EXPLOITING STRUCTURE OF AUTOREGRESSIVE PROCESSES IN CHANCE-CONSTRAINED MULTISTAGE STOCHASTIC LINEAR PROGRAMS

VINCENT GUIGUES AND CLAUDIA SAGASTIZÁBAL

Abstract. We consider an interstage dependent stochastic process whose components follow an autoregressive model with time varying order. At a given time, we give some recursive formulæ linking future values of the process with past values and noises. We then consider multistage stochastic linear programs with uncertain sets depending affinely on such processes. At each stage, dealing with uncertainty using probabilistic constraints, the recursive relations can be used to obtain explicit expressions for the feasible set.

Keywords: Stochastic processes, Generalized autoregressive models, Risk-averse optimization.

AMS subject classifications: 90C15, 91B30.

1. Introduction

In many real-life problems, the dynamics of uncertainty is modelled as a time series, reflecting the fact that observations close together in time will be more closely related than observations further apart. For instance, such is the case in mathematical finance for closing values of some stock market index. If, in addition, observations relate to geographical locations, uncertainty also exhibits some kind of spatial dependence. This situation arises when modelling the annual flow volume of different rivers in a given hydrological basin. Autoregressive processes with Gaussian noises are popular models to forecast uncertainty that is time and spatial dependent. Essentially, these models correspond to a multivariate discrete time stochastic process depending in an affine manner on previous values, and with nondiagonal covariance matrices. When controlling the evolution of a system with dynamics depending on this type of uncertainty, it is often desirable to limit the effects of bad outcomes, by means of some *risk measure*. Because of the time dependence, statistical moments (used to control risk) need to be conditioned to the past history of realizations. For this reason, it is crucial to have expressions relating future values with past ones, by means of a recursive application of the autoregressive model.

In this work, which is a companion paper of [1], we give explicit or recursive expressions for expressing uncertainty at a given time $t+i$ as a function of information available at time t. We then recall from [1] how these expressions can be used to make tractable a multistage stochastic program with random variables that are stagewise dependent, by means of a rolling-horizon implementation.

Our paper is organized as follows. The multistage stochastic program constraints are detailed in Section 2. In Section 3, we provide algorithms to decompose future values of the underlying stochastic process as a function of past realizations and noises. Finally, in Section 4, we explain how to use these results to build risk-averse policies using a rolling-horizon approach.

We adopt the following notation and conventions. For $t_2 \geq t_1$, the short form $v_{(t_1,t_2]}$ (resp., $v_{[t_1,t_2]}$) stands for the concatenation $(v_{t_1+1},\ldots,v_{t_2})$ (resp., $(v_{t_1},\ldots,v_{t_2})$), with $v_{(t,t]}$ vacuous and knowing that the concatenated objects v_j can be vectors or matrices, depending on the context. For sums and products, $\sum_{i=i_0}^{i_1} x_i = 0$ and $\prod_{i=i_0}^{i_1} x_i = 1$ whenever $i_0 > i_1$, knowing that for matrices X_i, if $i_0 > i_1$ then $\prod_{i=i_0}^{i_1} X_i = I$, the identity matrix. For a random variable ξ, $\tilde{\xi}$ denotes a particular realization, whereas $\mathbb{E}(\xi)$ and $\sigma(\xi)$ are the expected value and the standard deviation, respectively. Conditional expectations and probabilities are denoted by $\mathbb{E}(\xi_1|\tilde{\xi}_2) := \mathbb{E}(\xi_1|\xi_2 = \tilde{\xi}_2)$ and $\mathbb{P}(\xi_1 \in A|\tilde{\xi}_2) := \mathbb{P}(\xi_1 \in A|\xi_2 = \tilde{\xi}_2)$. The cumulative distribution function of the Gaussian random variable with mean 0 and standard deviation one is denoted by $F(\cdot)$. For process ξ, $\mathcal{F}_t$

is the sigma-algebra $\mathcal{F}_t := \sigma(\xi_j, j \le t)$ and Ω is a sample space equipped with sigma algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_T$ where T is given. Finally, for a continuous random variable X for which higher values are preferred, the Conditional Value-at-Risk of level $\varepsilon_{\tt p} \in (0,1)$ of X is defined by $CVaR_{\varepsilon_{\tt p}}(X) = -\mathbb{E}[X|X \le F_X^{-1}(\varepsilon_{\tt p})]$ while the Value-at-Risk of level $\varepsilon_{\tt p}$ of X is $VaR_{\varepsilon_{\tt p}}(X) = -F_X^{-1}(\varepsilon_{\tt p})$.

2. Motivation and general setting

Suppose we want to control over T stages the evolution of a dynamic system with state variable x_t and transition equation depending affinely on an interstage dependent stochastic process. Specifically, consider that

- uncertainty ξ_t is an M-dimensional random process. Each process component $\xi_t(m)$ follows a generalized autoregressive model with time varying order (each value of the process is an affine function of previous values plus a Gaussian noise); see Section 3 below. The realization $\tilde{\xi}_t$ becomes known at the beginning of time step t;
- for some $p \ge 2$, $x_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \Re^{N_x})$ is the state of the system at the end of time step t given a known x_0, with dynamics given by (1); and
- $u_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \Re^{N_u})$ is the control variable, applied to the system at time step t.

