temperature T_j and pressure P_j. The independent variables are the feed flow rate and composition $F_j, z_{Fi,j}$, heat flow Q_j and the flows and compositions from the upper as well as lower tray. The model equations include component and energy balances, vapor-liquid equilibrium equations, a liquid holdup equation as well as a pressure drop equation (hydraulics) for each tray of the column:

- Component balance:

$$\frac{d(M_j x_{i,j})}{dt} = L_{j-1} x_{i,j-1} + V_{j+1} y_{i,j+1} - L_j x_{i,j} - V_j y_{i,j} + F_j z_{Fi,j} \quad (1)$$

- Phase equilibrium:

$$y_{i,j} = \eta_j K_{i,j}(x_{i,j}, T_j, P_j) x_{i,j} + (1 - \eta_j) y_{i,j+1} \quad (2)$$

- Summation equation:

$$\sum_{i=1}^{NK} x_{i,j} = 1, \qquad \sum_{i=1}^{NK} y_{i,j} = 1 \quad (3)$$

- Energy balance:

$$\frac{d(M_j H_j^L)}{dt} = L_{j-1} H_{j-1}^L + V_{j+1} H_{j+1}^V - L_j H_j^L - V_j H_j^V + F_j H_{F,j}^L + Q_j \quad (4)$$
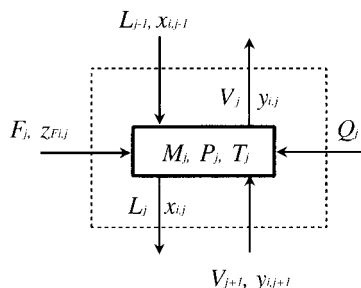
- Holdup correlation:

$$M_j = \varphi_j(x_{i,j}, T_j, L_j) \quad (5)$$

- Pressure drop equation:

$$P_j = P_{j-1} + \psi(x_{i,j-1}, y_{i,j}, L_{j-1}, V_j, T_j) \quad (6)$$

In addition to the equations (1)–(6), there are auxiliary relations to describe the vapor and liquid enthalpy H_j^V, H_j^L, phase equilibrium constant $K_{i,j}$, holdup correlation φ_j and pressure drop correlation ψ_j which are functions of the dependent variables. Parameters in these correlations can be found in chemical engineering handbooks like [9, 19]. Murphree tray efficiency η_j is introduced to describe the nonequilibrium behavior. This is a parameter that can be verified by comparing the simulation results with the operating data.

Figure 2. A general tray of the distillation column

Equations of all trays in the column lead to a complicated nonlinear DAE system. Moreover, some dependent variables are required to be kept at a predefined value (e.g., the bottom liquid level of the column). This will be realized by feedback control loops usually with PID (proportional-integral-derivative) controllers. Thus controller equations have to be added to the model equation system, if closed loop behaviors will be studied. Process simulation means, with given independent variables, to solve the DAE so as to gain the profiles of the dependent variables. In the framework of optimization, an objective function will be defined (e.g., minimizing the energy consumption during the operation). The above DAE system will be the equality constraints. The inequality constraints consist of the distillate and bottom product specifications as well as the physical limitations of vapor and liquid flow rates. Thus a dynamic nonlinear optimization problem is formulated. Approaches to solve dynamic optimization problems use a discretization method (either multiple-shooting or orthogonal collocation) to transform the dynamic system to a NLP problem. They can be classified into simultaneous approaches, where all discretized variables are included in a huge NLP problem, and sequential approaches, where a simulation step is used to compute the dependent variables and thus only the independents will be solved by NLP. Solution approaches to such problems can be found in [15, 23]. As a result, optimal operating policies for the manipulated variables can be achieved. It should be noted that some processes may have zero degree of freedom. In the above example, when the product specifications become equalities, it implies that the independent variables at the steady state must be fixed for fulfilling these specifications.

1.2   Uncertainties in Process Operation

Although deterministic approaches have been successfully applied to many complex chemical processes, their results are only applicable if the real operating conditions are included in the problem formulation. To deal with the unknown operating reality a priori, optimization under uncertainty has to be considered [13]. From the viewpoint of process operation there are two general types of uncertainties.
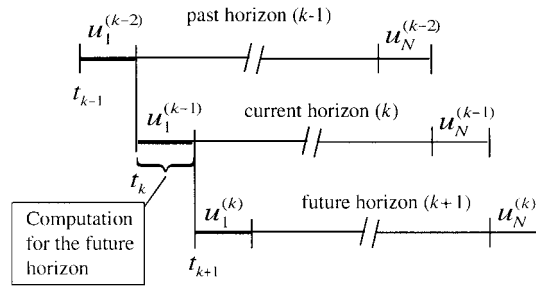
Internal Uncertainties

These uncertainties represent the unavailability of the knowledge of a process. The process model is only an approximation and thus can not describe the real behavior of the process exactly. Internal uncertainties include the model structure and the parameter uncertainty. For the description of a chemical or a thermodynamic phenomenon several representations always exist. The selection of a representation for the model leads to a structure uncertainty. Model parameters (such as parameters of reaction kinetics and vapor-liquid equilibrium) are usually estimated from a limited number of experimental data and hence the model may not be able to predict the actual process [28].

External Uncertainties

These uncertainties, mainly affected by market conditions, are from outside but have impacts on the process. These can be the flow rate and composition of the feedstock, product specifications as well as the supply of utilities. The outlet stream from an upstream unit and the recycle stream from a downstream unit are usually uncertain streams of the considered operating unit. For some processes which are sensitive to the surrounding conditions, the atmospheric temperature and pressure will be considered as external uncertain variables.

While some uncertain variables are treated as constants during the process operation, there are some time-dependent uncertain variables which are dependent on the process operating conditions. For instance, the tray efficiency of a distillation column often changes with its vapor and liquid load. Another example is the uncertainty of the feed streams coming from the upstream plants. In these cases a dynamic stochastic optimization problem will be formulated. For such problems, rather than individual stochastic parameters, continuous stochastic processes should be considered. Approximately, most of them can be considered as normal distributed stochastic processes. There may exist correlation between these variables. Operation data from historic records can be used to estimate these stochastic properties.

