Joint dynamic probabilistic constraints with projected linear decision rules

Vincent Guigues
FGV/EMAp,
22250-900 Rio de Janeiro, Brazil
vguigues@fgv.br

René Henrion
Weierstrass Institute Berlin
10117 Berlin, Germany
henrion@wias-berlin.de

Abstract

We consider multistage stochastic linear optimization problems combining joint dynamic probabilistic constraints with hard constraints. We develop a method for projecting decision rules onto hard constraints of wait-and-see type. We establish the relation between the original (infinite dimensional) problem and approximating problems working with projections from different subclasses of decision policies. Considering the subclass of linear decision rules and a generalized linear model for the underlying stochastic process with noises that are Gaussian or truncated Gaussian, we show that the value and gradient of the objective and constraint functions of the approximating problems can be computed analytically.

Keywords dynamic probabilistic constraints, multistage stochastic linear programs, linear decision rules.

1 Introduction

A probabilistic constraint is an inequality

$$\mathbb{P}\left(g(x,\xi) \leq 0\right) \geq p, \tag{1}$$

where g is a mapping defining a random inequality system, x is a decision vector, and ξ is a random vector living on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In many applications, the decision x has to be taken before the realization of the random parameter ξ is observed ('here-and-now decisions'). The meaning of (1) is the following: a decision x is feasible if and only if the random inequality system $g(x,\xi) \leq 0$ is satisfied at least with probability $p \in (0,1]$. Choosing p close to one reflects the wish for robust decisions which moreover can be interpreted in a probabilistic way. Probabilistic constraints have important applications in engineering optimization problems involving uncertain data, e.g., in water management, telecommunications, electricity network expansion, mineral blending, chemical engineering, etc. For a comprehensive overview on the theory, numerics and applications of probabilistic constraints, we refer to, e.g., [4, 13].

Often, decisions may depend on time, i.e., the vector x represents a discrete decision process. In such case, the 'here-and-now' setting of (1) means that decisions for the whole time period are taken prior to observing the random parameter, which is now a discrete

stochastic process. Then, inequality (1) represents a static probabilistic constraint because the decision process does not take into account the gain of information over time while observing the random process. To overcome this deficiency, one may pass from a decision vector $x = (x_1, \ldots, x_T)$ to a closed loop decision policy

$$x = (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \ldots, x_T(\xi_1, \ldots \xi_{T-1}))$$

each component of which represents a function of previously observed values of the random process for a given time. With this definition, (1) becomes a dynamic probabilistic constraint now acting on a variable x from an infinite dimensional space. In order to return to a numerically tractable problem in finite dimensions, the decision policies are often parameterized, the most common approach being the introduction of linear decision rules, i.e., $x_i(\xi) = A\xi + b$ for appropriate A, b which now become the finite-dimensional substitutes for the originally infinite dimensional variables. This strategy has been introduced to probabilistically constrained hydro reservoir problems as early as 1969 [15]. It was used there (and in subsequent publications) in the context of so-called individual probabilistic constraints where each component of the given random inequality system is individually turned into a probabilistic constraint:

$$\mathbb{P}\left(g_i(x, \xi) \leq 0\right) \geq p \quad (i = 1, \ldots, m).$$

The big advantage of such individual constraints is that - in case the component $g_i(x, \xi)$ is separable with respect to ξ - they are easily converted into explicit constraints via quantiles. In particular, if g happens to be a linear mapping and the objective is linear too, then all one has to do to solve such a probabilistic optimization problem is to apply linear programming. It is well known, however, that the chosen probability level p in an individual model may by far not correspond to the level in a joint model, given by (1), where the probability is taken over the entire inequality system. In [19] a hydro reservoir problem is presented where at an optimal release policy the level constraints are satisfied in each time interval with probability 90% individually, whereas the probability of keeping the level constraints through the whole time period is as low as 32%. This observation strongly suggests to deal with the joint model (1) albeit much more difficult to treat algorithmically.

The application of joint probabilistic constraints (with continuously distributed random vector) in hydro reservoir management has been pioneered in [14] where the authors considered a rolling-horizon type variant for probabilistic constraints. Joint probabilistic constraints in the closed loop sense discussed above, have been investigated in [1] again in the context of a reservoir problem. Here a highly flexible piecewise constant approximation of decision policies $x(\xi)$ was considered and it turned out that the optimal policies of the given problem were definitely not linear. However, a sufficiently fine piecewise approximation requires a big computational effort and limits the applicability of the model to a few time stages like three or four. Therefore, picking up again the idea of parameterized (in particular, linear) decision rules but now in the context of joint constraints appears to be reasonable. Other authors embed optimization problems with dynamic probabilistic constraints into a dynamic programming scheme of optimal control, however, typically imposing simplifications with regard to the joint system of constraints like the assumption of independent components, or of a discrete distribution (scenarios) or of an individualized (via Boole-Bonferroni inequality) surrogate model (e.g., [2, 11]).

The aim of the current paper is to discuss several modeling issues in the context of dynamic probabilistic constraints putting the emphasis on

- joint probabilistic constraints as in (1);
- continuous multivariate distributions of the random vector (in particular, Gaussian) with typically correlated components;
- parameterized decision rules (in particular, linear and projected linear ones);
- mixed probabilistic and hard (almost sure) constraints.

We do not intend to investigate the so-called time consistent models for dynamic probabilistic constraints as it was done, for instance, in [2]. This issue has been considered so far in the framework of Dynamic Programming, where the assumption of the random vector having independent components is paramount, e.g., [**?**]. Moreover, typically, a discrete distribution is assumed for numerical analysis. As pointed out above, we are interested here in continuously distributed distributions with potentially correlated variables. Though it seems possible to establish time consistent models for dynamic chance constraints under mutivariate Gaussian distribution, this issue would complicate the analysis we have in mind here and is yet to be explored in future research.

Moreover, the focus of this paper is not to develop a new algorithm neither the study of a concrete application, although a simple hydro reservoir problem will guide us as an illustration. Our idea is rather to provide a modeling framework taking into account the items listed above and yielding a link to algorithmic approaches for static probabilistic constraints. The latter have been successfully dealt with numerically in the context of linear probabilistic constraints under multivariate Gaussian (and Gaussian like) distribution (see, e.g., [13, 14, 19, 20]).

The paper is organized as follows: Section 2 presents a general linear multistage problem with probabilistic and hard constraints. It describes a method for projecting decision rules onto hard constraints of wait-and-see type. It finally establishes the relation between the original (infinite dimensional) problem and approximating problems working with projections from different subclasses of decision policies. These subclasses are kept very general in this section while they are specialized to linear decision rules in Section 3. In that same section the probabilistic time series model we intend to use for the discrete stochastic process is made precise. It is clarified, how the objective, the probabilistic constraint and the hard constraints look like under this probabilistic model and the assumed linear decision rules. Finally, Section 4 explicitly develops the shape of general optimization problems introduced in Section 2 when assuming multivariate Gaussian and truncated Gaussian models for the discrete process. Advantages and difficulties for the different problems are discussed.

2 A linear multistage problem with probabilistic constraints

2.1 The general model

For given $T \in \mathbb{N}$ with $T \geq 2$, we consider a T-stage stochastic linear minimization problem with the following random constraints:

$$\sum_{\tau=1}^{t} A_{t,\tau} y_\tau + \sum_{\tau=1}^{t} B_{t,\tau} \xi_\tau \leq b_t, \quad t = 1, \ldots, T. \tag{2}$$

Here, for $t = 1, \ldots, T$, y_t are n_t-dimensional decision vectors, ξ_t are M_t-dimensional random vectors, $A_{t,\tau}$ and $B_{t,\tau}$ are given matrices of orders (l_t, n_τ) and (l_t, M_τ), respectively, and $b_t \in \mathbb{R}^{l_t}$ are given vectors. In what follows, the index 't' will be interpreted as time and y_t and ξ_t represent discrete decision and stochastic processes, respectively, having finite horizon. In this time-dependent setting we shall assume that all components of the random process have the same dimension $M_1 = \cdots = M_T =: M$. The joint random vector $\xi = (\xi_1 \ldots, \xi_T) \in \mathbb{R}^{MT}$ is supposed to live in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Similarly to traditional multistage stochastic programming, we shall assume that the decision y_t is taken in the beginning of time interval $[t, t+1)$ but the random vector ξ_t is observed only at the end of that same interval. Therefore, the realization of ξ_t is unknown at the time one has to decide on y_t. On the other hand, in order to take into account the gain of information due to past observations of randomness, the decision y_t is allowed to depend on $\xi_{1:t-1} := (\xi_1, \ldots, \xi_{t-1})$ such that y_t is Borel measurable. In the following, we will refer to the $y_t(\xi_{1:t-1}), t = 1, \ldots, T$, (including the deterministic first stage decision $y_1(\xi_{1:0}) := y_1$) as decision policies rather than decision vectors in order to emphasize their functional character. Summarizing, we are dealing with the following problem:

$$\begin{gathered} \text{minimize} \quad \mathbb{E} \textstyle\sum_{t=1}^{T} \langle h_t, y_t(\xi_{1:t-1}) \rangle \quad \text{subject to} \\ \sum_{\tau=1}^{t} A_{t,\tau} y_\tau(\xi_{1:\tau-1}) + \sum_{\tau=1}^{t} B_{t,\tau} \xi_\tau \leq b_t, \quad t = 1, \ldots, T, \end{gathered} \tag{3}$$

where $\mathbb{E}$ is the expectation operator.

Example 2.1 *As an illustration, we consider a two-stage problem for the optimal release y of a hydro-reservoir under stochastic inflow ξ. The released water is used to produce and sell hydro-energy at a price p which is assumed to be known in advance. Given the two stages, these quantities have components $\xi = (\xi_1, \xi_2)$, $p = (p_1, p_2)$, $y = (y_1, y_2(\xi_1))$. The reservoir level is required to stay at both stages between given lower and upper limits ℓ^{lo}, ℓ^{up}, respectively. Finally, the release is supposed to be bounded by fixed operational limits y^{lo}, y^{up}, respectively, for turbining water at both time stages. Denoting by ℓ_0 the initial water level in the reservoir, the random cost is given by $-(p_1 y_1 + p_2 y_2(\xi_1))$ while the random constraints can be written*

$$\begin{gathered} \ell^{lo} \leq \ell_0 + \xi_1 - y_1 \leq \ell^{up} \\ \ell^{lo} \leq \ell_0 + \xi_1 + \xi_2 - y_1 - y_2(\xi_1) \leq \ell^{up} \\ y^{lo} \leq y_1 \leq y^{up} \\ y^{lo} \leq y_2(\xi_1) \leq y^{up}. \end{gathered} \tag{4}$$

It is easy to see that this is a special instance of problem (3) with data

$$h := -p,\ A_{1,1} := A_{2,2} := \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix},\ A_{2,1} := \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$B_{1,1} := B_{2,1} := B_{2,2} := \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix},\ b_1 := b_2 := \begin{pmatrix} \ell^{up} - \ell_0 \\ \ell_0 - \ell^{lo} \\ y^{up} \\ -y^{lo} \end{pmatrix}.$$

As far as the constraints are concerned, satisfying them in expectation only, would result in decisions leading to frequent violation of constraints which is not desirable for a stable operation say of technological equipment, etc. At the other extreme, constraints could be required to hold almost surely, thus yielding very robust decisions avoiding violation of constraints with probability one. In that case, we obtain the well-defined optimization problem

$$\begin{aligned} &\text{minimize } \mathbb{E} \textstyle\sum_{t=1}^{T} \langle h_t, y_t\left(\xi_{1:t-1}\right)\rangle \quad \text{subject to} \\ &\sum_{\tau=1}^{t} A_{t,\tau} y_\tau\left(\xi_{1:\tau-1}\right) + \sum_{\tau=1}^{t} B_{t,\tau}\xi_\tau \le b_t \quad t = 1,\dots,T, \qquad \mathbb{P}\text{-almost surely.} \end{aligned} \tag{5}$$

If in the constraints of (5) one had that $B_{T,T} = 0$, then the last component ξ_T of the random process would not enter the constraints and (5) would represent a conventional multistage stochastic linear program. Note, however, that $B_{2,2} \neq 0$ in the two-stage problem (4) and so the random inflow ξ_2 observed only after taking the last decision $y_2(\xi_1)$ plays a role in some of the (level) constraints. In such cases, insisting on almost sure satisfaction of constraints may be impossible in particular for unbounded random distributions. In (4), for instance, no matter what has been observed (ξ_1) or decided on (y_1,$y_2(\xi_1)$) until the beginning of the second time interval, the last unknown inflow ξ_2 could always be large enough to eventually violate the upper level constraint

$$\ell_0 + \xi_1 + \xi_2 - y_1 - y_2(\xi_1) \le \ell^{up}.$$

Therefore, one has to look for alternative models for such constraints leaving the possibility of a 'controlled' violation. These observations lead us to distinguish in (5) between *hard constraints* which have to be satisfied almost surely for physical or logical reasons and *soft constraints* which can be dealt with in a more flexible way. A typical example for hard constraints are the lower and upper limits for the amounts of turbined water ($y^{lo} \le y_1, y_2(\xi_1) \le y^{up}$) in (4): there is no turbining beyond the given operational limits just for physical reasons.