Then, given a vector $d_t \in \Re^{N_x}$, and matrices A_t, B_t, C_t, of respective orders $N_x \times N_x$, $N_x \times N_u$, $N_x \times M$, the state transition equation has the form

$$x_t = A_{t-1}x_{t-1} + B_t u_t + C_t \xi_t + d_t \,. \tag{1}$$

The feasible controls satisfy the scalar inequality

$$E_t x_t + \mathring{F}_t u_t \ge G_t \xi_t + h_t \,, \tag{2}$$

where $h_t \in \Re$, and the matrices E_t, $\mathring{F}_t$, and G_t have orders $1 \times N_x$, $1 \times N_u$, and $1 \times M$, respectively.

An example of a dynamic system governed by relation (1) is given by a reservoir endowed with a power plant. In this case, the state of the system is the volume of the reservoir while uncertainty corresponds to the incoming streamflow. Controls are the spilled and turbined water used to generate power to satisfy the demand of electricity, represented by h_t in (2). For this application, some constraints involve only the state x_t or only the control u_t. In this case, when certain rows in E_t are nonzero, the corresponding rows in $\mathring{F}_t$ are null, and reciprocally. To ease the presentation, we consider here only one scalar relation (2), knowing that the approach can be generalized to vectorial constraints, as in [1].

In order to handle uncertainty with some degree of risk aversion, at time step t some future constraints can be required to be satisfied in a probabilistic manner. For example, for some future time $\tau \in \{t+1, \ldots, T\}$, one could impose satisfaction of (2), written with t replaced by τ, with a sufficiently high probability, say $1 - \varepsilon_{\tt p}$:

$$\mathbb{P}\Big(E_\tau x_\tau(x_t, u_{(t:\tau]}, \xi_{(t:\tau]}) + \mathring{F}_\tau u_\tau \ge G_\tau \xi_\tau + h_\tau\Big) \ge 1 - \varepsilon_{\tt p} \,, \tag{3}$$

noting that controls are considered as variables of "here-and-now" type.

In the above constraint in variables $x_t, u_{(t:\tau]}$, the term $x_\tau(x_t, u_{(t:\tau]}, \xi_{(t:\tau]})$ represents the expression of x_τ as a function of variables $x_t, u_{(t:\tau]}$, and of random vectors $\xi_{(t:\tau]}$. This expression is obtained by applying recursively transition equation (1), between time steps $t+1$ and τ, as explained below.

For the hydro-reservoir example, control constraints such as $u_t \ge 0$, stating that the turbined outflow cannot be negative, need to be satisfied almost surely. By contrast, some operational lower bounds on the state, called "min-zone" and having the form $x_t \ge x_t^{\min}$ for some parameter $x_t^{\min} \ge 0$, need to be satisfied with high probability only. In general, if in (2) there is more than one random constraint, we set individual chance constraints for each one of them (for joint chance constraints we refer to [6]).

Since the chance constraint is set at time step t, it is natural to *condition* it to all the information available at time t. That is, not only the history of realizations $\tilde{\xi}_1, \ldots, \tilde{\xi}_t$ is known, but also the fact that the state variable evolves according to (1) until time τ.

The amount of available information is made explicit by writing each future state x_τ as a function of

- the current state x_t, and
- future controls and random vectors, from time step $t+1$ to time step τ, denoted by $u_{(t,\tau]} := (u_{t+1}, \dots, u_\tau)$ and $\xi_{(t,\tau]} := (\xi_{t+1}, \dots, \xi_\tau)$, respectively.

This is done by applying recursively (1), yielding the affine relation $x_\tau = x_\tau(x_t,\ u_{(t,\tau]}, \xi_{(t,\tau]})$ appearing in the probabilistic constraint above. In turn, this affine expression for x_τ is used to define the Gaussian random variable $X = X(x_t, u_{(t,\tau]}, \xi_{(t,\tau]})$, appearing in the chance constraint:

$$X := E_\tau x_\tau + \mathring{F}_\tau u_\tau - G_\tau \xi_\tau \,. \tag{4}$$

Therefore, since X is a Gaussian random variable, we obtain (see for instance [4])

$$\mathbb{P}\left(X \geq h_\tau \middle| \tilde{\xi}_{[t]}\right) \geq 1 - \varepsilon_{\mathtt{p}} \quad \Longleftrightarrow \quad \mathbb{E}[X|\tilde{\xi}_{[t]}] \geq h_\tau + F^{-1}(1-\varepsilon_{\mathtt{p}})\sigma(X|\tilde{\xi}_{[t]})\,, \tag{5}$$

and the chance constraint can be rewritten in terms of the conditional mean and standard deviation of the random variable X.