In deterministic optimization approaches the expected values of uncertain variables are usually employed. In the reality the uncertain variables will deviate from their expected values. Based on the realized uncertain variables a reoptimization can be carried out to correct the results from the last iteration. For dynamic optimization, a moving horizon with N time intervals will be introduced. Figure 3 shows the implementation of the three consecutive paces of the moving horizon. At the current horizon k only the values of the available policies for the first time interval u_1 which were developed in the past horizon $k - 1$, will be realized to the process. During this time interval a reoptimization is carried out to develop the operating policies for the future horizon $k + 1$. The method in which the expected values of the uncertainties are used in the problem formulation is the so-called wait-and-see strategy. The shortcoming of this strategy is that it can not guarantee holding inequality constraints.
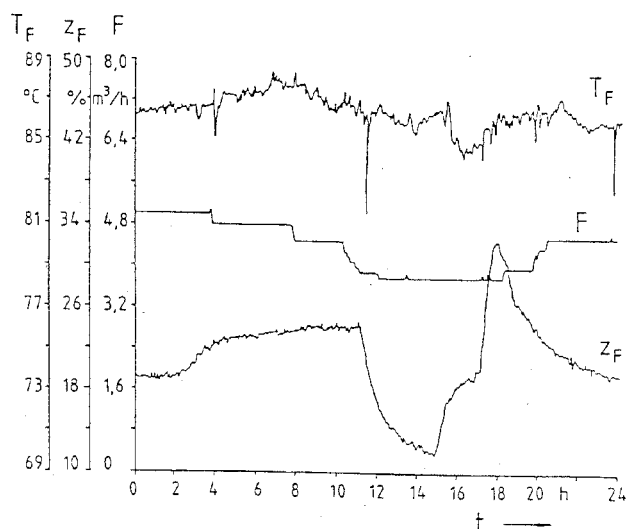
Figure 3. Reoptimization over a moving horizon

1.3   Distillation Column Operation under Uncertain Feed Streams

Now we consider again the industrial distillation process. The flows from the up-stream plants often change considerably due to the varying upstream operation. We may have high flow rates of the feed during the main working hours and decreased flow rates during the night hours or at the weekend. Figure 4 shows the measured profiles of the total feed flow, composition and temperature for 24 hours. Here we only focus on the impact of the variation of the flow rate. One consequence resulted from the fluctuating feed streams is that the tank level l may exceed the upper bound l^{max} (then a part of the liquid must be pumped out to an extra tank) or fall below the lower bound l_{min} (then a redundant feed stream must be added to the feed flow). Since the appearance of these cases will lead to considerable extra costs, a careful planning for the operation should be made to prevent these situations.

Another consequence of a large feed change is that it causes significant variations of the operating point of the distillation column. To guarantee the product quality (x_D, x_B), a conservative operating point is usually used for a higher purity than the required specification. This leads, however, to more energy consumption than required. The growth of energy requirement for a column operation is very sensitive to the product purity, especially for a high purity distillation.

Conventionally a feedback control loop is used in process industry to keep the level of the feed tank, using the outflow as the manipulated variable. The drawback of this control loop is that it can not guarantee the output constraints and it will propagate the inflow disturbance to the downstream distillation column.

To describe the continuous uncertain inflow this stochastic process will be discretized as multiple stochastic variables in fixed time intervals. We assume they have a multivariate normal distribution with an available expected profile and a covariance matrix in the considered time horizon. The reason for this assumption is that the total feed of the tank is the sum of several independent streams from the upstream plants. According to the central limit theorem [16], if a random variable is generated as the sum of effects of many independent random parameters, the distribution of the variable approaches a normal distribution, regardless of the distribution

Figure 4. Measured feed profiles of an industrial methanol-water distillation process

of each individual parameter. These parameters can be readily obtained by analyzing daily measured operating data. It is obvious that a wait-and-see strategy is not appropriate to be used in this process. Setting the feed flow with its expected profile in a deterministic optimization can not guarantee holding the tank level in the desired region. The product specification will also be easily violated by the drastically changing real feed flow. Therefore, a here-and-now strategy, which includes the uncertainties in the optimization problem, should be used. This will be discussed in the next sections.

2   MODELING UNCERTAINTY

As discussed in the previous section, a common technique of correcting random disturbances in chemical processes is *moving horizon* control (or *model predictive control*): states are measured (or estimated) in relatively small intervals, and optimal open-loop strategies are computed over a given planning horizon—"optimal" under the simplifying assumption that no further disturbances occur. In effect, the frequent repetition of this process implicitly generates a (possibly nonlinear) feedback controller that reacts to the measured disturbances.

The stochastic approaches described here are naturally applicable within such a moving horizon framework but differ in a fundamental aspect: rather than just reacting they *look ahead* by taking stochastic information on future events *explicitly* into account. This is possible if it is known which random events may occur and how

likely they are. In other words, a stochastic model of the disturbances is required, taking the form of a random process $\xi = (\xi_t)_{t \in [0,T]}$ defined on some underlying probability space $(\Omega, \mathcal{F}, P)$. Here T is the length of the planning horizon and 0 is the current time. In the present context, only $\mathbb{R}^k$-valued *discrete-time* processes for $t = 0, 1, \ldots, T$ are considered, and it is assumed that ξ_t is observed just before time t so that ξ_0 is known at $t = 0$. Thus, the processes can be seen as random variables $\xi = (\xi_1, \ldots, \xi_T)$ in $\mathbb{R}^{kT}$. Moreover, we consider either *discrete* distributions P_ξ or distributions with a continuous *density function* on $\mathbb{R}^{kT}$. (More details will be given below.) For a comprehensive treatment of the measure-theoretic and probability-theoretic foundations see, e.g., Bauer [1, 2].

Apparently the explicit modeling of uncertainty adds information to the optimization model and allows for more robust process control. The price one has to pay is the necessity of solving a *stochastic* optimization problem whose complexity may exceed the complexity of the underlying *deterministic* problem by orders of magnitude.

The precise nature of uncertainties (such as the time dependence and the significance in objective and constraints) leads to different classes of stochastic optimization models; we will describe two of them. The first approach yields a multistage recourse strategy consisting of optimal reactions to every observable sequence of random events. It minimizes expected costs while satisfying all constraints. This is appropriate if feasible solutions exist for every possible disturbance, or if costs for the violation of soft constraints can be quantified (as penalty terms). The second approach yields a single control strategy that does not react to random events but is guaranteed to satisfy the constraints with a prescribed probability. This is appropriate if constraint violations are unavoidable in certain extreme cases, or if they cause significant costs that cannot be modeled exactly. For detailed discussions of stochastic modeling aspects and problem classes we refer to the textbooks [5] and [14].
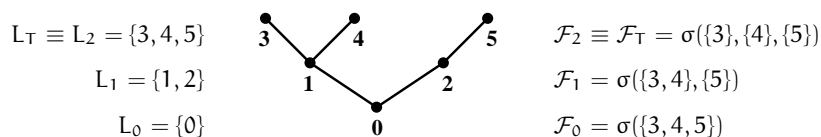
3   SCENARIO-BASED STOCHASTIC OPTIMIZATION

In scenario-based optimization, uncertainty is modeled as a finite set of possible realizations of the future with associated positive probabilities. Each realization is called a *scenario* and represents a certain event or, in our case, history of events. In precise probabilistic terms this corresponds to a discrete distribution given by a finite probability space $(\Omega, \mathcal{F}, P)$, $|\Omega| = N$. One may simply think of ω as the "number" of a scenario, which is often emphasized by using index notation. Thus, each elementary event $\omega \in \Omega$ labels a possible realization $\xi_\omega = (\xi_{\omega 1}, \ldots, \xi_{\omega T})$, and the distribution is given by N probabilities $(p_\omega)_{\omega \in \Omega}$, that is, $P_\xi(\xi_\omega) = P(\omega) = p_\omega$. (The σ-field is then simply the power set of the sample space, $\mathcal{F} = 2^\Omega$.)