On the other hand, the reservoir level constraints could be considered to be soft ones. Suppose, for instance, that ℓ^{lo} in (4) represents the physical lower limit of the reservoir below which no water is released and turbined. Then, a violation of the lower level constraint can never happen and so the corresponding two inequalities can be removed from (4). Doing so, one has to take into account, however, that not the total amount of

the release policies y_1 and $y_2(\xi_1)$, respectively, can be turbined and sold at the given prices but only the part not violating the lower level constraint, i.e., $\min\{y_1, \ell_0 + \xi_1 - \ell^{lo}\}$ in the first stage and $\min\{y_2(\xi_1), \ell_0 + \xi_1 + \xi_2 - l^{lo} - y_1\}$ in the second stage. This means that the original profits $p_1 y_1$ and $p_2 y_2(\xi_1)$ at the two stages have to be reduced by the amounts $p_1 \left(y_1 - \ell_0 - \xi_1 + \ell^{lo}\right)_+$ and $p_2 \left(y_2(\xi_1) - \ell_0 - \xi_1 - \xi_2 + \ell^{lo} + y_1\right)_+$, respectively, where the lower index '+' as usual represents the component-wise maximum of the given expression and zero. In this way, the original lower level constraints in (4) have been removed and compensated for by appropriate penalty terms in the objective.

Next, suppose that ℓ^{up} in (4) represents some upper limit of the reservoir which is considerably lower than the physical one and serves the purpose of keeping a flood reserve. Then we may neither be able to satisfy this upper limit almost surely (see above) nor to remove it in exchange for an appropriate penalty. In such cases it is reasonable to impose a probabilistic constraint instead:

$$\mathbb{P}\left(\ell_0 + \xi_1 - y_1 \leq \ell^{up},\, \ell_0 + \xi_1 + \xi_2 - y_1 - y_2(\xi_1) \leq \ell^{up}\right) \geq p,$$

where $p \in (0,1)$ is a specified probability level. Hence, the release policies $y_1, y_2(\xi_1)$ are defined to be feasible if the indicated set of random inequalities is satisfied at least with probability p. Observe that $p = 1$ would yield the almost sure constraints again, hence choosing p close to but smaller than one, offers us the possibility of finding a feasible release policy while keeping the soft upper level constraint in a very robust sense.

Example 2.2 *Taking into account all three kinds of hard and soft constraints in the (random) hydro reservoir model (4), one ends up with the following well-defined optimization problem:*

$$\begin{array}{l} \textit{minimize} \\ -\mathbb{E}(p_1 y_1 + p_2 y_2(\xi_1)) \\ +\mathbb{E}(p_1(y_1 - \ell_0 - \xi_1 + \ell^{lo})_+ + p_2(y_2(\xi_1) - \ell_0 - \xi_1 - \xi_2 + y_1 + \ell^{lo})_+) \end{array} \tag{6}$$

subject to

$$\mathbb{P}\left(\begin{array}{r} \ell_0 + \xi_1 - y_1 \leq l^{up} \\ \ell_0 + \xi_1 + \xi_2 - y_1 - y_2(\xi_1) \leq \ell^{up} \end{array}\right) \geq p$$

$$\left.\begin{array}{r} y^{lo} \leq y_1 \leq y^{up} \\ y^{lo} \leq y_2(\xi_1) \leq y^{up} \end{array}\right\} \quad \mathbb{P}\text{-}\textit{almost surely.}$$

Here, the group of soft lower level constraints has disappeared and entered the objective as a second penalization term, the group of soft upper level constraints (for which no penalization costs are available) has turned into a probabilistic constraint and the group of hard box constraints is formulated in the almost sure sense.

Applying this strategy to the general random constraints (5), we are led to partition the data matrices and vectors for $t = 1, \ldots, T$, and $\tau = 1, \ldots, t$, as

$$A_{t,\tau} = \left(A_{t,\tau}^{(1)}, A_{t,\tau}^{(2)}, A_{t,\tau}^{(3)}\right),\ B_{t,\tau} = \left(B_{t,\tau}^{(1)}, B_{t,\tau}^{(2)}, B_{t,\tau}^{(3)}\right),\ b_t = \left(b_t^{(1)}, b_t^{(2)}, b_t^{(3)}\right)$$

according to penalized soft constraints (upper index (1)), probabilistic soft constraints (upper index (2)) and almost sure hard constraints (upper index (3)). Accordingly, (5) turns into the well-defined optimization problem

$$\begin{aligned} &\text{minimize} \\ &\sum_{t=1}^{T} \mathbb{E}\left\{ \langle h_t, y_t\left(\xi_{1:t-1}\right)\rangle + \left\langle \mathcal{P}_t, \left(\sum_{\tau=1}^{t} A_{t,\tau}^{(1)} y_\tau(\xi_{1:\tau-1}) + \sum_{\tau=1}^{t} B_{t,\tau}^{(1)} \xi_\tau - b_t^{(1)} \right)_+ \right\rangle \right\} \\ &\text{subject to} \\ &\mathbb{P}\left(\sum_{\tau=1}^{t} A_{t,\tau}^{(2)} y_\tau\left(\xi_{1:\tau-1}\right) + \sum_{\tau=1}^{t} B_{t,\tau}^{(2)} \xi_\tau \le b_t^{(2)}, \quad t=1,\dots,T \right) \ge p \\ &\sum_{\tau=1}^{t} A_{t,\tau}^{(3)} y_\tau\left(\xi_{1:\tau-1}\right) + \sum_{\tau=1}^{t} B_{t,\tau}^{(3)} \xi_\tau \le b_t^{(3)}, \quad t=1,\dots,T, \quad \mathbb{P}\text{-almost surely.} \end{aligned} \tag{7}$$

Here, the $\mathcal{P}_t \ge 0$ refer to a cost vectors penalizing the violation of soft constraints with upper index (1).

2.2 Projection onto hard constraints of wait-and-see type

We will refer in (5) to *wait-and-see constraints* if $B_{t,t} = 0$ for all $t = 1, \dots, T$, and to *here-and-now constraints* otherwise. The distinction is made according to whether in the constraint of any stage t there is unobserved randomness ξ_t left or not. For example, in (4), the first two inequalities (level constraints) are here-and-now whereas the last two (operational limits) are wait-and-see. As mentioned earlier, the almost sure constraints in (7) don't have a good chance to be ever satisfied if $B_{T,T}^{(3)} \neq 0$ and the support of the random distribution is unbounded. We'll get back to such here-and-now constraints for bounded support of the random distribution in Section 4.5. First, let us deal with the case where all hard constraints are of wait-and-see type as in (6). In this case, owing to $B_{t,t}^{(3)} = 0$ for all $t = 1, \dots, T$, the constraint set of (7) can be written as

$$\begin{aligned} M_1 := \Big\{ &\left(y_t\left(\xi_{1:t-1}\right)\right)_{t=1,\dots,T} \,| \\ &\mathbb{P}\left(\sum_{\tau=1}^{t} A_{t,\tau}^{(2)} y_\tau\left(\xi_{1:\tau-1}\right) + \sum_{\tau=1}^{t} B_{t,\tau}^{(2)} \xi_\tau \le b_t^{(2)}, \quad t=1,\dots,T \right) \ge p \\ &\sum_{\tau=1}^{t} A_{t,\tau}^{(3)} y_\tau\left(\xi_{1:\tau-1}\right) + \sum_{\tau=1}^{t-1} B_{t,\tau}^{(3)} \xi_\tau \le b_t^{(3)}, \quad t=1,\dots,T, \quad \mathbb{P}\text{-almost surely} \Big\}. \end{aligned} \tag{8}$$

In the context of numerical solution approaches, one will usually not work in the infinite-dimensional setting of all Borel measurable policies but rather with a finite dimensional approximation which may be defined by some proper subset $\mathcal{K}$ of policies. Later in this paper we will deal with the class of *linear decision rules* (see Section 3.2). The feasible set of (7) will then become the intersection $M_1 \cap \mathcal{K}$ rather than just M_1. This intersection may turn out to be very small or even empty thus leading to a poor approximation of the infinite dimensional problem (7). If, for instance one of the hard constraints is given as $y_2(\xi_1) \in [1, 2]$ ($\mathbb{P}$-almost surely) and if, moreover, the class of policies is

$$\mathcal{K} := \{(y_1, y_2(\xi_1) | \exists a \in \mathbb{R} : y_2(\xi_1) = a\xi_1\},$$

then, clearly, $M_1 \cap \mathcal{K} = \emptyset$. One possibility to avoid this kind of problem is to operate with projections of policies onto the feasible domain of hard constraints.

Given a closed convex subset X of a finite dimensional space, we denote the uniquely defined projection onto this set by π_X. For $t = 1, \ldots, T$, we introduce the multifunctions

$$X_t\left(z_{1:t-1}, \xi_{1:t-1}\right) := \left\{ y | A_{t,t}^{(3)} y(\xi_{1:t-1}) \leq b_t^{(3)} - \sum_{\tau=1}^{t-1} B_{t,\tau}^{(3)} \xi_\tau - \sum_{\tau=1}^{t-1} A_{t,\tau}^{(3)} z_\tau(\xi_{1:\tau-1}) \right\}. \tag{9}$$

Here, we adopt the previous notation $z_{1:t-1} := (z_1, \ldots, z_{t-1})$ from ξ. By Π we denote the operator which maps a policy $y := \left(y_t\left(\xi_{1:t-1}\right)\right)_{t=1,\ldots,T}$ to a new policy $z := \Pi(y)$ defined iteratively by

$$z_t\left(\xi_{1:t-1}\right) := \pi_{X_t(z_{1:t-1}, \xi_{1:t-1})}\left(y_t\left(\xi_{1:t-1}\right)\right) \quad \forall \xi,\ \forall t = 1, \ldots, T, \tag{10}$$

starting from $z_1 := \pi_{X_1}\left(y_1\right)$. For example, for $t = 1, 2, 3, \ldots$ one gets successively that

$$\begin{aligned} z_1 &: = \pi_{X_1}\left(y_1\right), \\ z_2\left(\xi_1\right) &: = \pi_{X_2(z_1, \xi_1)}\left(y_2\left(\xi_1\right)\right), \quad \forall \xi_1, \\ z_3\left(\xi_1, \xi_2\right) &: = \pi_{X_3(z_1, z_2(\xi_1), \xi_1, \xi_2)}\left(y_3\left(\xi_1, \xi_2\right)\right), \quad \forall \xi_1\ \forall \xi_2, \end{aligned}$$

so that $\Pi(y)$ is correctly defined and by (9) satisfies the hard (almost sure) constraints of (7). (10) amounts to a scenario-wise projection onto the polyhedra (9) which can be carried out numerically by solving a convex quadratic program subject to linear constraints. In the special case of rectangular sets $[y^{lo}, y^{up}]$, which can be modeled as a hard constraint in (8) by putting for $t = 1, \ldots, T$ and $\tau = 1, \ldots, t-1$:

$$A_{t,t}^{(3)} := (I, -I)^T,\ b_t^{(3)} := \begin{pmatrix} y_t^{up} \\ -y_t^{lo} \end{pmatrix},\ A_{t,\tau}^{(3)} := 0,\ B_{t,\tau}^{(3)} := 0, \tag{11}$$

an explicit formula can be exploited: projection of a policy then just means cutting it off at the given lower and upper limits. For instance, in the context of the hard constraints in (6), one has that

$$\Pi(y) = \Pi(y_1, y_2(\cdot)) = \Big(\max\{y^{lo}, \min\{y_1, y^{up}\}\}, \max\{y^{lo}, \min\{y_2(\cdot), y^{up}\}\}\Big). \tag{12}$$

As mentioned above, projection via Π is a way to enforce the hard constraints. This offers several alternatives to the above-mentioned direct intersection of feasible policies from M_1 with a given (typically finite-dimensional) subclass $\mathcal{K}$. One option would consist in working from the very beginning with projected policies so that the feasible set would become $M_1 \cap \Pi(\mathcal{K})$ rather than $M_1 \cap \mathcal{K}$. Indeed, we shall see in Lemma 2.3 that the intersection with the original infinite-dimensional feasible set may be substantially larger by doing so (in particular it would be no more empty in the example discussed before). A second option would consist in relaxing the hard constraints to probabilistic constraints similar to the ones given from the beginning and projecting them afterwards onto the set defined by hard constraints. We formalize this idea by introducing the alternative

(infinite-dimensional) constraint set

$$M_2 := \Bigg\{ \left(y_t\left(\xi_{1:t-1}\right)\right)_{t=1,\ldots,T} \mid \tag{13}$$
$$\mathbb{P}\left(\left. \begin{array}{l} \sum\limits_{\tau=1}^{t} A_{t,\tau}^{(2)} y_\tau\left(\xi_{1:\tau-1}\right) + \sum\limits_{\tau=1}^{t} B_{t,\tau}^{(2)} \xi_\tau \leq b_t^{(2)} \\ \sum\limits_{\tau=1}^{t} A_{t,\tau}^{(3)} y_\tau\left(\xi_{1:\tau-1}\right) + \sum\limits_{\tau=1}^{t-1} B_{t,\tau}^{(3)} \xi_\tau \leq b_t^{(3)} \end{array} \right\} t = 1,\ldots,T \right) \geq p \quad \Bigg\}.$$

We shall see in Lemma 2.3 that the projection of M_2 onto the hard constraints yields the set M_1, so there is no difference in the solution of (7) in the original infinite-dimensional setting. When considering intersections with a subclass $\mathcal{K}$, however, a significant advantage over working with M_1 may be observed.