In [1, Sec. 4] it is shown that the conditional random variable $X|\tilde{\xi}_{[t]}$ is an affine function of the relevant future uncertainty, $\xi_{(t,\tau]}$. It is also shown that the parameters in the combination depend affinely on $(x_t, u_{(t,\tau]})$. So in (4) the dependence of X on $(x_t, u_{(t,\tau]}, \xi_{(t,\tau]})$ is also affine. In a manner similar, both the conditional mean and standard deviation of X involve affine combinations of the expectations of random variables $\mathbb{E}[\xi_{t+j}|\tilde{\xi}_{[t]}]$, for $j = 1, \dots, \tau - t$. As a result, an explicit reformulation of chance constraint (5) depends on the ability to express random variables $\xi_{t+j}|\tilde{\xi}_{[t]}$ as a function of $\tilde{\xi}_{[t]}$ and future noises, for $j = 1, \dots, \tau - t$.

The next section shows how such calculations can be done in an explicit or recursive manner, depending on the parameters defining the autoregressive process.

3. Recursive relations for the statistical model

We consider a multivariate discrete time stochastic process depending in an affine manner on previous values. More precisely, for $m = 1, \dots, M$, each component $\xi_t(m)$ is represented by a generalized autoregressive model, with varying orders $p_t(m) \geq 0$. Accordingly, for every integer t, there exist coefficients $\Phi_t^i(m)$ for $i = 1, \dots, p_t(m)$, with non-null $\Phi_t^{p_t(m)}(m)$, such that

$$\xi_t(m) = \sum_{i=1}^{p_t(m)} \Phi_t^i(m)\xi_{t-i}(m) + \eta_t(m). \tag{6}$$

We use the terminology *generalized autoregressive*, instead of autoregressive, to emphasize the fact that orders $p_t(m)$ depend on time and could be null. Indeed, autoregressive processes (see [3] for instance) correspond to the particular case of a nonzero order, constant in time.

In the expression above, noises $(\eta_1, \dots, \eta_T)$, are independent Gaussian vectors with each component having a mean $\mu_t := \mathbb{E}[\eta_t]$ and an $M \times M$ covariance matrix $\Gamma_t := Cov(\eta_t)$.

Let $t = 1, \dots, T-1$ and $j = 1, \dots, T-t$. In order to express a given value of the process, say $\xi_{t+j}(m)$, as a function of its past history, it is convenient to introduce integers

$$p_{t,j}^{\max}(m) = \max_{1 \leq k \leq j} \{p_{t+k}(m) - k\}, \tag{7}$$

as well as the corresponding past history of the process:

$$\tilde{\xi}_{[t]} = \left\{\tilde{\xi}_{t-i}(m),\ m = 1, \dots, M,\ i = 0, \dots, \max(p_{t,T-t}^{\max}(m), t-1)\right\}. \tag{8}$$

The index $p_{t,j}^{\max}(m)$ specifies the minimal amount of past information needed at time step t to express $\xi_{t+j}(m)$ as a function of $\tilde{\xi}_{[t]}$, for a process ξ_t modelled by (6).

Recursive expressions for ξ_{t+j} as a function of $\tilde{\xi}_{[t]}$ can be derived from the model data in (6), using two alternative algorithms described in the next lemmas. In these developments, a component m is fixed and $\xi_t(m)$ is denoted by ξ_t to alleviate notation (the dependence with respect to the m-component is suppressed).

Lemma 3.1. *[First recursive formula] Consider a process ξ_t modeled by (6). Then, for $t = 1,\dots,T-1$, and $j = 1,\dots,T-t$, the relation*

$$\xi_{t+j} = \sum_{\ell=0}^{p_{t,j}^{\max}} \alpha_{t,j}^{\ell} \xi_{t-\ell} + \sum_{\ell=1}^{j} \beta_{t+j}^{\ell}\, \eta_{t+j-\ell+1} \tag{9}$$

holds. The coefficients in this expression can be computed with the following algorithm:

For $t = 1,\dots,T-1$,

  For $m = 1,\dots,M$,

    $\alpha_{t,1}^{\ell} = \Phi_{t+1}^{\ell+1}, \ell = 0,\dots,p_{t,1}^{\max};\ \beta_{t+1}^{1} = 1;$

    For $j = 2,\dots,t$,

      $\beta_{t+1}^{j} = \alpha_{t+2-j,j-1}^{0};$

      For $\ell = 0,\dots,\min(p_{t+2-j,j-1}^{\max}, p_{t+2-j}) - 1$,

$$\alpha_{t+1-j,j}^{\ell} = \alpha_{t+2-j,j-1}^{\ell+1} + \alpha_{t+2-j,j-1}^{0} \Phi_{t+2-j}^{\ell+1},$$