3.1   Scenario Trees

As indicated, we have to deal with event histories rather than single events. This means that there is a finite number of realizations of ξ_1, each of which may lead to

$$L_T \equiv L_2 = \{3,4,5\} \qquad \qquad \mathcal{F}_2 \equiv \mathcal{F}_T = \sigma(\{3\},\{4\},\{5\})$$

$$L_1 = \{1,2\} \qquad \qquad \mathcal{F}_1 = \sigma(\{3,4\},\{5\})$$

$$L_0 = \{0\} \qquad \qquad \mathcal{F}_0 = \sigma(\{3,4,5\})$$

Figure 5. A small scenario tree with level sets and corresponding σ-fields

a different group of realizations of ξ_2, and so on. The repeated branching of partial event histories $\xi^t := (\xi_0,\ldots,\xi_t)$, called *stage* t *scenarios*, defines a *scenario tree* (or *event tree*) whose root represents ξ_0, the known observation at $t = 0$, and whose leaves represent the complete set of scenarios. Thus any node represents a group of scenarios that share a partial history ξ^t. We denote by V the set of nodes (or *vertices*) of the tree, by $L_t \subseteq V$ the level set of nodes at time t, and by $L \equiv L_T$ the set of leaves; further by $0 \in L_0$ the root, by $j \in L_t$ the "current" node, by $i \equiv \pi(j) \in L_{t-1}$ its unique predecessor (if $t > 0$), and by $S(j) \subseteq L_{t+1}$ its set of successors. The scenario probabilities are $p_j > 0$, $j \in L$. All other nodes also have a probability p_j satisfying $p_j = \sum_{k \in S(j)} p_k$. Hence, $\sum_{j \in L_t} p_j = 1$ holds for all t, and $p_0 = 1$.

Seen as a partitioning of the scenarios into groups, each level set L_t consists of *atoms* generating a sub-σ-field $\mathcal{F}_t = \sigma(L_t) \subseteq \mathcal{F}$ (where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_T = \mathcal{F}$), and ξ_t is measurable with respect to $\mathcal{F}_t$. The tree structure is thus reflected by the fact that these σ-fields form a *filtration* $\mathcal{F}_0 \subseteq \cdots \subseteq \mathcal{F}_T$ to which the process $(\xi_t)_{t=0}^T$ is *adapted*. For instance, in Figure 5 the nodes represent scenario sets as follows: $0 \leftrightarrow \{3,4,5\}$, $1 \leftrightarrow \{3,4\}$, $2 \leftrightarrow \{5\}$, $3 \leftrightarrow \{3\}$, $4 \leftrightarrow \{4\}$, and $5 \leftrightarrow \{5\}$. Since these abstract probability-theoretic notions are unnecessarily general for our purposes, we will use the more natural concept of scenario trees in the following. The notation $\xi_t = (\xi_j)_{j \in L_t}$ or $\xi = (\xi_j)_{j \in V}$ refers to the distinct realizations of ξ on level t or on the entire tree, respectively. (Here we include the deterministic initial event ξ_0 in ξ.)

3.2   Multistage Stochastic Programs

The main topic of this section are *multistage decision processes*, that is, sequences of alternating decisions and observations over the given planning horizon. The initial decision must be made without knowledge of the actual realizations of future events; hence it is based solely on ξ_0 and the probability distribution of ξ. As the future unfolds, the decision maker observes realizations ξ_j of random events ξ_t, thus collecting additional information which he or she takes into account from then on. The resulting sequence of decisions is therefore called *nonanticipative*. For instance, in controlling the feed tank of the distillation column in Section 1, we have to decide in each time step how much liquid to extract during the next period based on observations of the inflow during all previous periods and taking into account the probability distribution of future inflows. (For the initial decision, past obser-

vations do not appear explicitly in the problem but are implicitly modeled in the distribution.)

The specific class of problems considered here are (convex quadratic) *multistage recourse problems* on scenario trees, with decision vectors $y_j \in \mathbb{R}^n$, $j \in V$. Given are convex quadratic objective functions

$$\phi_j(y_j) := \frac{1}{2}y_j^* H_j y_j + f_j^* y_j,$$

and polyhedral feasible sets depending on the previous decision $y_i \equiv y_{\pi(j)}$,

$$Y_0 := \{\, y_0 \geq 0 : W_0 y_0 = h_0 \,\}, \tag{7}$$
$$Y_j(y_i) := \{\, y_j \geq 0 : W_j y_j = h_j - T_j y_i \,\}, \quad j \in V^* := V \setminus \{0\}. \tag{8}$$

These are the realizations of random costs $\phi_t(y_t)$ and random sets $Y_t(y_{t-1})$, that is, we take as random events the problem matrices and vectors

$$\xi_j = (H_j, f_j, W_j, h_j, T_j)$$

or, more generally, the functions and sets $\xi_j = (\phi_j, Y_j)$. (Conceptually we are thus allowing entirely random problem data. In practice, however, only a subset of matrix and vector elements will usually depend on an even smaller number of random influences.) Decisions y_j are to be made so as to minimize the immediate costs plus the expected costs of (optimal) future decisions; this is expressed in the general multistage recourse problem

$$\min_{y_0 \in Y_0} \phi_0(y_0)$$
$$+ \mathbb{E}_{L_0}\left[\min_{y_1 \in Y_1(y_0)} \phi_1(y_1) + \cdots + \mathbb{E}_{L_{T-1}}\left[\min_{y_T \in Y_T(y_{T-1})} \phi_T(y_T)\right]\ldots\right]. \tag{9}$$