2.3 Approximating the original problem by means of subclasses of decision rules

The following result clarifies the relations between the feasible sets M_1, M_2 introduced above and their intersection with (projections of) subclasses of decision rules:

Lemma 2.3 *If $\mathcal{K}$ is an arbitrary subset of Borel measurable policies $\left(y_t\left(\xi_{1:t-1}\right)\right)_{t=1,\ldots,T}$, then the following chain of inclusions holds true:*

$$M_1 \cap \mathcal{K} \subseteq \Pi(M_2 \cap \mathcal{K}) \subseteq M_1 \cap \Pi(\mathcal{K}) \subseteq M_1.$$

In particular, by setting $\mathcal{K}$ equal to the space of all Borel measurable policies, we derive that $\Pi(M_2) = M_1$.

Proof. Let $z \in M_1 \cap \mathcal{K}$. Then, the probabilistic constraint for the first and the almost sure constraints for the other inequality system in (8), respectively, guarantee that the joint probabilistic constraint in (13) is satisfied, hence $z \in M_2 \cap \mathcal{K}$. With z fulfilling the almost sure constraints in (8), we have that $z = \Pi(z)$, whence $z \in \Pi(M_2 \cap \mathcal{K})$. This proves the first inclusion in the above chain. Next, as for the second inequality, let $z \in \Pi(M_2 \cap \mathcal{K})$, hence $z = \Pi(y)$ for some $y \in M_2 \cap \mathcal{K}$. In particular, $z \in \Pi(\mathcal{K})$ and it remains to show that $z \in M_1$. As an image of the mapping Π, z satisfies the almost sure constraints of (8). By $y \in M_2$ and (13), there exists a measurable set $S \subseteq \Omega$ such that $\mathbb{P}(S) \geq p$ and

$$\begin{array}{rcl} \sum\limits_{\tau=1}^{t} A_{t,\tau}^{(2)} y_\tau\left(\xi_{1:\tau-1}(\omega)\right) + \sum\limits_{\tau=1}^{t} B_{t,\tau}^{(2)} \xi_\tau(\omega) & \leq & b_t^{(2)} \\ \sum\limits_{\tau=1}^{t} A_{t,\tau}^{(3)} y_\tau\left(\xi_{1:\tau-1}(\omega)\right) + \sum\limits_{\tau=1}^{t-1} B_{t,\tau}^{(3)} \xi_\tau(\omega) & \leq & b_t^{(3)} \end{array}$$

are satisfied for all $t = 1,\ldots,T$ and all $\omega \in S$. By (10), the second inequality system implies (successively for t from 1 to T) that

$$y_t\left(\xi_{1:t-1}(\omega)\right) \in X_t\left(z_{1:t-1}, \xi_{1:t-1}(\omega)\right) \quad \forall t = 1,\ldots,T,\ \forall \omega \in S.$$

Hence, again by (10), $(z_t(\xi_{1:t-1})(\omega))_{t=1,\dots,T} = (y_t(\xi_{1:t-1})(\omega))_{t=1,\dots,T}$ for all $\omega \in S$. Therefore, the first inequality system above can be written as

$$\sum_{\tau=1}^{t} A_{t,\tau}^{(2)} z_\tau(\xi_{1:\tau-1}(\omega)) + \sum_{\tau=1}^{t} B_{t,\tau}^{(2)} \xi_\tau(\omega) \le b_t^{(2)} \quad \forall t = 1,\dots,T,\ \forall \omega \in S.$$

Since $\mathbb{P}(S) \ge p$ it follows that z satisfies the probabilistic constraint in (8). Summarizing we have shown that also $z \in M_1$, whence the desired inclusion follows. The last inclusion is trivial. ■

The previous Lemma suggests to consider the following 4 optimization problems each of them being some relaxation of our original optimization problem (7):

$$\min\{h(y) | y \in M_1 \cap \mathcal{K}\}, \tag{14}$$

$$\min\{h(z) | z \in \Pi(\arg\min\{h(y) | y \in M_2 \cap \mathcal{K}\})\}, \tag{15}$$

$$\min\{h(y) | y \in \Pi(M_2 \cap \mathcal{K})\}, \tag{16}$$

$$\min\{h(y) | y \in M_1 \cap \Pi(\mathcal{K})\}. \tag{17}$$

Here h refers to the objective function of (7) and $\mathcal{K}$ is a given subclass of decision policies. The meaning of (14), (16) and (17) is clear and relates to the feasible sets considered in Lemma 2.3. In (15) we determine first the solution(s) of the inner optimization problem $\min\{h(y) | y \in M_2 \cap \mathcal{K}\}$ and then project them via Π. If this inner optimization problem has multiple solutions, then we choose those of their projections under Π yielding the smallest value of the objective. We observe that (16) has the same optimal value as the problem

$$\min\{h(\Pi(y)) | y \in M_2 \cap \mathcal{K}\}, \tag{18}$$

where the projection is shifted from the constraints to the objective, and that y is a solution of (18) if and only if $\Pi(y)$ is a solution of (16). Hence, (16) and (18) are equivalent and it may be a matter of convenience which of the two forms is preferred. The potential advantage of (15) say over (16) and (17) is that projections don't have to be dealt with in the constraints or in the objective directly but can be carried out after solving the problem.

Lemma 2.4 *Denote by φ_1, φ_2, φ_3, φ_4, respectively, the optimal values of problems (14)-(17) and by φ the optimal value of the originally given problem (7). Then, any solution of problems (14)-(17) is feasible for problem (7) and it holds that*

$$\varphi_1, \varphi_2 \ge \varphi_3 \ge \varphi_4 \ge \varphi.$$

Proof. From Lemma 2.3 we see that any feasible point and, hence, any solution of (14), (16) and (17) is feasible for (7). From the inclusions of Lemma 2.3 it follows that $\varphi_1 \ge \varphi_3 \ge \varphi_4 \ge \varphi$. Now, let z^* be a solution of (15). Then, there exists some $y^* \in M_2 \cap \mathcal{K}$ such that $z^* = \Pi(y^*)$ and y^* solves the problem $\min\{h(y) | y \in M_2 \cap \mathcal{K}\}$. In particular, $z^* \in \Pi(M_2 \cap \mathcal{K})$ is feasible for (16). This implies first $z^* \in M_1$ by Lemma 2.3 and, hence, the asserted feasibility of z^* for (7). Second, it implies the desired remaining relation $\varphi_2 = h(z^*) \ge \varphi_3$. ■

Lemma 2.4 can be interpreted as follows: Problem (14) reflects the pure transition to a subclass $\mathcal{K}$ of policies in the originally given problem (7). The resulting loss in optimal value equals $\varphi_1 - \varphi \geq 0$. In contrast, using projections onto hard constraints in the one or other way as in (16) and (17) may lead to smaller losses in the optimal values. Of course, this advantage of working with projections requires that the computational gain by passing to an interesting subclass $\mathcal{K}$ is not destroyed by the projection procedure. This is why in Section 3.2 we shall introduce the class of linear decision rules as a suitable one harmonizing well to a certain degree with projections onto polyhedral sets. The following example illustrates Lemma 2.4:

Example 2.5 *Consider the following problem with policies $y_1, y_2(\xi_1)$ as variables:*

$$\begin{aligned} &\min\ y_1 \quad \text{subject to} \\ &\mathbb{P}(\xi_1 \leq y_1,\ \xi_2 \leq y_2(\xi_1)) \geq p \\ &y_1,\ y_2(\xi_1) \in [0,1], \quad \mathbb{P}-\text{almost surely.} \end{aligned}$$

We assume that the random vector $\xi = (\xi_1, \xi_2)$ follows a uniform distribution over the set $\Theta = ([-1,1] \times [0,1]) \cup ([0,1] \times [0,-1])$ and that $p = 1/3$. As a subclass of policies, we consider (purely) linear second stage decisions:

$$\mathcal{K} := \{(y_1, y_2(\xi_1)) \,|\exists a \geq -1 : y_2(\xi_1) = a\xi_1\}.$$

- ***Solution of the original problem** (7):*

  We claim that the optimal value φ of the original problem equals 0. Indeed, it cannot be smaller than 0 due to the constraint $y_1 \geq 0$. On the other hand, $y_1 := 0$ and $y_2(\xi_1) := 1$ for all ξ_1 represents a feasible policy because it clearly satisfies the almost sure constraints and the set of ξ satisfying $\xi_1 \leq 0$ and $\xi_2 \leq 1$ covers one third of the support of ξ. Hence the probabilistic constraint is satisfied too. The objective value associated with this feasible policy equals $y_1 = 0$, so $\varphi = 0$ as asserted.

- ***Solution of problem** (14):*

  The feasible set here is $M_1 \cap \mathcal{K}$ and a feasible second stage policy $y_2(\xi_1) = a\xi_1$ has to be trivial ($a = 0$) in order to satisfy the almost sure constraint $0 \leq y_2(\xi_1) \leq 1$. Then, the only choice for y_1 such that $(y_1, 0)$ satisfies the probabilistic constraint is $y_1 := 1$ (only then, the set of ξ satisfying $\xi_1 \leq y_1$ and $\xi_2 \leq 0$ covers one third of the support of ξ). Hence the feasible set in this problem reduces to a singleton and its optimal value equals to the objective value of this singleton: $\varphi_1 = y_1 = 1$.

- ***Solution of problems** (15) and (16):*

  As stated above, (16) is equivalent with (18). In our example, h is the projection onto the first component, hence we seek to minimize $(\Pi(y))_1$ over the constraint set

$$\begin{aligned} M_2 \cap \mathcal{K} &= \{(y_1, a\xi_1) \mid a \geq -1, \mathbb{P}(y_1, a\xi_1 \in [0,1],\ \xi_1 \leq y_1,\ \xi_2 \leq a\xi_1) \geq 1/3\} \\ &= \{(y_1, a\xi_1) \mid y_1 \in [0,1], a \geq -1, \psi(y_1, a) \geq 1/3\} \end{aligned} \tag{19}$$

  where $\psi(y_1, a) := \mathbb{P}((\xi_1, \xi_2) \in \mathcal{S}(a, y_1))$ with

$$\mathcal{S}(a, y_1) := \{(\xi_1, \xi_2) \in \Theta \mid \xi_1 \leq y_1, \xi_2 \leq a\xi_1, 0 \leq a\xi_1 \leq 1\}$$

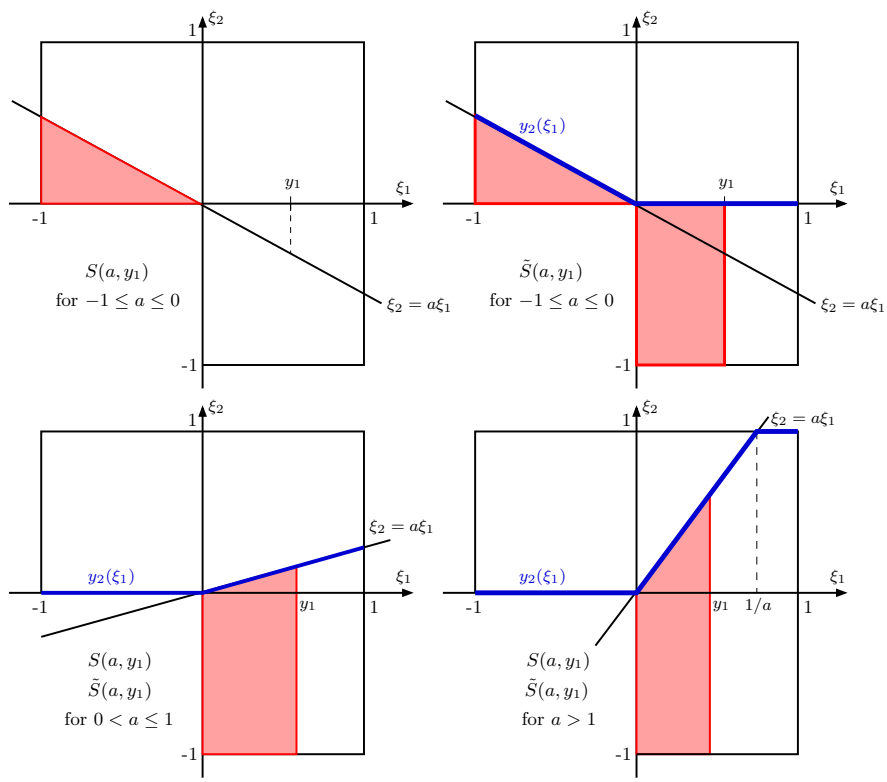

Figure 1: Representations of $\mathcal{S}(a, y_1)$ and $\tilde{\mathcal{S}}(a, y_1)$: top figures for $-1 \le a \le 0$, bottom left for $0 < a \le 1$, and bottom right for $a > 1$.