      End For

      For $\ell = p_{t+2-j},\dots,p_{t+2-j,j-1}^{\max} - 1$,

$$\alpha_{t+1-j,j}^{\ell} = \alpha_{t+2-j,j-1}^{\ell+1},$$

      End For

      For $\ell = p_{t+2-j,j-1}^{\max},\dots,p_{t+2-j} - 1$,

$$\alpha_{t+1-j,j}^{\ell} = \alpha_{t+2-j,j-1}^{0} \Phi_{t+2-j}^{\ell+1},$$

      End For

    End For

  End For

End For

Proof. Take fixed $t \in \{1,\dots,T-1\}$. In these recursive formulæ, calculations are done in a manner that ξ_{t+1} can progressively be written as a function of $\xi_{[t]}$, of $\xi_{[t-1]},\dots,\xi_{[t-j-1]}$ and, finally, of $\xi_{[1]}$. To prove the statement, we check the validity of coefficient expressions for all possible values of j. For $j = 0$, (6) gives the following relation between ξ_{t+1} and $\xi_{[t]}$: $\xi_{t+1} = \sum_{\ell=0}^{p_{t+1}-1} \Phi_{t+1}^{\ell+1} \xi_{t-\ell} + \eta_{t+1}$. Since, by (7), the identity $p_{t+1} - 1 = p_{t,1}^{\max}$ holds, the initial values $\alpha_{t,1}^{\ell} = \Phi_{t+1}^{\ell+1}$ for $\ell = 0,\dots,p_{t,1}^{\max}$ and $\beta_{t+1}^{1} = 1$ coincide with those starting the recursion.
Now suppose that the expression for ξ_{t+1} as a function of $\xi_{[t+2-j]}$ for some $j \in \{2,\dots,t\}$ is available. That is to say, there are known coefficients $\alpha_{t+2-j,j-1}^{\ell},\ \ell = 0,\dots,p_{t+2-j,j-1}^{\max}$ and $\beta_{t+1}^{\ell},\ \ell = 1,\dots,j-1$, satisfying the identity

$$\xi_{t+1} = \sum_{\ell=0}^{p_{t+2-j,j-1}^{\max}} \alpha_{t+2-j,j-1}^{\ell}\, \xi_{t+2-j-\ell} + \sum_{\ell=1}^{j-1} \beta_{t+1}^{\ell}\, \eta_{t+2-\ell}. \tag{10}$$

We want to express ξ_{t+1} as a function of $\xi_{[t+1-j]}$. We start by replacing ξ_{t+2-j} in (10) with its corresponding expression from (6): $\sum_{\ell=0}^{p_{t+2-j}-1} \Phi_{t+2-j}^{\ell+1} \xi_{t+1-j-\ell} + \eta_{t+2-j}$. Then, we consider separately the right-hand side terms with ξ and η and write $\xi_{t+1} = \xi_{t+1}^{a} + \xi_{t+1}^{b}$, for

$$\xi_{t+1}^{a} := \sum_{\ell=0}^{p_{t+2-j,j-1}^{\max}-1} \alpha_{t+2-j,j-1}^{\ell+1}\, \xi_{t+1-j-\ell} + \sum_{\ell=0}^{p_{t+2-j}-1} \alpha_{t+2-j,j-1}^{0}\, \Phi_{t+2-j}^{\ell+1} \xi_{t+1-j-\ell},$$

$$\xi_{t+1}^{b} := \sum_{\ell=1}^{j-1} \beta_{t+1}^{\ell}\, \eta_{t+2-\ell} + \alpha_{t+2-j,j-1}^{0} \eta_{t+2-j}.$$

The desired formulæ for computing β_{t+1}^{j} and $\alpha_{t+1-j,j}^{\ell}$ for $\ell = 0,\dots,p_{t+1-j,j}^{\max}$ follow from identifying the terms above with the corresponding ones in (9). □

The affine relation in (9) can be expressed in a matrix-vector form, thus simplifying the proof of Lemma 3.1. However, in this case the involved matrices and vectors would have many null elements carrying no information at all. This is an important matter for practitioners, since from an implementation point of view, the matrix formulation considerably increases the computational bulk, especially regarding use of memory. This is the reason why we chose the more involved form (9), which is the most economical one because it uses the minimal amount of past information, via the integers $p_{t,j}^{\max}$ from (7). The explicit knowledge of coefficients α and β in (9) can be plugged in the formulas for $\mathbb{E}[X|\tilde{\xi}_{[t]}]$ and $\sigma(X|\tilde{\xi}_{[t]})$ given in Lemmas 4.1 and 4.2 of [1]. The chance constraint (5) can then be rewritten as a deterministic linear constraint. At step t, the algorithm in Lemma 3.1 computes coefficients $\alpha_{t,1}^{\ell}(m)$, $\alpha_{t-1,2}^{\ell}(m)$, …, $\alpha_{1,t}^{\ell}(m)$ as well as coefficients $\beta_{t+1}^{\ell}(m)$ for all possible values of m, ℓ. These coefficients give decompositions of ξ_{t+1} on $\xi_{[t]}$, then on $\xi_{[t-1]}$, …, and finally on $\xi_{[1]}$. For some calculations, for instance those in Lemma 3.3 or to implement the rolling-horizon approach from Section 4, it may be preferable instead to compute at once all the coefficients for each time step. These calculations can be done by the alternative recursive algorithm below.