Here $\mathbb{E}_{L_t}$ denotes the conditional expectation with respect to L_t,

$$\mathbb{E}_{L_t}(X_{t+1}) = \left\{\sum_{k \in S(j)} \frac{p_k}{p_j} X_k\right\}_{j \in L_t}.$$

The recourse structure of this class of stochastic programs is induced by the stage-coupling equations in (8); it is best seen in the *deterministic equivalent* form. Defining $\mathcal{Q}_{T+1}(y_T) := 0$ and then recursively

$$\mathcal{Q}_t(y_{t-1}) := \mathbb{E}_{L_{t-1}}\left[\min_{y_t \in Y_t(y_{t-1})} \phi_t(y_t) + \mathcal{Q}_{t+1}(y_t)\right], \quad t = T(-1)1,$$

or, in terms of the realizations,

$$\mathcal{Q}_t(y_i) := \sum_{j \in S(i)} \frac{p_j}{p_i}\left[\min_{y_j \in Y_j(y_i)} \phi_j(y_j) + \mathcal{Q}_{t+1}(y_j)\right], \quad t = T(-1)1, \quad i \in L_{t-1},$$

the deterministic equivalent problem reads

$$\min_{y_0 \in Y_0} \phi_0(y_0) + \mathcal{Q}_1(y_0).$$

This has the form of a deterministic optimization problem (hence the name), but $\mathcal{Q}_1$ is nonlinear and in general non-smooth, so it is not necessarily an appropriate formulation for numerical computations. In the case of interest one can actually unwrap the nesting of minimizations to obtain a single objective; the deterministic equivalent then takes the form of a large but structured convex quadratic program in the decision variables $y = (y_j)_{j \in V}$,

$$\min_y \sum_{j \in V} p_j \left[\frac{1}{2} y_j^* H_j y_j + f_j^* y_j \right], \tag{10}$$

$$\text{s.t. } W_0 y_0 = h_0, \tag{11}$$

$$W_j y_j = h_j - T_j y_i \quad \forall j \in V^*, \tag{12}$$

$$y_j \geq 0 \quad \forall j \in V. \tag{13}$$

This is called the *extensive form*. In stochastic notation with $y = (y_0, \ldots, y_T)$ the same problem reads

$$\min_y \sum_{t=0}^{T} \mathbb{E} \left[\frac{1}{2} y_t^* H_t y_t + f_t^* y_t \right] \tag{14}$$

$$\text{s.t. } W_0 y_0 = h_0, \tag{15}$$

$$W_t y_t = h_t - T_t y_{t-1} \quad \forall t \in \{1, \ldots, T\}, \tag{16}$$

$$y_t \geq 0 \quad \forall t \in \{0, \ldots, T\}. \tag{17}$$

Problem (14–17) and its deterministic equivalent (10–13) represent a standard problem class in stochastic programming. Especially the *linear* case with *fixed recourse* (i.e., objective $\sum_t \mathbb{E}[f_t^* y_t]$ and *deterministic* W_t) is very well-understood and widely used in practice. An important property of the deterministic equivalent is that, except for the recourse *sub*-structure, it has the form of a standard mathematical program (LP, QP, CP, or NLP). Thus, even though the scenario tree may cause exponential growth of the problem size, standard solution approaches are applicable when combined with suitable techniques that exploit the sparsity induced by the stochastic nature. The most prominent such techniques are *decomposition* approaches which split the large stochastic program into smaller problems associated with clusters of nodes (or scenarios). For a discussion of these techniques we refer the reader to the excellent survey articles [4, 21]; our own approach combines interior point methods with specially developed sparse-matrix techniques.

3.3 Dynamic Structure

The stage-coupling equations (16) or (12) define (implicitly) an underlying dynamic process, usually combined with further equality constraints. More precisely, the

rows of conditions (16) can be categorized into *dynamic equations* and certain types of constraints which possess natural interpretations and satisfy associated regularity conditions. In [25, 26] we have developed complete such categorizations for several formulations of stochastic programs, accompanied by solution algorithms that employ natural pivot orders resulting from the refined sparse structure.

In processes governed by differential (or difference) equations there is typically also a natural partitioning of the decision variables $y = (x, u)$ into (independent) *control* variables u and (dependent) *state* variables x, the former representing the actual degrees of freedom available to the decision maker. The dynamic equations are then often given in explicit form,

$$x_j = G_j x_i + E_j u_i + h_j, \tag{18}$$

which is equivalent to (12) if we define $W_j := (I \ \ 0)$ and $T_j := -(G_j \ \ E_j)$. In this notation (and with the convention $x_{\pi(0)}, u_{\pi(0)} \in \mathbb{R}^0$), the multistage stochastic program of interest takes the form

$$\min_{(x,u)} \sum_{j \in V} p_j \left[\frac{1}{2} \begin{pmatrix} x_j \\ u_j \end{pmatrix}^* \begin{pmatrix} H_j & J_j^* \\ J_j & K_j \end{pmatrix} \begin{pmatrix} x_j \\ u_j \end{pmatrix} + \begin{pmatrix} f_j \\ d_j \end{pmatrix}^* \begin{pmatrix} x_j \\ u_j \end{pmatrix} \right], \tag{19}$$

$$\text{s.t.} \quad x_j = G_j x_i + E_j u_i + h_j \quad \forall j \in V, \tag{20}$$

$$x \in [x^{min}, x^{max}], \tag{21}$$

$$u \in [u^{min}, u^{max}], \tag{22}$$

$$\sum_{j \in V} p_j (F_j x_j + D_j u_j + e_j) = 0. \tag{23}$$

Apart from the form of dynamic equations, the major difference to the standard formulation consists in the additional equality constraint (23). This condition represents a sum of expectations; we call it a *global* constraint since it may couple all nodes of the tree. In the standard formulation (10–13), such a condition cannot be modeled directly; it would require surplus variables and additional constraints.

The natural interpretation of the dynamics (18) is that the decision u_i at time $t-1$ controls all subsequent states $x_j, j \in S(i)$, at time t. This is the typical situation in discretized continuous-time processes: actually u_i determines a control action for the entire interval $(t-1, t)$ which becomes effective in x_j one period later. In other application contexts (particularly in the financial area), decisions become effective immediately, leading to dynamic equations

$$x_j = G_j x_i + E_j u_j + h_j. \tag{24}$$

Here each state x_j has "its own" control u_j rather than sharing u_i with the siblings $S(i)$.

The problem classes and solution algorithms associated with the possible formulations of dynamics are closely related; we refer to them collectively as *tree-sparse*. (For details see [24–26].) Applications are not only in discrete-time deterministic and stochastic optimal control but also in other dynamic optimization problems with

an underlying tree topology; extensions to network topologies with "few" cycles are straightforward. A very general related problem class is investigated in [20] using a similar formulation of dynamics but σ-fields and probability spaces rather than scenario trees.