(see Figure 1). Note that in (19) *we were allowed to extract the deterministic constraint* $y_1 \in [0, 1]$ *from the probabilistic constraint.*

As $(\Pi(y))_1$ *is the projection of* y_1 *onto the first stage almost sure constraint set* $X_1 = [0, 1]$ *(see (10) and (9)), we get that* $(\Pi(y))_1 = y_1$*. Consequently, according to (18), we want to minimize* y_1 *for all policies* $(y_1, a\xi_1)$ *belonging to (19). We consider three cases:*

(i) For $-1 \le a \le 0$ *(see top left in Figure 1), we have* $\psi(y_1, a) = -a/6 < 1/3$.

(ii) For $0 < a \le 1$ *(see bottom left in Figure 1), we have* $\psi(y_1, a) = \frac{1}{3}(y_1 + ay_1^2/2)$*. The smallest value of* y_1 *satisfying* $\psi(y_1, a) \ge 1/3$ *is obtained taking* $a = 1$ *and* $y_1 = -1 + \sqrt{3} > 2/3$.

(iii) For $a > 1$ *(see bottom right in Figure 1), we get*

$$\psi(y_1, a) = \begin{cases} \frac{1}{3}(y_1 + ay_1^2/2) & \text{if } y_1 \le 1/a, \\ \frac{1}{2a} & \text{otherwise.} \end{cases}$$

In particular, $\psi(\frac{2}{3}, \frac{3}{2}) = \frac{1}{3}$*. We distinguish the two subcases:*

(1) $a > 3/2$*: if* $y_1 > 1/a$ *then* $\psi(y_1, a) = \frac{1}{2a} < \frac{1}{3}$ *and if* $0 \le y_1 \le \frac{1}{a}$ *then* $\psi(y_1, a) \le \psi(\frac{1}{a}, a) = \frac{1}{2a} < \frac{1}{3}$.

(2) $1 \le a \le 3/2$*: If* $0 \le y_1 < \frac{2}{3}$ *then* $y_1 \le 1/a$ *and, hence,*

$$\psi(y_1, a) < \frac{1}{3}\left(\frac{2}{3} + \frac{3}{2}\frac{4}{2 \cdot 9}\right) = \frac{1}{3}$$

Summarizing, the best value of the objective at an admissible solution of (18) equals $\frac{2}{3}$ *and is realized uniquely by the optimal policy* $\left(\frac{2}{3}, \frac{3}{2}\xi_1\right)$*. The latter is therefore the unique optimal solution of (18). According to our observation above, its projection*

$$\Pi\left(\frac{2}{3}, \frac{3}{2}\xi_1\right) = \left(\frac{2}{3}, \max\{0, \min\{\frac{3}{2}\xi_1, 1\}\}\right) \tag{20}$$

onto the almost sure constraints in our example is an optimal solution of (16). The associated function value equals $h\left(\frac{2}{3}, \frac{3}{2}\xi_1\right) = 2/3$ *which therefore is the optimal value of* (16)*. It follows that* $\varphi_3 = 2/3$*.*

On the other hand, as we have already observed that $h(y) = h(\Pi(y)) = y_1$ *due to* $0 \le y_1 \le 1$*, it follows that the unique optimal solution* $\left(\frac{2}{3}, \frac{3}{2}\xi_1\right)$ *of (18) yields the unique optimal solution to the problem*

$$\min\{h(y) | y \in M_2 \cap \mathcal{K}\}$$

at the same time. Hence, its projection onto the almost sure constraints is the already identified solution (20) of problem (16) implying that the optimal value of problem (15) is the same as that of (16): $\varphi_2 = \varphi_3 = 2/3$*.*

- ***Solution of problem*** *(17): By virtue of* (12)*, the policies belonging to the set* $\Pi(\mathcal{K})$ *have the form* $(y_1, \max\{0, \min\{a\xi_1, 1\}\})$ *for some* $y_1 \in [0, 1]$ *and* $a \ge -1$ *(see Figure 1). Since these policies already satisfy the almost sure constraints, all one has to add in order to get a policy feasible for (17) is the satisfaction of the probabilistic constraint. Observe that*

$$M_1 \cap \Pi(\mathcal{K}) = \{(y_1, \max\{0, \min\{a\xi_1, 1\}\}) \mid y_1 \in [0, 1],\ a \ge -1,\ \tilde{\psi}(y_1, a) \ge 1/3\}$$

where $\tilde{\psi}(y_1, a) := \mathbb{P}((\xi_1, \xi_2) \in \tilde{\mathcal{S}}(a, y_1))$ *with*

$$\tilde{\mathcal{S}}(a, y_1) = \{(\xi_1, \xi_2) \in \Theta \mid \xi_1 \le y_1, \xi_2 \le \max(0, \min(a\xi_1, 1))\}$$

(see Figure 1). For $-1 \le a \le 0$ *(see Figure 1), we have*

$$\tilde{\psi}(y_1, a) = \frac{1}{3}(y_1 - \frac{a}{2}) < \tilde{\psi}(1/2, -1) = \frac{1}{3} \quad \forall y_1 < 1/2.$$

For $0 < a \le 1$ *(see bottom left in Figure 1), we have* $\tilde{\psi}(y_1, a) = \frac{1}{3}(y_1 + ay_1^2/2)$*. The smallest value of* y_1 *satisfying* $\tilde{\psi}(y_1, a) \ge 1/3$ *is obtained taking* $a = 1$ *and* $y_1 = -1 + \sqrt{3} > 1/2$*. Finally, for* $a > 1$ *(see bottom right in Figure 1), we assume that* $y_1 \le 1/2$*. Then,*

$$y_1 > 1/a \implies \tilde{\psi}(y_1, a) = \frac{1}{3}(2y_1 - \frac{1}{2a}) < \frac{y_1}{3} \le \frac{1}{3}$$

$$y_1 \le 1/a \implies \tilde{\psi}(y_1, a) = \frac{1}{3}(y_1 + \frac{ay_1^2}{2}) \le \frac{1}{3}(\frac{1}{2} + \frac{1}{2a}) < \frac{1}{3}.$$

This means that there is no feasible policy with $y_1 \le 1/2$ *and* $a > 1$*. Consequently, the optimal value of (17) equals* $\varphi_4 = 1/2$ *and is realized by the policy* $(1/2, \max\{0, \min\{-\xi_1, 1\}\})$ *which is the projection of the decision rule* $(1/2, -\xi_1) \in \mathcal{K}$ *onto the hard box constraints.*

3 Probabilistic Model and Linear decision rules

Example 2.5 has illustrated the different approximating optimization problems with respect to the given one (7). In order to formulate these ideas in a practically meaningful framework, one has to specify the probabilistic model for the random vector ξ and a suitable subclass $\mathcal{K}$ of decision rules in Lemma 2.3.

3.1 Probabilistic model

We introduce in this section the class of stochastic processes (ξ_t) we consider. Each component $\xi_t(m)$ of ξ_t follows a linear model of the form

$$\sum_{k=0}^{p_t(m)} \alpha_{t,k}(m)\xi_{t-k}(m) = \mu_t(m) + \sum_{k=0}^{q_t(m)} \beta_{t,k}(m)\varepsilon_{t-k}(m),\ m = 1, \ldots, M, \tag{21}$$

where lags $p_t(m), q_t(m)$ are nonnegative and depend on time. We assume that for every t, the coefficients $\alpha_{t,0}(m), \alpha_{t,p_t(m)}(m)$, and $\beta_{t,q_t(m)}(m)$ are nonzero.

Finally, the noises are supposed to obey centered Gaussian laws $\varepsilon_t \sim \mathcal{N}(0, \Sigma_t)$, pairwise independent for different time steps. We recall the notation $\mathcal{N}(\mu, \Sigma)$ for referring to a multivariate Gaussian distribution with mean μ and covariance matrix Σ. Hence, $\varepsilon := (\varepsilon_1, \ldots, \varepsilon_T) \sim \mathcal{N}(0, \Sigma)$, where Σ is a block-diagonal covariance matrix whose blocks are the covariance matrices Σ_t of the components ε_t.

Remark 3.1 *We assume that the parameters of model* (21) *are known. In its full generality, model* (21) *is not identifiable. Additional assumptions are needed to identify lags $p_t(m), q_t(m)$ and calibrate the model parameters. As special cases, the identifiable SARIMA (with constant lags) and Periodic Autoregressive (PAR, with periodic time dependent lags) models can be considered.*

Using iteratively model equation (21), for each instant $t = 1, \ldots, T$, we can decompose $\xi_t(m)$ as a function of noises $\varepsilon_1, \ldots, \varepsilon_t$ and of past observations of the process (ξ_t) and of the noises (observations for instants $0, -1, -2, \ldots$). More precisely, for every $t = 1, \ldots, T$ and for every component m, we have for $\xi_t(m)$ a decomposition of the form

$$\xi_t(m) = c_t(m) + \sum_{k=1}^{r_t(m)} \gamma_{t,k}(m)\xi_{1-k}(m) + \sum_{k=1}^{s_t(m)} \delta_{t,k}(m)\varepsilon_{1-k}(m) + \sum_{k=1}^{t} \theta_{t,k}(m)\varepsilon_k(m) \tag{22}$$

for some lags $r_t(m)$ and $s_t(m)$ that represent the minimal number of past observations of respectively the stochastic processes (ξ_t) and (ε_t) that are necessary to decompose $\xi_t(m)$ over its past. This decomposition will be used in the next sections. In this decomposition, the first two sums gather the past realizations of process (ξ_t) and of the noises. Lemma 5.1 stated and proved in the Appendix, provides the formulae to compute iteratively the coefficients appearing in the decompositions of $\xi_1(m), \xi_2(m), \ldots, \xi_T(m), m = 1, \ldots, M$, of the form (22) above. The computation of these coefficients is necessary when one is interested in solving the optimization problems we consider in the next sections when (ξ_t) is of the form (21). A similar decomposition for less general models was given in [7], [8].

It is convenient to write (22) in the compact form

$$\xi_t = \tilde{\mu}_t + \Theta_t \varepsilon \quad (t = 1, \ldots, T), \tag{23}$$

where for each $t = 1, \ldots, T$,

- $\tilde{\mu}_t$ is a constant vector in $\mathbb{R}^M$ with component m given by

$$\tilde{\mu}_t(m) = c_t(m) + \sum_{k=1}^{r_t(m)} \gamma_{t,k}(m)\xi_{1-k}(m) + \sum_{k=1}^{s_t(m)} \delta_{t,k}(m)\varepsilon_{1-k}(m),$$

- Θ_t is the $M \times MT$ matrix

$$\Theta_t = \big(\mathrm{diag}(\theta_{t,1}(1), \ldots, \theta_{t,1}(M)), \ldots, \mathrm{diag}(\theta_{t,t}(1), \ldots, \theta_{t,t}(M)), 0_{M \times M(T-t)}\big)$$

  where the coefficients $\theta_{t,j}(m)$ are given in Lemma 5.1.