Lemma 3.2. *[Alternative recursive formula] Consider a process ξ_t modeled by* (6)*. The coefficients in* (9) *can be computed with the following algorithm:*
For $t = 1, \dots, T-1$,
set the initial conditions $\beta_{t+1}^{1} = 1$ *and* $\alpha_{t,1}^{\ell} = \Phi_{t+1}^{\ell+1}$, *for* $\ell = 0, \dots, p_{t,1}^{\max}$.
If $t \geq 2$, *for* $\ell = 2, \dots, t$,

$$\beta_{t+1}^{\ell} = \sum_{k=0}^{\min(\ell-2, p_{t+1}-1)} \Phi_{t+1}^{k+1} \beta_{t-k}^{\ell-k-1}. \tag{11}$$

End If
If $t \leq T-2$, *for* $j = 1, \dots, T-t-1$,
Let $L_{t,j} = \min(p_{t,j}^{\max}, \ p_{t+j+1} - (j+1))$ *and let* $I_{t,j}^{\ell}$ *be the set*

$$I_{t,j}^{\ell} = \{k : 0 \leq k \leq \min(p_{t+j+1}, j) - 1 \text{ and } \ell \leq p_{t,j-k}^{\max}\}.$$

Let $J_1(t) \cup J_2(t) \cup J_3(t)$ *be the partition of the set* $\{1, \dots, T-t-1\}$ *given by*

$$\begin{aligned} J_1(t) &= \Big\{ j \in \{1, \dots, T-t-1\} : p_{t+j+1} \leq j \Big\}, \\ J_2(t) &= \Big\{ j \in \{1, \dots, T-t-1\} : j+1 \leq p_{t+j+1} \leq j+1+p_{t,j}^{\max} \Big\}, \text{ and} \\ J_3(t) &= \Big\{ j \in \{1, \dots, T-t-1\} : p_{t+j+1} > j+1+p_{t,j}^{\max} \Big\}. \end{aligned}$$

Then for all $j \in J_2(t) \cup J_3(t)$,

$$\alpha_{t,j+1}^{\ell} = \Phi_{t+j+1}^{j+\ell+1} + \sum_{k \in I_{t,j}^{\ell}} \Phi_{t+j+1}^{k+1} \alpha_{t,j-k}^{\ell} \text{ for } \ell = 0, \dots, L_{t,j}, \tag{12}$$

whereas

$$\alpha_{t,j+1}^{\ell} = \sum_{k \in I_{t,j}^{\ell}} \Phi_{t+j+1}^{k+1} \alpha_{t,j-k}^{\ell} \text{ if } \begin{array}{l} (j \in J_1(t) \text{ and } \ell = 0, \dots, p_{t,j+1}^{\max}) \text{ or} \\ (j \in J_2(t) \text{ and } \ell = L_{t,j}+1, \dots, p_{t,j+1}^{\max}). \end{array} \tag{13}$$

Finally, if $j \in J_3(t)$ *and* $\ell = L_{t,j}+1, \dots, p_{t,j+1}^{\max}$, *then*

$$\alpha_{t,j+1}^{\ell} = \Phi_{t+j+1}^{j+\ell+1} \text{ if } j \in J_3(t) \text{ and } \ell = L_{t,j}+1, \dots, p_{t,j+1}^{\max}. \tag{14}$$

End If
End For

Proof. We first check the computation of α-coefficients by fixing t such that $t \in \{1, \dots, T-1\}$. The initial conditions provide coefficients α in the decomposition of ξ_{t+1} on $\xi_{[t]}$. We claim that the inner loop on $j \in \{1, \dots, T-t-1\}$ provides successively the α-values that appear in the decomposition of ξ_{t+2} on $\xi_{[t]}$, then of ξ_{t+3} on $\xi_{[t]}$, …, and finally of ξ_T on $\xi_{[t]}$. To prove our claim, we proceed by induction on j, assuming that for some $j \in \{1, \dots, T-t-1\}$, the coefficients α

in the decompositions of ξ_{t+k} on $\xi_{[t]}$ for $k = 1, \ldots, j$, are available. These coefficients are $\alpha^{\ell}_{t,k}$ for $k = 1, \ldots, j,\ \ell = 0, \ldots, p^{\max}_{t,k}$. To show that at iteration j of the loop in j, the algorithm computes the coefficients α necessary for writing the decomposition of ξ_{t+j+1} on $\xi_{[t]}$, we first write, using model (6),

$$\begin{aligned} \xi_{t+j+1} &= \sum_{k=0}^{p_{t+j+1}-1} \Phi^{k+1}_{t+j+1}\xi_{t+j-k} + \eta_{t+j+1} \\ &= \sum_{k=0}^{\min(p_{t+j+1}-1, j-1)} \Phi^{k+1}_{t+j+1}\xi_{t+j-k} + \sum_{k=\min(p_{t+j+1}, j)}^{p_{t+j+1}-1} \Phi^{k+1}_{t+j+1}\xi_{t+j-k} + \eta_{t+j+1}. \end{aligned} \tag{15}$$

Next, observe that for $0 \le k \le \min(p_{t+j+1}-1, j-1)$ we have $t+1 \le t+j-k \le t+j$. As a result, in relation (15), we have, for each ξ_{t+j-k} appearing in the first sum, a decomposition of the form (9). Therefore

$$\begin{aligned} \xi_{t+j+1} = &\sum_{k=0}^{\min(p_{t+j+1}-1, j-1)} \Phi^{k+1}_{t+j+1} \left(\sum_{\ell=0}^{p^{\max}_{t,j-k}} \alpha^{\ell}_{t,j-k}\xi_{t-\ell} + \sum_{\ell=1}^{j-k} \beta^{\ell}_{t+j-k}\, \eta_{t+j-k-\ell+1} \right) \\ &+ \sum_{k=\min(p_{t+j+1}, j)}^{p_{t+j+1}-1} \Phi^{k+1}_{t+j+1}\xi_{t+j-k} + \eta_{t+j+1}. \end{aligned} \tag{16}$$