3.4   Convex Programs

Since we are concerned with convex quadratic stochastic programs, we recall here some basic definitions and facts of convex optimization. A convex optimization problem has the general form

$$\min_{y \in Y} f(y) \tag{25}$$

where Y is a convex set and $f \colon Y \to \mathbb{R}$ is a convex function, that is,

$$(1-t)y_0 + ty_1 \in Y \quad \text{and} \quad f((1-t)y_0 + ty_1) \le (1-t)f(y_0) + tf(y_1)$$

for all $y_0, y_1 \in Y$ and $t \in [0,1]$. A *convex program* (CP) is the special case

$$\min_y f(y) \quad \text{s.t.} \quad g(y) = 0, \quad h(y) \ge 0, \tag{26}$$

where $g \colon \mathbb{R}^n \to \mathbb{R}^m$ is an affine mapping, $g(y) \equiv Ay + a$, and $h \colon \mathbb{R}^n \to \mathbb{R}^k$ is a (component-wise) concave mapping. If f, g, h are twice continuously differentiable, this means that the Hessians of f and $-h_i$ are positive semidefinite, $D^2 f(y) \ge 0$ and $D^2 h_i(y) \le 0$. The convex quadratic case (with $H \ge 0$) reads

$$\min_y \frac{1}{2} y^* H y + f^* y \quad \text{s.t.} \quad Ay + a = 0, \quad By + b \ge 0. \tag{27}$$

The feasible set Y is a *polyhedron* if and only if it is given by finitely many linear equalities and inequalities, as in (27). It is easily seen that all level sets $N_c := \{\, y \in Y : f(y) \le c \,\}$ of (25) are convex. Moreover, every local solution is automatically a global solution, and the set $\mathcal{S}$ of all solutions is convex. In the general case $\mathcal{S}$ may be empty even if feasible solutions exist. This happens either if f is unbounded below, $\inf_{y \in Y} = -\infty$, or if f is bounded below but the level sets N_c are unbounded for $c \downarrow \inf_{y \in Y} > -\infty$. Both situations are impossible in the convex quadratic case (27): existence of a solution $\hat{y} \in Y$ is then always guaranteed (unless the problem is infeasible). Uniqueness of a solution $\hat{y}$ holds under standard conditions. For the convex QP (27), a sufficient condition is positive definiteness of H on the null space $N(A)$ or, more generally, on its intersection with the null spaces of the rows of B associated with *strictly active* inequalities at $\hat{y}$. All this applies in particular to the stochastic problems (10–13) and (19–23). For an exhaustive treatment of the theory and numerical aspects of convex and (nonconvex) nonlinear programming we refer the reader to standard textbooks, such as [7, 8, 17].

4   STOCHASTIC OPTIMIZATION UNDER PROBABILISTIC CONSTRAINTS

An important instance of optimization problems with uncertain data occurs if the constraints depend on a stochastic parameter, such as the inequality system

$$h(x, \xi) \geq 0, \tag{28}$$

where $h \colon \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$, ξ is an s-dimensional random variable defined on some probability space $(\Omega, \mathcal{A}, P)$ and the inequality sign has to be understood component-wise. Written as such, the constraint set is not a well-defined part of an optimization problem since, usually, the decision on the variables x has to be taken before ξ can be observed. It is clear that, in order to arrive at an implementable form of the constraints, one has to remove in an appropriate way the dependence of h on specific outcomes of ξ. The most prominent approaches to do so are

- the expected value approach
- the compensation approach
- the worst case approach
- the approach by probabilistic constraints

Using expected values, the system (28) is replaced by $\mathbb{E} \, h(x, \xi) \geq 0$, which now can be understood as an inequality system depending on x only, as the expectation operator acts as an integrator over ξ. An even simpler form is obtained when the random variable itself is replaced by its expectation: $h(x, \mathbb{E}\xi) \geq 0$ (both forms coincide in case that h depends linearly on ξ). The last form corresponds to the naive idea of substituting random parameters by average values. It seems obvious (and will be demonstrated later) that such reduction to first moment information ignores substantial information about ξ. Indeed, the expectation approach guarantees the inequality system to be satisfied on the average only, but a decision x leading to a failure of a system for about half of the realizations of ξ is usually considered as unacceptable. On the other extreme, the worst case approach enforces a decision to be feasible under all possible outcomes of ξ: $h(x, \xi) \geq 0 \; \forall \xi$. This puts emphasis on absolute safety which is frequently either not realizable in the strict sense or is bought by extreme increase of costs. Although diametrically opposed in their modeling effects, both the expected value and worst case approach share some ignorance of the stochastic nature of ξ.

The basic idea of compensation relies on the possibility to adjust constraint violations in the system (28) after observation of ξ by later compensating actions. Accordingly, the set of variables splits into first stage decisions x (to be fixed before realization of ξ) and second stage decisions y (to be fixed after realization of ξ). As an example, one may think of power scheduling where an optimal load pattern of power generating units has to be designed prior to observing the unknown demand, and, where possible later gaps between supply and demand can be corrected by additional resources (e.g., hydro-thermal units, contracts etc.). The adjustment of constraint violation is modeled by an inequality system $H(x, \xi, y) \geq 0$, connecting all three types of variables and it causes additional costs $g(y, \xi)$ for the second stage

decisions. Of course, given x and ξ, y should be chosen as to minimize second stage costs among all feasible decisions. Summarizing, compensation models replace the original problem

$$\min\{\, f(x) \mid h(x, \xi) \geq 0 \,\}$$

of minimizing the costs of first stage decision under stochastic inequalities by a problem where the sum of first stage costs and expected optimal second stage costs is minimized:

$$\min\{f(x) + Q(x)\}, \ \ Q(x) = \mathbb{E}q(x, \xi), \ \ q(x, \xi) = \min\{\, g(y, \xi) \mid H(x, \xi, y) \geq 0 \,\}.$$

The compensation approach, however, requires that compensating actions exist at all and can be reasonably modeled. In many situations this is not the case. For instance, operating an abundance of inflow in a continuous distillation process may cause adjusting actions which are inconvenient to carry out or the costs of which are hard to specify. In such circumstances, emphasis is shifted towards the reliability of a system by requiring a decision to be feasible at high probability. More precisely, (28) is replaced by the probabilistic constraint

$$P(h(x, \xi) \geq 0) \geq p.$$

Here, P is the probability measure of the given probability space and $p \in (0, 1]$ is some probability level. Of course, the higher p the more reliable is the modeled system. On the other hand, the set of feasible x is more and more shrunk with $p \uparrow 1$ which makes increase the optimal value of the objective function at the same time. The extreme case $p = 1$ is similar to the worst case approach mentioned before. Fortunately, in a typical application, considerable increase in reliability can be obtained—for instance when contrasted to the expected value approach—at a small expense of the objective function and it is only for requirements close to certainty that the optimal value of the objective function worsens critically. This makes the use of probabilistic constraints a good compromise between the afore-mentioned methods. For a detailed introduction into various models of stochastic optimization the reader is referred to the monographs [5], [14] and [18].