3.2 Linear decision rules

As mentioned in Section 2.2 the numerical solution of problem (7) requires to reduce the space of all Borel measurable decision policies to some convenient finite-dimensional subspace. A simple and widely used way to do so consists in considering so-called *linear decision rules* as policies which are defined as the set

$$\mathcal{K} := \{(y_t(\xi_{1:t-1}))_{t=1,\ldots,T} \mid \exists F_t, f_t : \, y_t(\xi_{1:t-1}) = F_t \xi_{1:t-1} + f_t \quad (t = 1, \ldots, T)\}, \tag{24}$$

with matrices F_t and vectors f_t of appropriate size. Since the first stage decision y_1 is deterministic, we convene about fixing $F_1 := 0$.

3.2.1 The random inequality system under linear decision rules

Under linear decision rules and the probabilistic model (23), our generic random inequality system

$$\sum_{\tau=1}^{t} A_{t,\tau} y_\tau(\xi_{1:\tau-1}) + \sum_{\tau=1}^{t} B_{t,\tau}\xi_\tau \le b_t \quad t = 1, \ldots, T \tag{25}$$

turns into (for $t = 1, \ldots, T$)

$$\underbrace{\left(\sum_{\tau=1}^{t} A_{t,\tau} F_\tau \Theta_{1:\tau-1} + B_{t,\tau}\Theta_\tau\right)}_{H_t(x)} \varepsilon \le \underbrace{b_t - \sum_{\tau=1}^{t} B_{t,\tau}\tilde{\mu}_\tau - \sum_{\tau=1}^{t} A_{t,\tau} f_\tau - \sum_{\tau=1}^{t} A_{t,\tau} F_\tau \tilde{\mu}_{1:\tau-1}}_{\alpha_t(x)}. \tag{26}$$

In this system, ε is the transformed random vector, whereas now $x := (F_t, f_t)_{t=1,\ldots T}$ represents a finite-dimensional decision vector approximating the original decision policies $(y_t(\xi_{1:t-1}))_{t=1,\ldots,T}$. With the notation introduced below the corresponding expressions, we may compactly rewrite (26) in the form

$$H_t(x)\varepsilon \le \alpha_t(x) \quad (t = 1, \ldots, T), \tag{27}$$

where the H_t, h_t are affine linear mappings of x. When relating these mappings not to the generic system (25) but to the concrete systems of hard and soft constraints in (7) labeled by upper indices (1), (2), (3), we shall use the corresponding upper indices for the mappings H_t and h_t as well.

We observe that, thanks to affine linearity of H_t, h_t, the set of x satisfying (27) is convex for each fixed ε.

3.2.2 The objective function under linear decision rules

From (23) and ε having a centered distribution, it follows that the expectation of ξ_t equals $\tilde{\mu}_t$. Therefore, the objective of our problem (7) takes under linear decision rules the form

$$\underbrace{\sum_{t=1}^{T}\langle h_t, F_t\tilde{\mu}_{1:t-1} + f_t\rangle}_{\beta_1(x)} + \sum_{t=1}^{T}\left\langle \mathcal{P}_t, \mathbb{E}\left(\sum_{\tau=1}^{t} A_{t,\tau}^{(1)} y_\tau(\xi_{1:\tau-1}) + \sum_{\tau=1}^{t} B_{t,\tau}^{(1)}\xi_\tau - b_t^{(1)}\right)_+\right\rangle$$

where in the definition of β_1 we used once more the convention $x := (F_t, f_t)_{t=1,\dots T}$. Now, applying (27) with upper index (1) referring to the inequality subsystem penalized in the objective, we can rewrite the objective of (7) under linear decision rules as $\beta(x) := \beta_1(x) + \beta_2(x)$, where

$$\beta_2(x) := \sum_{t=1}^{T}\left\langle \mathcal{P}_t, \mathbb{E}\left(H_t^{(1)}(x)\varepsilon - \alpha_t^{(1)}(x)\right)_+\right\rangle$$

Lemma 3.2 *β is convex.*

Proof. Since β_1 is linear, it suffices to check convexity of β_2. As mentioned earlier, the mappings $H_t^{(1)}, \alpha_t^{(1)}$ are affine linear, whence the mapping $H_t^{(1)}(x)\varepsilon - \alpha_t^{(1)}(x)$ is affine linear in x. In particular, each component of this mapping is convex in x which remains true upon passing to its maximum with zero. It follows that the components of $\mathbb{E}\left(H_t^{(1)}(x)\varepsilon - \alpha_t^{(1)}(x)\right)_+$ (depending only on x) are convex. Now, the result follows from $\mathcal{P}_t \geq 0$. ∎

For implementation purposes, it is useful to have an analytic expression of the objective function. For this purpose, we need the folloming Lemma:

Lemma 3.3 *Let X be a one-dimensional Gaussian random variable distributed according to $\mathcal{N}(m, \sigma^2)$ and let $a, b \in \bar{\mathbb{R}}$ with $a \leq b$. Then, with Φ referring to the one-dimensional standard normal distribution function, it holds that*

$$\begin{aligned}\mathbb{E}[\max\{a, \min\{X, b\}\}] &= \frac{\sigma}{\sqrt{2\pi}}\left(\exp\left(-\frac{(a-m)^2}{2\sigma^2}\right) - \exp\left(-\frac{(b-m)^2}{2\sigma^2}\right)\right) + \\ &\quad (a-m)\Phi(\frac{a-m}{\sigma}) + (m-b)\Phi(\frac{b-m}{\sigma}) + b.\end{aligned}$$

Proof. With $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$ being the density of X and with $\tilde{\Phi}$ being the associated cumulative distribution function, we have

$$\begin{aligned}
\mathbb{E}[\max\{a, \min\{X, b\}\}] &= \\
\int_{-\infty}^{a} a f_X(x)dx + \int_a^b x f_X(x)dx + \int_b^{\infty} b f_X(x)dx &= \\
a\tilde{\Phi}(a) + \int_a^b (x-m) f_X(x)dx + m\int_a^b f_X(x)dx + b(1-\tilde{\Phi}(b)) &= \\
a\tilde{\Phi}(a) + \left[\frac{-\sigma}{\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\right]_a^b + m(\tilde{\Phi}(b) - \tilde{\Phi}(a)) + b(1-\tilde{\Phi}(b)) &= \\
\frac{\sigma}{\sqrt{2\pi}}\left(\exp\left(-\frac{(a-m)^2}{2\sigma^2}\right) - \exp\left(-\frac{(b-m)^2}{2\sigma^2}\right)\right) + (a-m)\tilde{\Phi}(a) + (m-b)\tilde{\Phi}(b) + b.
\end{aligned}$$

On the other hand, since $\sigma^{-1}(X-m) \sim \mathcal{N}(0,1)$, we have that, for all z,

$$\tilde{\Phi}(z) = \mathbb{P}(X \leq z) = \mathbb{P}(\sigma^{-1}(X-m) \leq \sigma^{-1}(z-m)) = \Phi(\sigma^{-1}(z-m))$$

and the result follows. ■

The only non-explicit part in our objective function $\beta(x)$ is the vector of expectations in the definition of $\beta_2(x)$. Its $i$thcomponent is given by

$$\mathbb{E}[\max(X(x), 0)] = \mathbb{E}[\max\{0, \min\{X(x), +\infty\}]; \quad X(x) := \left(H_t^{(1)}(x)\varepsilon - \alpha_t^{(1)}(x)\right)_i.$$

According to the transformation rules of Gaussian distributions, we know that

$$X(x) \sim \mathcal{N}(m, \sigma^2); \quad m := -(\alpha_t^{(1)}(x))_i; \quad \sigma := \sqrt{\left(H_t^{(1)}(x)\Sigma[H_t^{(1)}(x)]^T\right)_{ii}},$$

where Σ is the block-diagonal covariance matrix of ε (see Section 3). With these data, Lemma 3.3 can be employed (with $a := 0, b := +\infty$) to make the objective $\beta(x)$ fully explicit in terms of the initial data of the problem.

3.2.3 Projection of linear decision rules onto hard constraints

The solution of problems (15), (16), (17), (18) is intimately related with the ability to either explicitly or numerically compute projections $\Pi(y)$ of policies $y \in \mathcal{K}$ according to (10). In the case of linear decision rules introduced in (24), the projected policy $z := \Pi(y)$ is obtained for $y = (F_t\xi_{1:t-1} + f_t)_{t=1,\ldots,T}$ as the successive (unique) solution of (scenario-dependent) quadratic optimization problems:

$$z_t(\xi_{1:t-1}) = \begin{cases} \underset{u}{\operatorname{argmin}} \|F_t\xi_{1:t-1} + f_t - u\|^2 \\ A_{t,t}^{(3)} u \leq b_t^{(3)} - \sum_{\tau=1}^{t-1} B_{t,\tau}^{(3)}\xi_\tau - \sum_{\tau=1}^{t-1} A_{t,\tau}^{(3)} z_\tau(\xi_{1:\tau-1}), \end{cases} \tag{28}$$

$$\forall \xi, \quad \forall t = 1, \ldots, T.$$

Here, starting from $t = 1$, previously obtained solutions for z_τ are plugged in on the right-hand side of (28). Hence, for instance the first two components of z are obtained as

$$z_1 = \operatorname*{argmin}_u \left\{ \|f_1 - u\|^2 | A^{(3)}_{1,1} u \leq b^{(3)}_1 \right\}$$
$$z_2(\xi_1) = \operatorname*{argmin}_u \left\{ \|F_2\xi_1 + f_2 - u\|^2 | A^{(3)}_{2,2} u \leq b^{(3)}_2 - B^{(3)}_{2,1}\xi_1 - A^{(3)}_{2,1} z_1 \right\} \quad \forall \xi_1.$$

In the special case of box constraints

$$y_t(\xi_{1:t-1}) \in [\underline{y}_t, \overline{y}_t] \quad \mathbb{P}\text{-almost surely } t = 1, \ldots, T, \tag{29}$$

an explicit formula for the projection of $y = (F_t\xi_{1:t-1} + f_t)_{t=1,\ldots,T}$ can be provided:

$$\Pi(y) = \Big(\max\{(\underline{y}_t)_i, \min\{(F_t\xi_{1:t-1} + f_t)_i, (\overline{y}_t)_i\}\} \Big)_{t=1,\ldots,T;\, i=1,\ldots,n_t}. \tag{30}$$

3.2.4 Probabilistic constraints under Linear Decision Rules and Gaussian distribution

Under the assumption of linear decision rules (24), the originally dynamic probabilistic constraint

$$\mathbb{P}\left(\sum_{\tau=1}^{t} A_{t,\tau} y_\tau(\xi_{1:\tau-1}) + \sum_{\tau=1}^{t} B_{t,\tau}\xi_\tau \leq b_t \quad t = 1, \ldots, T \right) \geq p$$

associated with (25) and occuring in problems (7) turns into a conventional static probabilistic constraint

$$\mathbb{P}\left(H_t(x)\varepsilon \leq \alpha_t(x) \quad (t = 1, \ldots, T)\right) \geq p, \tag{31}$$

with finite-dimensional decisions $x := (F_t, f_t)_{t=1,\ldots T}$. (31) represents a joint linear probabilistic constraint under Gaussian distribution. This class has been intensively studied with respect to its analytical properties and numerical solution approaches, see, e.g., [4, 13]. For an algorithmic treatment of such probabilistic constraints within the framework of nonlinear optimization it is important to have required information about the probability function

$$\varphi(x) := \mathbb{P}\left(H_t(x)\varepsilon \leq \alpha_t(x) \quad (t = 1, \ldots, T)\right)$$

defining the inequality constraint $\varphi(x) \geq p$ in (31). In particular, procedures computing or, better, approximating values and gradients of φ are needed. As shown in [20], both tasks can be realized simultaneously by reduction to the computation of multivariate Gaussian distribution functions. The latter can be quite efficiently done using Genz' code as described in [5]. An alternative approach consists in the use of the so-called spheric-radial decomposition of Gaussian random vectors [3, 16, 18]. Another important property for algorithmic purposes is convexity of the feasible set described by (31). While this is well-known to be true in case of constant matrices H_t and mappings α_t having concave components [13, Theorem 10.2.1], the same does not hold true in general for (31), in particular not for arbitrary probability levels p. Apart from special cases, such as the presence of one single random inequality in the system [10, 21] or specially structured covariance matrices [12, 9], where convexity for sufficiently large p could be guaranteed, no general result on this issue seems to be available so far.