Since $p^{\max}_{t,j+1} = \max(p^{\max}_{t,j}, p_{t+j+1} - (j+1))$, the sequence $(p^{\max}_{t,j})_{j \ge 1}$ is non-decreasing and, hence, $\max\Big(p^{\max}_{t,j-k},\ k = 0, \ldots, \min(p_{t+j+1}-1, j-1)\Big) = p^{\max}_{t,j}$. It follows that the terms depending on ξ in the right-hand side of (16) can be written

$$\sum_{\ell=0}^{p^{\max}_{t,j}} \left(\sum_{k \in I^{\ell}_{t,j}} \Phi^{k+1}_{t+j+1}\alpha^{\ell}_{t,j-k} \right) \xi_{t-\ell} + \sum_{\ell=\min(p_{t+j+1}-j, 0)}^{p_{t+j+1}-j-1} \Phi^{j+\ell+1}_{t+j+1}\xi_{t-\ell}. \tag{17}$$

We now consider three cases, depending on whether j belongs to $J_1(t), J_2(t)$, or $J_3(t)$.

a) If $j \in J_1(t)$, then $p^{\max}_{t,j+1} = p^{\max}_{t,j}$ and the second summation in (17) vanishes. Therefore, by identifying the above expression with (9), we obtain the first part in (13).
b) If $j \in J_2(t)$, then the second part in (13) and (12) follow from (17) and the fact that $p^{\max}_{t,j+1} = p^{\max}_{t,j}$.
c) If $j \in J_3(t)$, then (12) and (14) follow from (17) and the fact that $p^{\max}_{t,j+1} = p_{t+j+1} - j - 1$.

We now check the computation of coefficients β. For fixed $t \in \{1, \ldots, T-1\}$, we show by induction that the algorithm computes $\beta^1_{t+1}, \ldots, \beta^t_{t+1}$, i.e., the coefficients appearing in the decomposition of ξ_{t+1} on $\xi_{[1]}$. For $t = 1$, due to (6), we have $\beta^1_2 = 1$, as given by the initial conditions of the algorithm. For some $t \in \{2, \ldots, T-1\}$, suppose that the algorithm has correctly computed $\beta^1_2, \beta^1_3, \beta^2_3, \ldots, \beta^1_t, \ldots, \beta^{t-1}_t$. To see that the formulæ giving β^{ℓ}_{t+1} for $\ell = 1, \ldots, t$ hold, first note that for $t \in \{2, \ldots, T-1\}$, relation (9) with t and j respectively replaced by 1 and $t-1$, gives that

$$\xi_t = \sum_{i=0}^{p^{\max}_{1,t-1}} \alpha^i_{1,t-1}\xi_{1-i} + \sum_{i=1}^{t-1} \beta^i_t \eta_{t-i+1}, \tag{18}$$

where all coefficients β in the expression above have already been computed. Using once again (6), we see that

$$\xi_{t+1} = \sum_{k=0}^{\min(t-2, p_{t+1}-1)} \Phi^{k+1}_{t+1}\xi_{t-k} + \sum_{k=\min(t-1, p_{t+1})}^{p_{t+1}-1} \Phi^{k+1}_{t+1}\xi_{t-k} + \eta_{t+1}. \tag{19}$$

The second sum above is an affine combination of terms in $\xi_{[t]}$. As for the first sum, note that $t-k \ge 2$, because $k \in \{0, \ldots, \min(t-2, p_{t+1}-1)\}$. Therefore, a decomposition of the form (18) with t substituted by $t-k$ is available for ξ_{t-k}. Plugging this decomposition into (19), the terms in the right-hand side of (19) that depend on η are

$$\sum_{k=0}^{\min(t-2,p_{t+1}-1)} \Phi_{t+1}^{k+1}\left(\sum_{i=1}^{t-k-1} \beta_{t-k}^{i}\eta_{t-k-i+1}\right)+\eta_{t+1}.$$

The change of variable $\ell = i+k+1$ then yields the expression

$$\sum_{\ell=2}^{t}\left(\sum_{k=0}^{\min(\ell-2,p_{t+1}-1)} \Phi_{t+1}^{k+1}\beta_{t-k}^{\ell-k-1}\right)\eta_{t-\ell+2}+\eta_{t+1}.$$

Comparing the terms in η in the above expression with those in (18), written with t replaced by $t+1$, we obtain for β_{t+1}^{ℓ} the expression given in the algorithm, i.e., (11) for $\ell \in \{2,\ldots,t\}$ and $\beta_{t+1}^{1}=1$ as initial condition. □

Some particular instances yielding explicit expressions for the coefficients are given below without proof.