4.1   Types of Probabilistic Constraints

Both for theoretical and practical reasons it is a good idea to identify different types of probabilistic constraints. First let us recall, that (28) is a system of inequalities given in components by $h_1(x, \xi) \geq 0, \ldots, h_m(x, \xi) \geq 0$. Now, when passing to probabilistic constraints as described before, one has the choice of integrating or separating these components with respect to the probability measure P:

$$P[h_1(x, \xi) \geq 0, \ldots, h_m(x, \xi) \geq 0] \geq p \qquad \text{or}$$
$$P[h_1(x, \xi) \geq 0] \geq p, \quad \ldots, \quad P[h_m(x, \xi) \geq 0] \geq p.$$

These alternatives are referred to as joint and individual probabilistic constraints, respectively. It is easily seen that feasibility in the first case entails feasibility in the

second case while the reverse statement is false. In the context of control problems, the components of ξ may relate to a discretization of the time interval. Then, joint probabilistic constraints express the condition that at minimum probability p certain trajectories satisfy the given constraints over the whole interval whereas individual ones mean that this statement holds true for each fixed time of the discretized interval. From the formal point of view, passing from joint to individual constraints may appear as a complication as a single inequality (with respect to the decision variables x) is turned into a system of m inequalities. However, introducing one-dimensional random variables (depending on x) $\eta_i(x) := h_i(x, \xi)$, it can be seen that the joint constraints involve all components η_i simultaneously, whereas in each of the individual constraints just one specific component η_i figures as a scalar random variable. Taking into account that the numerical treatment of probability functions involving high-dimensional random vectors is much more delicate than in dimension one, where typically a reduction to quantiles of one-dimensional distributions can be carried out, the increase in the number of inequalities is more than compensated by a much simpler implementation. Of course, the choice between both formulations is basically governed by the modeling point of view.

Another important structure of probabilistic constraints occurs if in the constraint function h decision and random variables are separated in the sense that $h(x, \xi) = \tilde{h}(x) - \hat{h}(\xi)$. Using the distribution function $F_\eta(z) := P(\eta \leq z)$ for the transformed random variable $\eta = \hat{h}(\xi)$, the resulting (joint) probabilistic constraint may be equivalently written as

$$P(h(x, \xi) \geq 0) \geq p \iff P(\tilde{h}(x) \geq \hat{h}(\xi)) \geq p \iff F_\eta(\tilde{h}(x)) \geq p.$$

In this way, the originally implicit constraint function on x has been transformed into a composed function $F_\eta \circ \tilde{h}$. Taking into account that $\tilde{h}$ is analytically given from the very beginning and that there exist satisfactory approaches of evaluating distribution functions (in particular multivariate normal distribution), one has arrived at an explicit, implementable constraint. Thus it makes sense to speak of explicit probabilistic constraints here.

In the general implicit case, the evaluation of probabilities $P(h(x, \xi) \geq 0)$ as well as of their gradients with respect to x may become very difficult and efficient only in lower dimension. Nevertheless, there is some good chance for special cases like $h_i(\cdot, \xi)$ concave. Another option for solution is passing from joint to individual constraints.

4.2   Storage Level Constraints

An important instance of probabilistic constraints arises with the control of stochastic storage levels. Here it is assumed that some reservoir storing water or energy or anything similar is subject to lower and upper capacity levels $l^{\min}$ and $l^{\max}$. The reservoir is continuously fed and emptied. The feed ξ is assumed to be stochastic whereas extraction x is carried out in a controlled way. We consider this process over a fixed time horizon $[t_a, t_b]$ and discretize ξ and x according to subintervals of

time as $(\xi_1, \ldots, \xi_s)$ and $(x_1, \ldots, x_s)$, where ξ_i and x_i denote the amount of substance directed to or extracted from the reservoir, respectively, during the i-th time subinterval. Accordingly, the current capacity level after the i-th interval amounts to $l^0 + \xi_1 + \cdots + \xi_i - x_1 - \cdots - x_i$, where l^0 refers to the initial capacity at t_a. Thus, the stochastic storage level constraints may be written as

$$l^{min} - l^0 \le \xi_1 + \cdots + \xi_i - x_1 - \cdots - x_i \le l^{max} - l^0 \quad (i = 1, \ldots, s)$$

or more compactly as the system $l^1 \le L\xi - Lx \le l^2$, where L is a lower left triangular matrix filled with '1'. Obviously, decision and random variables are separated here and, according to the preceding section, the resulting probabilistic constraints become explicit and can basically be reduced to level sets of s-dimensional distribution functions in case of joint constraints. The problem becomes particularly simple if the constraints are considered individually both with respect to the upper and lower level and to time index i. For instance, the i-th upper level constraint writes as

$$P(\eta_i \le l_i^2 + x_1 + \cdots + x_i) \ge p \iff F_{\eta_i}(l_i^2 + x_1 + \cdots + x_i) \ge p$$
$$\iff x_1 + \cdots + x_i \ge (F_{\eta_i})^{-1}(p) - l_i^2,$$

where $\eta_i = \xi_1 + \cdots + \xi_i$, F_{η_i} refers to the 1-dimensional distribution function of η_i and $(F_{\eta_i})^{-1}(p)$ denotes the (usually tabulated) p-quantile of this distribution. Consequently, the probabilistic constraints can be transformed to a system of simple linear inequalities in the decision variable x then. Storage level constraints will be considered later in the context of controlling a continuous distillation process where the role of the reservoir is played by the so-called feed tank which acts as a buffer between stochastic inflows and the operating distillation unit.

4.3 Numerical Treatment

The solution of optimization problems involving probabilistic constraints requires at least the ability of evaluating the function $\varphi(x) = P(h(x, \xi) \ge 0)$. Thinking of discretized control problems which are typically large dimensional, efficient methods like SQP have to be employed. Then, of course, the gradient of φ has to be provided as well if not even second partial derivatives.