4 Approximating optimization problems under Linear Decision Rules and Gaussian and truncated Gaussian distribution

4.1 First optimization problem

The first optimization problem we address is (14), i.e., the original problem (7) but with the feasible set intersected with the class of linear decision rules (24). Making recourse to the compact notation introduced in Section 3.2, Problem (14) writes

$$\begin{aligned} \min\{\beta(x) \quad | \quad & \mathbb{P}(H_t^{(2)}(x)\varepsilon \leq \alpha_t^{(2)}(x) \quad (t=1,\ldots,T)) \geq p, \\ & H_t^{(3)}(x)\varepsilon \leq \alpha_t^{(3)}(x) \quad (t=1,\ldots,T), \quad \mathbb{P}\text{-almost surely}\}. \end{aligned} \tag{32}$$

In the definition of $\alpha_t^{(3)}$ according to (26) we have to recall that $B_{t,t}^{(3)} = 0$ for all $t = 1, \ldots, T$ according to our wait-and-see perspective on hard constraints (see Section 2.2). (32) is a nonlinear optimization problem with a joint probabilistic and a (linear) semi-infinite constraint ($\mathbb{P}$-almost surely could be replaced by 'for $\mathbb{P}$-almost all $\varepsilon \in \Xi$', where Ξ is the support of the random vector ε).

Proposition 4.1 *The hard constraint in problem* (32) *can be explicitly represented in terms of the original data (see* (26)*) as the system of linear (in-)equalities for* $t = 1, \ldots, T$*:*

$$\begin{aligned} \sum_{\tau=1}^{t} \left(A_{t,\tau}^{(3)} F_\tau \Theta_{1:\tau-1} + B_{t,\tau}^{(3)} \Theta_\tau \right) &= 0, \\ \sum_{\tau=1}^{t-1} B_{t,\tau}^{(3)} \tilde{\mu}_\tau + \sum_{\tau=1}^{t} A_{t,\tau}^{(3)} f_\tau + \sum_{\tau=1}^{t} A_{t,\tau}^{(3)} F_\tau \tilde{\mu}_{1:\tau-1} &\leq b_t^{(3)}. \end{aligned}$$

Proof. As mentioned above, the hard constraint in problem (32) can be replaced by

$$H_t^{(3)}(x)\varepsilon \leq \alpha_t^{(3)}(x) \quad \text{for } \mathbb{P}\text{-almost all } \varepsilon \in \Xi \quad (t=1,\ldots,T). \tag{33}$$

Since ε follows a mutivariate Gaussian distribution, its support is the whole space. As a consequence, some x can be feasible for (33) only if $H_t^{(3)}(x) = 0$ which in turn implies that $\alpha_t^{(3)}(x) \geq 0$. Conversely, any x satisfying these two relations is feasible for (33). Thus, we have shown that (33) is equivalent with the system $H_t^{(3)}(x) = 0$, $\alpha_t^{(3)}(x) \geq 0$. Now, (26) yields the assertion of the proposition. ■

By virtue of Proposition 4.1, the hard constraints in (32) define a polyhedral constraint set for the decision vector x. Recalling Lemma 3.2, (32) would be a convex optimization problem provided that the probabilistic constraint defines a convex feasible region. As discussed in Section 3.2.4, this can be guaranteed, however, only in certain special cases. Moreover, the range of applicability of Proposition 4.1 is potentially small:

Corollary 4.2 *Assume that all coefficients* $\theta_{t,k}$ *in* (22) *have all components different from zero. Then, if the hard constraints in* (32) *represent simple box constraints, the only feasible linear decision rules are static ones.*

Proof. For box constraints $y \in [y^{lo}, y^{up}]$, we are dealing with the data specified in (11). Accordingly, the equation derived in Proposition 4.1 yields that

$$A_{t,t}^{(3)} F_t \Theta_{1:t-1} + B_{t,t}^{(3)} \Theta_t = 0 \quad t = 1, \ldots, T.$$

Recalling that, by the assumed wait-and-see structure for the hard constraints, we have $B_{t,t}^{(3)} = 0$ for $t = 1, \ldots, T$ (see Section 2.2), and taking into account that $A_{t,t}^{(3)} = (I, -I)^T$, we derive in particular the relations $F_t \Theta_{1:t-1} = 0$ for $t = 1, \ldots, T$. Now, our assumption on coefficients $\theta_{t,k}$ ensures that the matrices $\Theta_{1:t-1}$ are surjective. As a consequence, $F_t = 0$ for $t = 1, \ldots, T$, which means that the linear decision rules in (24) reduce to $y_t(\xi_{1:t-1}) = f_t$ for $t = 1, \ldots, T$. In other words, one is back to a static decision problem. ■

In order to avoid the restrictive consequences following from the last corollary, one may pass from Gaussian to truncated Gaussian distributions having a bounded support. This will be discussed in Section 4.5.

4.2 Second optimization problem

The second optimization problem to be discussed is (15). We will focus our attention on the inner optimization problem

$$\min\{h(y) \mid y \in M_2 \cap \mathcal{K}\}. \tag{34}$$

If this problem happens to have a unique solution, then its projection via Π onto the hard constraints will be unique and thus will be a solution of the overall problem too. Otherwise, the outer optimization problem in (15) just serves the purpose of selecting the best solution among projected solutions of the inner problem possibly realizing different values of the objective function h. We will not address the issue of possible non-uniqueness of (34) here.

By (13), and using once more the compact notation of Section 3.2 along with the definition (24) of linear decision rules, problem (34) writes

$$\min\{\beta(x) \mid \mathbb{P}(H_t^{(2)}(x)\varepsilon \leq \alpha_t^{(2)}(x),\ H_t^{(3)}(x)\varepsilon \leq \alpha_t^{(3)}(x) \quad (t = 1, \ldots, T)) \geq p\}. \tag{35}$$

This problem has the same objective as (32) but the feasible set differs by the absence of hard constraints and the presence of an enlarged inequality system in the joint chance constraint. Once, a solution x^* of (35) has been determined, it is projected onto the hard constraints (either using an explicit formula if possible or by solving a quadratic optimization problem as described in Section 3.2.3) in order to yield a decision policy $\Pi(x^*)$ which is feasible for the original infinite-dimensional problem (7).

4.3 Third optimization problem

The third optimization problem we consider is (16) or its equivalent form (18). Observe first, that (16) can be written

$$\min\{h(z) | z = \Pi(y),\ y \in M_2 \cap \mathcal{K}\}.$$

The inclusion in the constraint set of this optimization problem is the same as in (34) and can thus be formulated as the probabilistic constraint in (35) under our convention $x := (F_t, f_t)_{t=1,\dots T}$. Taking into account formula (28) for the projection $z = \Pi(y)$, we arrive at the following description for problem (16):

$$\begin{aligned} \min\{h(z) \quad | \quad & z_t(\xi_{1:t-1}) = \operatorname*{argmin}_u\{\vartheta_t(x,u,\xi) \mid \gamma_t(u,\xi) \le 0 \quad \forall \xi,\ \forall t = 1,\dots,T\}, \\ & \mathbb{P}(H_t^{(2)}(x)\varepsilon \le \alpha_t^{(2)}(x),\ H_t^{(3)}(x)\varepsilon \le \alpha_t^{(3)}(x) \quad (t = 1,\dots,T)) \ge p\}, \end{aligned} \tag{36}$$

where

$$\begin{aligned} \vartheta_t(x,u,\xi) &:= \|F_t\xi_{1:t-1} + f_t - u\|^2 \\ \gamma_t(u,\xi) &:= A_{t,t}^{(3)}u + \sum_{\tau=1}^{t-1} B_{t,\tau}^{(3)}\xi_\tau + \sum_{\tau=1}^{t-1} A_{t,\tau}^{(3)} z_\tau(\xi_{1:\tau-1}) - b_t^{(3)} \end{aligned}$$

(recall that due to successive resolution of constraints in (28) the terms $z_\tau(\xi_{1:\tau-1})$ are known in step t for $\tau = 1,\dots,t-1$). Formally, (36) represents a kind of bilevel problem in variables (x, z), where the upper level variable x is subjected to a joint probabilistic constraint and the lower level variable z is subjected to a continuum of lower level problems depending on x. As such, this optimization problem appears to be very hard to solve. On the other hand, for given x satisfying the probabilistic constraint, the solutions z_t of the parametric lower level quadratic problem are piecewise linear in $\xi_{1:t-1}$ with an identifiable polyhedral decomposition of their domain. This would allow us to apply algorithms from multiparametric quadratic programming (see [17]) in order to determine the z_t.

The problem simplifies significantly if the hard constraints are simple box constraints (29) such that the explicit formula (30) can be applied. In this case, one may directly pass to the equivalent problem (18) which in our compact notation reads

$$\begin{gathered} \text{minimize} \\ \sum_{t=1}^{T} \mathbb{E}\left\{ \langle h_t, \delta_t(x,\xi)\rangle + \left\langle \mathcal{P}_t, \left(\sum_{\tau=1}^{t} A_{t,\tau}^{(1)}\delta_\tau(x,\xi) + \sum_{\tau=1}^{t} B_{t,\tau}^{(1)}\xi_\tau - b_t^{(1)} \right)_+ \right\rangle \right\} \\ \text{subject to} \\ \mathbb{P}(H_t^{(2)}(x)\varepsilon \le \alpha_t^{(2)}(x),\ H_t^{(3)}(x)\varepsilon \le \alpha_t^{(3)}(x) \quad (t = 1,\dots,T)) \ge p, \end{gathered} \tag{37}$$

where $x := (F_t, f_t)_{t=1,\dots T}$ and the components of $\delta_t(x,\xi)$ are defined as

$$(\delta_t(x,\xi))_i := \Big(\max\{(\underline{y}_t)_i, \min\{(F_t\xi_{1:t-1} + f_t)_i, (\overline{y}_t)_i\}\}\Big)_{i;\, t=1,\dots,T}. \tag{38}$$

The first part of the expectation in the objective of this problem requires just to compute the expectations $\mathbb{E}(\delta_t(x,\xi))_i$ which can be made fully explicit thanks to Lemma 3.3 upon putting there (see (23))

$$a := (\underline{y}_t)_i;\ b := (\overline{y}_t)_i;\ m := (F_t\tilde{\mu}_{1:t-1} + f_t)_i;\ \sigma := \sqrt{\left(F_t\Theta_{1:t-1}\Sigma\Theta_{1:t-1}^T F_t^T\right)_{ii}}.$$

Consequently, in the absence of penalty terms in the objective, the whole problem reduces to a standard optimization problem subject to joint linear probabilistic constraints with multivariate Gaussian distribution. It may be difficult to obtain an analytic expression for the expectation of the penalty terms applied to projected linear decision rules. In this case, more elementary techniques like Sample Average Approximation may be used to approximate these expectations numerically.

4.4 Fourth optimization problem

The last optimization problem we consider is (17). The difference with the previous optimization problems is that here decision variables are projections onto hard constraints from the very beginning. Similarly to the previous optimization problem, (17) can be written

$$\min\{h(z)|z = \Pi(y),\, z \in M_1,\, y \in \mathcal{K}\}. \tag{39}$$

Since the projection $z = \Pi(y)$ arleady ensures the hard constraint in the inclusion $z \in M_1$, it is sufficient to impose the probabilistic constraint in (8) on z. Following the idea and the notation of (36) in the previous optimization problem, one may reformulate (17) as

$$\begin{aligned} \min\{h(z) \quad | \quad & z_t(\xi_{1:t-1}) = \operatorname*{argmin}_u\{\vartheta_t(x,u,\xi) \mid \gamma_t(u,\xi) \le 0 \quad \forall \xi,\ \forall t = 1,\dots,T\}, \\ & \mathbb{P}\left(\sum_{\tau=1}^{t} A_{t,\tau}^{(2)} z_\tau\left(\xi_{1:\tau-1}\right) + \sum_{\tau=1}^{t} B_{t,\tau}^{(2)} \xi_\tau \le b_t^{(2)}, \quad t = 1,\dots,T\right) \ge p\}. \end{aligned} \tag{40}$$

Again, we are dealing with a bilevel problem in variables (x, z), where the lower level variable z is subjected to a continuum of lower level problems depending on the upper level variable x. This time, however, the probabilistic constraint does not operate on the upper but rather on the lower level variable. Moreover, it involves only the system of soft constraints (labeled by the upper index '(2)'). Evidently, in solving (40) one is faced with the same difficulties as for problem (36).