Lemma 3.3. *Consider a stochastic process as in* (6). *There are closed-form expressions for coefficients α and β in* (9) *in the following cases.*

(i) *If, for every t, the order p_t is null, then $\xi_{t+j}=\eta_{t+j}$.*

(ii) *If, for every t, the order $p_t = 1$, then*

$$\xi_{t+j}=\left(\prod_{i=1}^{j}\Phi_{t+i}^{1}\right)\xi_t+\sum_{\ell=1}^{j}\left(\prod_{i=j-\ell+1}^{j-1}\Phi_{t+i+1}^{1}\right)\eta_{t+j-\ell+1}\,.$$

(iii) *If, for every m, the orders $p_t = p$ are constant, then α and β are explicit functions of the roots of the characteristic polynomial $P(X)=X^p-\sum_{k=1}^{p}\Phi^k X^{p-k}$.* □

4. Using the autoregressive structure in an optimization setting

We now show how the developments of the previous sections can be used to build risk-averse policies for multistage stochastic linear programs.

Suppose that at any given time stage t, knowing

- the trajectory $\tilde{\xi}_{[t]}=(\tilde{\xi}_1,\ldots,\tilde{\xi}_t)$, of process ξ up to this stage, and
- the state $x_{t-1}(\tilde{\xi}_{[t-1]})$ of the system at the beginning of this stage,

the system volatility is controlled by requiring that future constraints, that is constraints for $\tau = t+1,\ldots,T$, are satisfied with certain probability, as in (3). This chance constraint (3) involves the variable X from (4), that we now write $X_{t,\tau}$ instead of X, to put in evidence the dependence on both t and τ: $X_{t,\tau} := E_\tau x_\tau(x_t, u_{(t,\tau]},\xi_{(t,\tau]})+\mathring{F}_\tau u_\tau - G_\tau\xi_\tau$. With this notation, and using (5), the probabilistic constraint writes down as

$$\mathbb{P}\left(X_{t,\tau}\geq h_\tau\middle|\tilde{\xi}_{[t]}\right)\geq 1-\varepsilon_{\mathtt{p}} \iff \mathbb{E}[X_{t,\tau}|\tilde{\xi}_{[t]}]\geq h_\tau^t$$

where we defined

$$h_\tau^t := h_\tau + F^{-1}(1-\varepsilon_{\mathtt{p}})\sigma(X_{t,\tau}|\tilde{\xi}_{[t]})\,. \tag{20}$$

Since controls u_τ appearing in chance constraint (3) are of the here-and-now type, in view of the definition of $X_{t,\tau}$, its conditional expected value equals

$$E_\tau x_\tau \mathbb{E}[(x_t,u_{(t,\tau]},\xi_{(t,\tau]})]+\mathring{F}_\tau u_\tau - G_\tau\mathbb{E}[\xi_\tau|\tilde{\xi}_{[t]}].$$

Therefore, using the transition equation (1), we see that we can obtain the identity $\bar{x}_\tau := \mathbb{E}[(x_t,u_{(t,\tau]},\xi_{(t,\tau]})]$ through the recursion

$$\begin{array}{rcl}\bar{x}_k &=& A_{k-1}\bar{x}_{k-1}+B_k u_k + C_k\mathbb{E}[\xi_k|\tilde{\xi}_{[t]}]+d_k,\ k=t,t+1,\ldots,\tau, \text{ with}\\ \bar{x}_{t-1} &:=& x_{t-1}(\tilde{\xi}_{[t-1]}) \text{ known at time step } t.\end{array}$$

As a result, controls $(u_t, u_{t+1}, \ldots, u_T)$ are feasible if and only if there exists a state $(\bar{x}_t, \bar{x}_{t+1}, \ldots, \bar{x}_T)$ such that

$$\left\{\begin{array}{l} \bar{x}_t = A_{t-1}x_{t-1}(\tilde{\xi}_{[t-1]}) + B_t u_t + C_t\tilde{\xi}_t + d_t \\ E_t\bar{x}_t + \mathring{F}_t u_t \geq G_t\tilde{\xi}_t + h_t, \\ \text{and for } \tau = t+1, \ldots, T, \ \ (\bar{x}_\tau, u_\tau) \in \mathcal{S}^t_\tau(\bar{x}_{\tau-1}), \end{array}\right.$$

where

$$\mathcal{S}^t_\tau(\bar{x}_{\tau-1}) = \left\{(\bar{x}_\tau, u_\tau) : \begin{array}{l} \bar{x}_\tau = A_{\tau-1}\bar{x}_{\tau-1} + B_\tau u_\tau + C_\tau\mathbb{E}[\xi_\tau|\tilde{\xi}_{[t]}] + d_\tau \\ E_\tau\bar{x}_\tau + \mathring{F}_\tau u_\tau \geq G_\tau\mathbb{E}[\xi_\tau|\tilde{\xi}_{[t]}] + h^t_\tau \end{array}\right\}.$$

The interest of Lemma 3.1 (or Lemma 3.2), lies precisely on the fact that, combined with Lemmas 4.1 and 4.2 in [1], the calculations therein provide an explicit representation for the constraints above.