Assuming ξ to have a density f_ξ, the function φ is formally defined as the parameter-dependent multivariate integral

$$\varphi(x) = \int_{h(x,z) \ge 0} f_\xi(z) \, dz, \tag{29}$$

where integration takes place over an s-dimensional domain. Thinking of discretized control problems again, the dimension s of the random variable may correspond to the discretization of a time interval, hence values of $s = 20$ are more than moderate. In such dimension, however, an 'exact' evaluation of the above integral by numerical integration is far from realistic. Rather, two principal 'inexact' strategies have

proven powerful in the past, namely bounding and simulation. Some rough ideas can be illustrated for the example of distribution functions, i.e., the special case where the domain of integration becomes a cell $h(x) + \mathbb{R}^s_-$. As mentioned in the previous sections, the evaluation of distribution functions is crucial for the important special case of explicit probabilistic constraints.

The generic representatives of the bounding and simulation procedures are the Bonferroni bounds and the crude Monte-Carlo estimator. The Bonferroni bounds refer to the determination of the probability $P(\bigcup_{k=1}^{s} A_k)$ of the union of s abstract probability events A_k, and they are based on the inequalities

$$\sum_{k=1}^{2m} (-1)^{k-1} S_k \leq P(\bigcup_{k=1}^{s} A_k) \leq \sum_{k=1}^{2m+1} (-1)^{k-1} S_k,$$

where $m = 1, \ldots, \lfloor s/2 \rfloor$ on the left hand side and $m = 0, \ldots, \lfloor (s-1)/2 \rfloor$ on the right hand side, and

$$S_k = \sum_{1 \leq i_1 < \cdots < i_k \leq s} P(A_{i_1} \cap \cdots \cap A_{i_k})$$

denotes the summarized probability of all possible intersections of order k. In case of $s = 2$, for instance, the very properties of a measure yield

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2) = S_1,$$

so we have recovered the first Bonferroni upper bound in a trivial case. For the evaluation of a distribution function one has

$$F(z) = P(\xi_1 \leq z_1, \ldots, \xi_s \leq z_s) = P(A_1 \cap \cdots \cap A_s) = 1 - P(\bigcup_{k=1}^{s} A_k),$$

hence the Bonferroni bounds can be applied to the last expression. Specifying these bounds for m up to 2, one gets

$$1 - S_1 \leq F(z) \leq 1 - S_1 + S_2.$$

Increasing m, these bounds become sharper and sharper until the maximum possible value of m exactly realizes the desired probability. On the other hand, the determination of S_k becomes increasingly complex. For instance, in the context of F being a multivariate normal distribution, the determination of probabilities $P(A_{i_1} \cap \cdots \cap A_{i_k})$ leads to k-dimensional integration of that distribution. This can be efficiently done for $k = 1, 2$ but gets quickly harder with higher dimension. At the same time, the number of such probability terms to be summed up in the determination of S_k equals $\binom{s}{k}$ and thus makes the numerical effort soon explode. That is why in the determination of distribution functions, one has to be content with the very few first terms S_k. Often, the gap between the resulting Bonferroni bounds is too large for practical purposes then.

Fortunately, sharper bounds can be derived on the basis of appropriate linear programs (see [18]). For $k \leq 4$ there even exist explicit expressions for these improved bounds, for instance $1 - S_1 + \frac{2}{s}S_2 \leq F(z)$ provides a much better lower bound on the basis of S_k with $k \leq 2$ than the Bonferroni counterpart $1 - S_1 \leq F(z)$ (where S_2 does not figure at all in the first lower bound). Still the gap may remain unsatisfactory. Another strategy of deriving bounds relies on graph-theoretical arguments. The prominent Hunter bound (see [12]), for instance, is based on finding a maximum weight spanning tree in a graph the vertices of which are represented by the single events A_k and the edges of which correspond to pairwise intersections of events $A_k \cap A_l$. The weight of an edge is given by the probability $P(A_k \cap A_l)$ which is easily calculated for all edges. The Hunter bound can be shown to be at least as good as, but frequently much better than the (improved) lower Bonferroni bound $1 - S_1 + \frac{2}{s}S_2$ mentioned above, although calculated with basically the same effort. The idea behind the Hunter bound has been continuously generalized towards more complex graph structures (hypertrees defined by hyperedges) in the last few years resulting in amazingly efficient lower and upper bounds. Excellent results for the multivariate normal distribution are reported in [6] with dimension up to $s = 40$.

The simplest scheme of Monte-Carlo simulation for evaluating (29) consists in generating a sample of N realizations $z_1, \ldots, z_N$ of ξ and to take then the ratio k/N as an estimate for the desired probability, where $k = \#\{ i \mid h(x, z_i) \geq 0 \}$. For larger dimension s, the variance of this estimate becomes quite large which makes it unsatisfactory soon. Similar to the starting point of Bonferroni bounds, more efficient simulation schemes have been developed as well. At this point, we may refer to Szántai's simulation scheme (see [27], related approaches are described in [18]) which is based on the knowledge of the first two terms S_1, S_2 of probabilities of single events and pairwise intersections. Using the same sample as already generated for the crude Monte-Carlo estimator, these terms allow immediately to calculate two additional Monte-Carlo estimators, the reason behind being a simple cancellation rule of binomial expressions. Now, the main idea is to convexly combine these three Monte-Carlo estimators (including the crude one) and to exploit correlations between them in order to minimize the variance of the combined estimator. In this way, simulation results become considerably more precise. Finally, an extension to incorporating Hunter's and the other mentioned graph-theoretical bounds into this scheme has been successfully carried out.

The procedures described so far are related to the evaluation of functional values of φ in (29) with special emphasis on distribution functions. As for gradients or higher order derivatives, these can be reduced analytically to the determination of functional values again at least in case of a multivariate normal distribution (for details see [18]). Hence, the same basic strategies apply although with repeated effort now (n components for the gradient and $n(n + 1)/2$ components for the Hessian if wanted).

4.4 Probability Maximization

As already mentioned above, increasing the probability level p in a probabilistic constraint shrinks the feasible set. Typically, the feasible set becomes empty starting from a critical value $\bar{p}$ which may be less than 1. In particular, a user of some implemented solution method dealing with probabilistic constraints might unintentionally have chosen a value of p above that critical value. Then, for instance, SQP codes working with infeasible iterates and enforcing feasibility in convergence only, will consume a lot of computing time in vain due to operating on an empty constraint set. This effect is particularly undesirable in an environment of on-line optimization. Therefore, one has good reason prior to the optimization problem itself to determine $\bar{p}$ by probability maximization over the constraints:

$$\max\{\, p \mid P(h(x,\xi) \geq 0) \geq p \,\}.$$

As long as probabilistic constraints are considered alone in this auxiliary problem, it can be solved rather quickly as compared to the original optimization problem. However, one has to take into account that the obtained maximum value of p is just an upper bound for $\bar{p}$ since the other constraints of the optimization problem (usually related to the dynamics of the underlying control problem) are not involved here. At least, this bound gives an indication for a probability level which cannot be exceeded at all. In order to calculate the exact bound, one would have to include all constraints which, of course, is almost as time consuming as the original problem.