As before, there is motivation to investigate the special case of box constraints (29). Since in this case the projection $\Pi(y)$ can be made explicit via (30), we may equivalently write (39)

$$\min\{h(\Pi(y)) \mid \Pi(y) \in M_1,\, y \in \mathcal{K}\}.$$

This problem has the same objective as problem (18) and, hence, can be made explicit exactly the same way as described in the previous section for (37). The difference now comes with the occurence of *projected* linear decision rules (38) as variables in the probabilistic constraint of (40). More precisely, we are led to the following optimization problem (where again $x := (F_t, f_t)_{t=1,\dots T}$):

$$\begin{aligned} & \text{minimize} \\ & \sum_{t=1}^{T} \mathbb{E}\left\{\langle h_t, \delta_t(x,\xi)\rangle + \left\langle \mathcal{P}_t, \left(\sum_{\tau=1}^{t} A_{t,\tau}^{(1)} \delta_\tau(x,\xi) + \sum_{\tau=1}^{t} B_{t,\tau}^{(1)} \xi_\tau - b_t^{(1)}\right)_+ \right\rangle\right\} \\ & \text{subject to} \\ & \mathbb{P}\left(\sum_{\tau=1}^{t} A_{t,\tau}^{(2)} \delta_\tau(x,\xi) + \sum_{\tau=1}^{t} B_{t,\tau}^{(2)} \xi_\tau \le b_t^{(2)}, \quad t = 1,\dots,T\right) \ge p. \end{aligned} \tag{41}$$

The challenge now is to deal with the projected linear decision rules inside the probabilistic constraint and to reduce this issue to a tractable linear structure of type (31). To this aim, with each index tuple

$$(i_{1,1},\dots,i_{1,n_1},\dots,i_{T,1},\dots i_{T,n_T}) \in \{1,2,3\}^{\sum_{t=1}^{T} n_t}$$

we associate the following $x-$dependent partition of the space of events:

$$S_{(i_{1,1},\ldots,i_{1,n_1},\ldots,i_{T,1},\ldots i_{T,n_T})}(x) :=$$

$$\left\{ \omega \in \Omega \mid \left\{ \begin{array}{ll} (F_t \xi_{1:t-1}(\omega) + f_t)_j \leq (\underline{y}_t)_j & \text{if } i_{t,j} = 1 \\ (\underline{y}_t)_j \leq (F_t \xi_{1:t-1}(\omega) + f_t)_j \leq (\overline{y}_t)_j & \text{if } i_{t,j} = 2 \\ (F_t \xi_{1:t-1}(\omega) + f_t)_j \geq (\overline{y}_t)_j & \text{if } i_{t,j} = 3 \end{array} \right. \right\} .$$

Actually, this not a partition in the strict sense because the case distinction in its definition allows some overlap for nonstrict inequality signs. Due to ξ having a density, however, this overlap is of measure zero. Therefore, we are allowed to reformulate the probability function in (39) as

$$\sum_{(i_{1,1},\ldots,i_{1,n_1},\ldots,i_{T,1},\ldots i_{T,n_T}) \in \{1,2,3\}^{\sum_{t=1}^T n_t}} \mathbb{P} \left(\begin{array}{l} \xi \in S_{(i_{1,1},\ldots,i_{1,n_1},\ldots,i_{T,1},\ldots i_{T,n_T})}(x), \\ \sum\limits_{\tau=1}^{t} \sum\limits_{j=1}^{n_\tau} (\delta_\tau(x,\xi))_j (A_{t,\tau}^{(2)})_j + \sum\limits_{\tau=1}^{t} B_{t,\tau}^{(2)} \xi_\tau \leq b_t^{(2)} \\ (t = 1, \ldots, T) \end{array} \right),$$

where $(A_{t,\tau}^{(2)})_j$ refers to column j of the matrix $A_{t,\tau}^{(2)}$. Observing that, by definition,

$$(\delta_\tau(x,\xi))_j = \left\{ \begin{array}{ll} (\underline{y}_\tau)_j & \text{if } i_{\tau,j} = 1 \\ (F_\tau \xi_{1:\tau-1} + f_\tau)_j & \text{if } i_{\tau,j} = 2 \\ (\overline{y}_\tau)_j & \text{if } i_{\tau,j} = 3 \end{array} \right. ,$$

we realize that each event over which the probability is taken above, is decsribed by a system of random inequalities which is linear in the random vector ξ. Consequently, the probability of each such event above can be described by

$$\mathbb{P} \left(\tilde{H}_t^{(i_{1,1},\ldots,i_{1,n_1},\ldots,i_{T,1},\ldots i_{T,n_T})}(x) \xi \leq \tilde{\alpha}_t^{(i_{1,1},\ldots,i_{1,n_1},\ldots,i_{T,1},\ldots i_{T,n_T})}(x) \quad (t = 1, \ldots, T) \right).$$

With ξ being an affine linear mapping of ε according to (23), we may finally write the probabilistic constraint in (39) as

$$\sum_{\substack{(i_{1,1},\ldots,i_{1,n_1},\ldots,i_{T,1},\ldots i_{T,n_T}) \\ \in \{1,2,3\}^{\sum_{t=1}^T n_t}}} \mathbb{P} \left(\begin{array}{l} \tilde{H}_t^{(i_{1,1},\ldots,i_{1,n_1},\ldots,i_{T,1},\ldots i_{T,n_T})}(x) \xi \leq \\ \tilde{\alpha}_t^{(i_{1,1},\ldots,i_{1,n_1},\ldots,i_{T,1},\ldots i_{T,n_T})}(x) \quad (t = 1, \ldots, T) \end{array} \right) \geq p,$$

which now involves similar terms as (31).

Clearly this approach for dealing with the probabilistic constraint in (39) quickly becomes prohibitive due to the number $3^{\sum_{t=1}^T n_t}$ of terms in the sum above. Even if every decision policy is one-dimensional ($n_t = 1$ for all t), this yields 3^T summands and limits the applicability of the approach to say $T = 6, 7$ stages. An alternative option would consist in the application of spherical-radial decomposition as mentioned in Section 3.2.4 which is not restricted to linear probabilistic constraints and would not suffer from the complexity issue.

4.5 Optimization problem under truncated Gaussian distribution

After introducing our original optimization problem (7), we have passed immediately to hard constraints of wait-and-see type in Section 2.2 because otherwise the hard constraints wouldn't have any good chance of ever being satisfied under distributions with unbounded support, e.g., Gaussian. This change became apparent by requiring $B^{(3)}_{t,t} = 0$ in (7), leading to the hard constraints of (8). When discussing our first optimization problem (32), we noticed that even for hard constraints of wait-and-see type, the unboundedness of the support of the random vector generates a strong restriction on the feasible decisions (see Corollary 4.2). In this section we come back to the first optimization problem but with a Gaussian random vector truncated to a bounded region. This approach will allow us not only to circumvent the mentioned restriction of problem (32) but even to admit the original hard constraints in (7) with possibly $B^{(3)}_{t,t} \neq 0$.

Definition 4.3 *We say that a random vector ε follows a normal distribution with parameters μ, Σ which is truncated to a Borel measurabls set S and then write $\varepsilon \sim \mathcal{TN}(\mu, \Sigma, S)$ if there exists a Gaussian random vector $\tilde{\varepsilon} \sim \mathcal{N}(\mu, \Sigma)$ such that*

$$\mathbb{P}(\varepsilon \in B) = \frac{\mathbb{P}(\tilde{\varepsilon} \in S \cap B)}{\mathbb{P}(\tilde{\varepsilon} \in S)} \quad \textit{for all Borel sets } B.$$

In the following we shall assume in contrast with the previous sections that the noises ε_t in the probabilistic model (21) are independent and distributed according to $\varepsilon \sim \mathcal{TN}(0, \Sigma, S)$, where Σ is the block-diagonal matrix introduced in Section 3.

We are now going to check the impact of truncating the Gaussian distribution on the structure of optimization problem (32).

The terms $\mathbb{E}[\langle h_t, y_t(\xi_{1:t-1})\rangle]$ in the objective function can be computed analytically since closed-form expressions are available for the expectation of truncated normal one-dimensional random variables.

Similarly to problem (37), the expectation of the penalty terms can be approximated using Sample Average Approximation.

As far as the probabilistic constraint in (32) is concerned, the underlying probability function can be written

$$\begin{aligned} \mathbb{P}(H^{(2)}_t(x)\varepsilon \leq \alpha^{(2)}_t(x) \ (t = 1, \ldots, T)) &= \\ \frac{\mathbb{P}(\{H^{(2)}_t(x)\tilde{\varepsilon} \leq \alpha^{(2)}_t(x) \ (t = 1, \ldots, T)\} \cap \{\tilde{\varepsilon} \in S\})}{\mathbb{P}(\tilde{\varepsilon} \in S)} &= \frac{\mathbb{P}(\tilde{H}(x)\tilde{\varepsilon} \leq \tilde{\alpha}(x))}{\mathbb{P}(\tilde{\varepsilon} \in S)}, \end{aligned}$$

where, with I referring to the identity matrix of appropriate size,

$$\tilde{H}(x) := \begin{pmatrix} H^{(2)}_1(x) \\ \vdots \\ H^{(2)}_T(x) \\ I \\ -I \end{pmatrix}, \quad \tilde{\alpha}(x) := \begin{pmatrix} \alpha^{(2)}_1(x) \\ \vdots \\ \alpha^{(2)}_T(x) \\ \overline{S} \\ -\underline{S} \end{pmatrix}.$$

Consequently, the probabilistic constraint in (32) turns into

$$\mathbb{P}(\tilde{H}(x)\tilde{\varepsilon} \leq \tilde{\alpha}(x)) \geq \tilde{p}, \quad \text{where } \tilde{p} := p \cdot \mathbb{P}(\tilde{\varepsilon} \in S). \tag{42}$$

Due to $\tilde{\varepsilon}$ being a Gaussian random vector, this probabilistic constraint is exactly of the same nature as the original one in (32) which was discussed in Section 3.2.4.

Addressing finally the almost sure constraints in (32), they can be equivalently formulated as

$$\max_{\varepsilon \in S} \left(H_t^{(3)}(x) \right)^j \varepsilon \leq \alpha_{t,j}^{(3)}(x), \quad \forall t, \ \forall j, \tag{43}$$

where $\left(H_t^{(3)}(x) \right)^j$ refers to the jth line of $H_t^{(3)}(x)$.

We consider two cases for S: a box and an ellipsoid. If $S := [\underline{S}, \overline{S}]$ is a box, the maximum in the left-hand-side of (43) can be computed analytically using the following lemma:

Lemma 4.4 ([6], Lemma 2) *For any x we have that*

$$\max_{y \in S} \ x^\top y = \frac{1}{2} \left(x^\top (\underline{S} + \overline{S}) + |x|^\top (\overline{S} - \underline{S}) \right).$$

As a result, if S is a box, since $\left(H_t^{(3)}(x) \right)^j$ and $\alpha_{t,j}^{(3)}(x)$ are affine functions of x, the almost sure constraints in (32) can be reformulated as explicit convex constraints in x.

Now taking for S the ellipsoid

$$S = \{x \in \mathbb{R}^T : (x - \mu)^\top \Sigma^{-1} (x - \mu) \leq \kappa^2\},$$

if vector $w_{t,j}(x)$ is the transpose of $\left(H_t^{(3)}(x) \right)^j$ then constraint (43) can be reformulated as the explicit conic quadratic (convex) constraint

$$\mu^\top w_{t,j}(x) + \kappa \sqrt{w_{t,j}(x)^\top \Sigma w_{t,j}(x)} \leq \alpha_{t,j}^{(3)}(x).$$

We end up again with a convex optimization problem.

Finally, observe that the term $\mathbb{P}(\tilde{\varepsilon} \in S)$ in (42) can be computed numerically both in the case when S is a box (using Genz' code as described in [5] for instance) and when S is an ellipsoid.

Acknowledgment: The first author's research was supported by an FGV grant, CNPq grant 307287/2013-0, FAPERJ grants E-26/110.313/2014, and E-26/201.599/2014. The second author gratefully acknowledges support by the *FMJH Program Gaspard Monge in optimization and operations research* including support to this program by EDF as well as support by the *Deutsche Forschungsgemeinschaft* within Projekt B04 in CRC TRR 154.

References

[1] L. Andrieu, R. Henrion, and W. Römisch. A model for dynamic chance constraints in hydro power reservoir management. European Journal of Operational Research, 207:579–589, 2010.

[2] P. Carpentier, J.-P. Chancelier, G. Cohen, M. de Lara, and P. Girardeau. Dynamic consistency for stochastic optimal control problems. Annals of Operations Research, 200:247–263, 2012.

[3] I. Deák. Subroutines for computing normal probabilities of sets - computer experiences. Annals of Operations Research, 100:103–122, 2000.