If controlling the dynamic system defined by (1)-(2) involves a linear cost of the form $c_t^\top u_t$ at time step t, then the corresponding chance-constrained optimization problem has the form

$$\left\{\begin{array}{l} \min\limits_{\bar{x}_{[t:T]}, u_{[t,T]}} \ \sum\limits_{\tau=t}^{T} c_\tau^\top u_\tau \\ \bar{x}_t = A_{t-1}x_{t-1}(\tilde{\xi}_{[t-1]}) + B_t u_t + C_t\tilde{\xi}_t + d_t, \\ E_t\bar{x}_t + \mathring{F}_t u_t \geq G_t\tilde{\xi}_t + h_t, \\ (\bar{x}_\tau, u_\tau) \in \mathcal{S}^t_\tau(\bar{x}_{\tau-1}), \text{ for } \tau = t+1, \ldots, T\,, \end{array}\right. \tag{21}$$

which is a deterministic linear programming problem. As such, (21) can be solved directly if the number of stages is not too high. By contrast, when T is large, it may be preferable to apply a decomposition method. Decomposition may be achieved in a stagewise manner, by writing Dynamic Programming equations for (21), as explained below.

More precisely, because the objective function is separable by time steps, problem (21) is solvable by Dual Dynamic Programming, by introducing cost-to-go functions $\mathcal{Q}^t_{T+1} \equiv 0$ and for $\tau = t+1, \ldots, T$,

$$\mathcal{Q}^t_\tau(\bar{x}_{\tau-1}, \tilde{\xi}_{[t]}) = \min\left\{c_\tau^\top u_\tau + \mathcal{Q}^t_{\tau+1}(\bar{x}_\tau, \tilde{\xi}_{[t]}) : (\bar{x}_\tau, u_\tau) \in \mathcal{S}^t_\tau(\bar{x}_{\tau-1})\right\}. \tag{22}$$

As a result, problem (21) is equivalent to the following linear program:

$$\left\{\begin{array}{l} \min\limits_{\bar{x}_t, u_t} \ c_t^\top u_t + \mathcal{Q}^t_{t+1}(\bar{x}_t, \tilde{\xi}_{[t]}) \\ \bar{x}_t = A_{t-1}x_{t-1}(\tilde{\xi}_{[t-1]}) + B_t u_t + C_t\tilde{\xi}_t + d_t \\ E_t\bar{x}_t + \mathring{F}_t u_t \geq G_t\tilde{\xi}_t + h_t. \end{array}\right. \tag{23}$$

In this problem, the feasible set depends on the initial state x_{t-1}, while the cost-to-go function $\mathcal{Q}^t_{t+1}$ depends on both the state of the system at the end of the stage t and the history of realizations of the process, because ξ is interstage dependent, by (6).

After solving for $t = 1, \ldots, T$ all the chance-constrained problems (21), in a rolling-horizon mode, a risk-averse policy can be built as follows. Having a solution $(\bar{x}^t_{[t:T]}, \bar{u}^t_{[t,T]})$ to (21), the controls $u_{Rob} := (\bar{u}^1_1, \ldots, \bar{u}^t_t, \ldots, \bar{u}^T_T)$, define such risk-averse policy (the remaining optimal controls $u_{t+1}, \ldots, u_T$ of problem (21) are not used). This policy is implementable under the assumption of relatively complete recourse, because, by construction, a solution to the t^{th} chance-constrained optimization problem (21) satisfies the constraints for time step t. The policy is also time consistent, [5]. Finally, nonanticipativity of the policy results from the fact that in (21) the solution depends on $\tilde{\xi}_{[t]}$, but not on future realizations $\tilde{\xi}_{(t,T]}$.

We refer the reader to [2] for a numerical assessment of a related rolling-horizon risk-averse policy for a large-scale stochastic linear program, arising in energy optimization.

Acknowledgments. We are grateful to the Associate Editor and the reviewer for constructive comments. The first author's research was supported by CNPq grant No. 382851/2007-4. The work of the second author was partially supported by grants CNPq 303840/2011-0, AFOSR FA9550-08-1-0370, NSF DMS 0707205, PRONEX-Optimization, and FAPERJ.

References

[1] V. Guigues and C. Sagastizábal. Risk-averse feasible policies for large-scale multistage stochastic linear programs. *Submitted to Math. Program.* Available for download at Optimization Online.

[2] V. Guigues and C. Sagastizábal. The value of rolling horizon policies for risk-averse hydro-thermal planning. *European Journal of Operational Research*, 217:129–140, 2012.

[3] J. Hamilton. *Time Series Analysis.* Princeton University Press, 1994.

[4] A. Prékopa. *Stochastic programming*, volume 324 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995.

[5] A. Shapiro. On a time consistency concept in risk averse multi-stage stochastic programming. *Oper. Res. Letters*, 37:143–147, 2009.

[6] W. Van Ackooij, R. Henrion, A. Möller, and R. Zorgati. On joint probabilistic constraints with Gaussian coefficient matrix. *Oper. Res. Lett.*, 39(2):99–102, 2011.

Vincent Guigues: UFRJ, Escola Politécnica, Departamento de Engenharia Industrial, Ilha do Fundão, Rio de Janeiro, Brazil, Fundação Getulio Vargas, Escola de Matemática Aplicada, praia de Botafogo, Rio de Janeiro, Brazil, IMPA, Instituto de Matemática Pura e Aplicada, 110 Estrada Dona Castorina, Jardim Botanico, Rio de Janeiro, Brazil, vguigues@impa.br and Claudia Sagastizábal: On leave from INRIA Rocquencourt, France, sagastiz@impa.br.