4.5 Structural Properties

For an efficient treatment of probabilistic constraints, it is crucial to have some insight into their analytical, geometrical and topological structure. While corresponding statements are well-known and immediate for usual (analytical) constraints of the type $g(x) \leq 0$ (e.g., when g is linear, convex or differentiable), there are no obvious relations between the quality of data and the structure of probabilistic constraints. Most results in this direction are concerned with convexity issues which have direct consequences for numerics and theoretical analysis. A corresponding important statement in simplified form is the following one (cf. [18]):

Theorem 1. *In (28), let the components h_i of h be convex and assume that ξ has a density the logarithm of which is concave. Then, the function $\varphi(x) = P(h(x,\xi) \geq 0)$ is concave and, hence, the corresponding probabilistic constraint may be convexly described, i.e., $P(h(x,\xi) \geq 0) \geq p \iff -\varphi(x) \leq p$.*

Many but not all of the prominent multivariate distribution share the property of having a log-concave density as required in the last theorem (e.g., multivariate normal distribution or uniform distribution on bounded convex sets, cf. [18]).
An alternative structural characterization relates to the weaker property of connectedness (cf. [11]):

Theorem 2. *The constraint set $\{ x \mid P(\tilde{h}(x) \geq \hat{h}(\xi)) \geq \underline{p} \}$ of an explicit probabilistic constraint is connected whenever the components $\tilde{h}_i$ are concave and the constraint qualification*

$$\operatorname{Im}(\tilde{h}) \cap (t \cdot (1, \ldots, 1) + \mathbb{R}^m_+) \neq \emptyset \quad \forall t \in \mathbb{R}.$$

In the affine linear case $\tilde{h}(x) = Ax + b$, this constraint qualification reduces to the positive linear independence of the rows of A.

Note that this last result does not require any assumptions on the distribution of the random variable. Applying the previous theorems to the specific situation of joint storage level constraints to be considered later on in the context of a distillation process, one may infer that the feasible set is convex for many and connected for all distributions of the random variable ξ.

REFERENCES

1. H. Bauer. *Maß- und Integrationstheorie. (Measure and integration theory)*. de Gruyter Lehrbuch. De Gruyter, Berlin, 2nd, revised edition, 1992. (In German.).
2. H. Bauer. *Probability Theory*. de Gruyter Studies in Mathematics 23. de Gruyter, Berlin, 1996. (Translated from the German by Robert B. Burckel.).
3. L. T. Biegler, I. E. Grossmann, and A. W. Westerberg. *Systematic Methods of Chemical Process Design*. Prentice-Hall, Englewood Cliffs, NJ, 1997.
4. J. R. Birge. Stochastic programming computations and applications. *INFORMS J. Comput.*, 9(2):111–133, 1997.
5. J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, New York, 1997.
6. J. Bukszár and A. Prékopa. Probability bounds with cherry trees. *Math. Oper. Res.*
7. A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley, New York, 1968. Reprinted by SIAM Publications, 1990.
8. P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Addison Wesley, 1981.
9. J. Gmehling, U. Onken, and W. Arlt. *VLE Data Collection*. DECHEMA, 1997.
10. I. E. Grossmann and A. W. Westerberg. Research challenges in process systems engineering. *AIChE J.*, 46:1700–1703, 2000.
11. R. Henrion. A note on the connectedness of chance constraints. Preprint 21, Stochastic Programming Eprint Series (SPEPS), 2000. Submitted to J. Optim. Theory Appl.
12. D. Hunter. Bounds for the probability of a union. *J. Appl. Probab.*, 13:597–603, 1976.
13. M. G. Ierapetritou, J. Acevedo, and E. N. Pistikopoulos. An optimization approach for process engineering problems under uncertainty. *Comput. Chem. Eng.*, 20:703–709, 1996.
14. P. Kall and S. W. Wallace. *Stochastic Programming*. Wiley, New York, 1994.
15. P. Li, H. A. Garcia, G. Wozny, and E. Reuter. Optimization of a semibatch distillation process with model validation on the industrial site. *Ind. Eng. Chem. Res.*, 37:1341–1350, 1998.
16. M. Loeve. *Probability Theory*. Van Nostrand-Reinhold, Princeton, NJ, 1963.
17. J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.
18. A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

19. R. C. Reid, J. M. Prausnitz, and B. E. Poling. *The Properties of Gases and Liquids*. McGraw-Hill, New York, 1987.

20. R. T. Rockafellar and R. J.-B. Wets. Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time. *SIAM J. Control Optim.*, 28(4):810–822, 1990.

21. A. Ruszczyński. Decomposition methods in stochastic programming. *Math. Programming*, 79(1-3):333–353, 1997.

22. W. D. Seider, J. D. Seader, and D. R. Lewin. *Process Design Principles: Synthesis, Analysis and Evaluation*. Wiley, New York, 1998.

23. M. C. Steinbach. *Fast Recursive SQP Methods for Large-Scale Optimal Control Problems*. Ph. D. dissertation, University of Heidelberg, Germany, 1995.

24. M. C. Steinbach. Recursive direct algorithms for multistage stochastic programs in financial engineering. In P. Kall and H.-J. Lüthi, editors, *Operations Research Proceedings 1998*, 241–250, New York, 1999. Springer-Verlag.

25. M. C. Steinbach. Hierarchical sparsity in multistage convex stochastic programs. In S. P. Uryasev and P. M. Pardalos, editors, *Stochastic Optimization: Algorithms and Applications*, 363–388, Kluwer Academic Publishers, 2001. Dordrecht, The Netherlands.

26. M. C. Steinbach. Tree-sparse convex programs. Technical Report ZR-01-08, ZIB, 2001. Submitted for publication.

27. T. Szántai and A. Habib. On the k-out-of-r-from-n : F system with unequal element probabilities. In F. Gianessi et al., editor, *New Trends in Mathematical Programming*, 289–303. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

28. W. B. Whiting. Effects of uncertainties in thermodynamic data and models on process calculations. *J. Chem. Eng. Data*, 41:935–941, 1996.