[4] D. Dentcheva, A. Ruszczyński, and A. Shapiro. Lectures on stochastic programming. SIAM, Philadelphia, 2009.

[5] A. Genz and F. Bretz. Computation of Multivariate Normal and t Probabilities. Springer, Heidelberg, 2009.

[6] V. Guigues. Robust production management. Optimization and Engineering, 10:505–532, 2009.

[7] V. Guigues and C. Sagastizábal. Exploiting the structure of autoregressive processes in chance-constrained multistage stochastic linear programs. Operations Research Letters, 40:478–483, 2012.

[8] V. Guigues and C. Sagastizábal. Risk-averse feasible policies for large-scale multistage stochastic linear programs. Mathematical Programming, 138:167–198, 2013.

[9] R. Henrion and C. Strugarek. Convexity of chance constraints with independent random variables. Computational Optimization and Applications, 41:263–276, 2008.

[10] S. Kataoka. A stochastic programming model. Econometrica, 31:181–196, 1963.

[11] M. Ono, M. Pavone, Y. Kuwata, and J. Balaram. Chance-constrained dynamic programming with application to risk-aware robotic space exploration. Autonomous Robots, 39:555–571, 2015.

[12] A. Prékopa. Programming under probabilistic constraints with a random technology matrix. Mathematische Opemtionsforschung und Statistik, Series Optimization, 5:109–116, 1974.

[13] A. Prékopa. Stochastic Programming. Kluwer, Dordrecht, 1995.

[14] A. Prékopa and T. Szántai. Flood control reservoir system design using stochastic programming. Mathematical Programming Study, 9:138–151, 1978.

[15] C. Revelle, E. Joeres, and W. Kirby. The linear decision rule in reservoir management and design. 1, development of the stochastic model. Water Resources Research, 5(4):767–777, 1969.

[16] J. O. Royset and E. Polak. Extensions of stochastic optimization results to problems with system failure probability functions. Journal of Optimization Theory and Applications, 133:1–18, 2007.

[17] P. Tøndel, T. A. Johansen, and A. Bemporad. An algorithm for multi-parametric quadratic programming and explicit mpc solutions. Automatica, 39:489497, 2003.

[18] W. van Ackooij and R. Henrion. Gradient formulae for nonlinear probabilistic constraints with gaussian and gaussian-like distributions. SIAM Journal on Optimization, 24:1864–1889, 2014.

[19] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati. On probabilistic constraints induced by rectangular sets and multivariate normal distributions. Mathematical Methods of Operations Research, 71:535–549, 2010.

[20] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati. On joint probabilistic constraints with Gaussian coefficient matrix. Operations Research Letters, 39:99–102, 2011.

[21] C. van de Panne and W. Popp. Minimum-cost cattle feed under probabilistic protein constraints. Managment Science, 9:405–430, 1963.

5 Appendix

Lemma 5.1 *Let ξ_t satisfy* (21) *and for any positive integers t, j, let $I_{t,j}, J_{t,j}$, and $H_{t,j}$ be the sets given by*

$$\begin{aligned} I_{t,j}(m) &= \{k \in \mathbb{N} \,:\, 1 \le k \le \min(p_{t+1}(m), t+1-j)\}, \\ J_{t,j}(m) &= \{k \in \mathbb{N} \,:\, 1 \le k \le \min(t, p_{t+1}(m)),\ j \le r_{t+1-k}(m)\}, \\ H_{t,j}(m) &= \{k \in \mathbb{N} \,:\, 1 \le k \le \min(t, p_{t+1}(m)),\ j \le s_{t+1-k}(m)\}. \end{aligned}$$

We also define

$$\begin{aligned} X_t(m) &= \max\left(r_{t+1-k}(m), k = 1, \ldots, \min(t, p_{t+1}(m))\right), \\ Y_t(m) &= \max\left(s_{t+1-k}(m), k = 1, \ldots, \min(t, p_{t+1}(m))\right). \end{aligned}$$

The coefficients c, γ, δ, and θ in the decompositions of $\xi_1(m), \xi_2(m), \ldots, \xi_T(m), m = 1, \ldots, M$, of the form (22) *are computed iteratively as follows:*

Initialization: *For $m = 1, \ldots, M$, set $c_1(m) = \frac{\mu_1(m)}{\alpha_{1,0}(m)}$, $r_1(m) = p_1(m)$, $\gamma_{1,k}(m) = -\frac{\alpha_{1,k}(m)}{\alpha_{1,0}(m)}, k = 1, \ldots, p_1(m)$, $s_1(m) = q_1(m)$, $\delta_{1,k}(m) = \frac{\beta_{1,k}(m)}{\alpha_{1,0}(m)}, k = 1, \ldots, q_1(m)$, and $\theta_{1,1}(m) = \frac{\beta_{1,0}(m)}{\alpha_{1,0}(m)}$.*

Loop: *For $m = 1, \ldots, M$ and for $t = 1, \ldots, T-1$,*

$$c_{t+1}(m) = \frac{\mu_{t+1}(m)}{\alpha_{t+1,0}(m)} - \sum_{k=1}^{\min(t, p_{t+1}(m))} \frac{\alpha_{t+1,k}(m)}{\alpha_{t+1,0}(m)} c_{t+1-k}(m).$$

$$\theta_{t+1,j}(m) = \begin{cases} \frac{\beta_{t+1,0}(m)}{\alpha_{t+1,0}(m)} \ \underline{for}\ j = t+1, \\ \frac{\beta_{t+1,t+1-j}(m)}{\alpha_{t+1,0}(m)} - \sum\limits_{k \in I_{t,j}(m)} \frac{\alpha_{t+1,k}(m)}{\alpha_{t+1,0}(m)} \theta_{t+1-k,j}(m) \ \underline{for}\ t+1-\min(t, q_{t+1}(m)) \le j \le t, \\ -\sum\limits_{k \in I_{t,j}(m)} \frac{\alpha_{t+1,k}(m)}{\alpha_{t+1,0}(m)} \theta_{t+1-k,j}(m) \ \underline{for}\ 1 \le j \le t - \min(t, q_{t+1}(m)). \end{cases}$$

Coefficient $\gamma_{t+1,j}(m)$ is given by

$$\begin{cases} -\frac{\alpha_{t+1,j+t}(m)}{\alpha_{t+1,0}(m)} - \sum_{k\in J_{t,j}(m)} \frac{\alpha_{t+1,k}(m)}{\alpha_{t+1,0}(m)}\gamma_{t+1-k,j}(m) \ \underline{for}\ 1 \le j \le \min(p_{t+1}(m)-t, X_t(m)), \\ -\frac{\alpha_{t+1,t+j}(m)}{\alpha_{t+1,0}(m)} \ \underline{for}\ 1+\min(p_{t+1}(m)-t, X_t(m)) \le j \le p_{t+1}(m)-t, \\ -\sum_{k\in J_{t,j}(m)} \frac{\alpha_{t+1,k}(m)}{\alpha_{t+1,0}(m)}\gamma_{t+1-k,j}(m) \ \underline{for}\ \max(1, 1+\min(p_{t+1}(m)-t, X_t(m))) \le j \le X_t(m). \end{cases}$$

Coefficient $\delta_{t+1,j}(m)$ is given by

$$\begin{cases} \frac{\beta_{t+1,j+t}(m)}{\alpha_{t+1,0}(m)} - \sum_{k\in H_{t,j}(m)} \frac{\alpha_{t+1,k}(m)}{\alpha_{t+1,0}(m)}\delta_{t+1-k,j}(m) \ \underline{for}\ 1 \le j \le \min(q_{t+1}(m)-t, Y_t(m)), \\ \frac{\beta_{t+1,t+j}(m)}{\alpha_{t+1,0}(m)} \ \underline{for}\ 1+\min(q_{t+1}(m)-t, Y_t(m)) \le j \le q_{t+1}(m)-t, \\ -\sum_{k\in H_{t,j}(m)} \frac{\alpha_{t+1,k}(m)}{\alpha_{t+1,0}(m)}\delta_{t+1-k,j}(m) \ \underline{for}\ \max(1, 1+\min(q_{t+1}(m)-t, Y_t(m))) \le j \le Y_t(m). \end{cases}$$

Finally,

$$r_{t+1}(m) = \max(p_{t+1}(m)-t, X_t(m)) \ \ \textit{and}\ s_{t+1}(m) = \max(q_{t+1}(m)-t, Y_t(m)).$$

Proof. We fix a component m and to alleviate notation, we drop (m) in the proof. The initialization is immediate, writing (21) for $t=1$. Now assume that for some $t<T$, the decompositions of $\xi_1,\dots,\xi_t$ of the form (22) are available. To obtain the decomposition of ξ_{t+1}, we use (21) to obtain

$$\begin{aligned} \xi_{t+1} &= \frac{\mu_{t+1}}{\alpha_{t+1,0}} - \sum_{k=1}^{\min(t,p_{t+1})} \frac{\alpha_{t+1,k}}{\alpha_{t+1,0}}\xi_{t+1-k} + \sum_{k=0}^{\min(t,q_{t+1})} \frac{\beta_{t+1,k}}{\alpha_{t+1,0}}\varepsilon_{t+1-k} \\ &\quad - \sum_{k=1+\min(t,p_{t+1})}^{p_{t+1}} \frac{\alpha_{t+1,k}}{\alpha_{t+1,0}}\xi_{t+1-k} + \sum_{k=1+\min(t,q_{t+1})}^{q_{t+1}} \frac{\beta_{t+1,k}}{\alpha_{t+1,0}}\varepsilon_{t+1-k}. \end{aligned}$$

In the first sum, since for all index $k\in\{1,2,\dots,\min(t,p_{t+1})\}$ we have $1\le t+1-k\le t$, we know for ξ_{t+1-k} a decomposition of the form (22) with known coefficients c,γ,δ, and θ. Using these expressions of ξ_{t+1-k}, this first sum can be written

$$-\sum_{k=1}^{\min(t,p_{t+1})} \frac{\alpha_{t+1,k}}{\alpha_{t+1,0}} \left(c_{t+1-k} + \sum_{j=1}^{r_{t+1-k}} \gamma_{t+1-k,j}\xi_{1-j} + \sum_{j=1}^{s_{t+1-k}} \delta_{t+1-k,j}\varepsilon_{1-j} + \sum_{j=1}^{t+1-k} \theta_{t+1-k,j}\varepsilon_j \right).$$

Gathering the terms that depend neither on noise ε nor on ξ, we obtain the expression of c_{t+1}.

The portion depending on $\varepsilon_1,\dots,\varepsilon_{t+1}$ can be written

$$\sum_{j=t+1-\min(t,q_{t+1})}^{t+1} \frac{\beta_{t+1,t+1-j}}{\alpha_{t+1,0}}\varepsilon_j - \sum_{j=1}^{t} \left(\sum_{k\in I_{t,j}} \frac{\alpha_{t+1,k}}{\alpha_{t+1,0}}\theta_{t+1-k,j} \right) \varepsilon_j.$$

We then consider the decomposition of ξ_{t+1} obtained replacing t by $t+1$ in (22). Identifying the portion of this decomposition depending on $\varepsilon_1, \ldots, \varepsilon_{t+1}$ with the expression above, we obtain the expressions of the coefficients $\theta_{t+1,j},\ j = 1, \ldots, t+1$.

The portion that depends on $\xi_0, \xi_{-1}, \ldots,$ can be written

$$-\sum_{j=1+\min(0,p_{t+1}-t)}^{p_{t+1}-t} \frac{\alpha_{t+1,t+j}}{\alpha_{t+1,0}} \xi_{1-j} - \sum_{j=1}^{X_t} \left(\sum_{k \in J_{t,j}} \frac{\alpha_{t+1,k}}{\alpha_{t+1,0}} \gamma_{t+1-k,j} \right) \xi_{1-j}.$$

From that expression, we obtain the desired value of r_{t+1} as well as the announced formulas for coefficients $\gamma_{t+1j}, j = 1, \ldots, r_{t+1}$.

Finally, the portion depending on $\varepsilon_0, \varepsilon_{-1}, \ldots,$ can be written

$$\sum_{j=1+\min(0,q_{t+1}-t)}^{q_{t+1}-t} \frac{\beta_{t+1,t+j}}{\alpha_{t+1,0}} \varepsilon_{1-j} - \sum_{j=1}^{Y_t} \left(\sum_{k \in H_{t,j}} \frac{\alpha_{t+1,k}}{\alpha_{t+1,0}} \delta_{t+1-k,j} \right) \varepsilon_{1-j}.$$

From that expression, we obtain the desired value of s_{t+1} as well as the announced formulas for coefficients $\delta_{t+1,j}, j = 1, \ldots, s_{t+1}$. ■