Abstract—We develop a generalized stability framework for stochastic discrete-time systems, where the generality pertains to the ways in which the distribution of the state energy can be characterized. We use tools from finance and operations research called risk functionals (i.e., risk measures) to facilitate diverse distributional characterizations. In contrast, classical stochastic stability notions characterize the state energy on average or in probability, which can obscure the variability of stochastic system behavior. After drawing connections between these risk-aware stability concepts for nonlinear systems, we specialize to linear systems and derive sufficient conditions for the satisfaction of some risk-aware stability properties. These results pertain to real-valued coherent risk functionals and a mean-conditional-variance functional. The results reveal novel noise-to-state stability properties, which assess disturbances in ways that reflect the chosen measure of risk. We illustrate the theory through examples about robustness, parameter choices, and state-feedback controllers.

Index Terms—Stochastic stability theory, risk functionals, stochastic discrete-time systems, linear systems

I. INTRODUCTION

Stability theory of stochastic systems has a long-standing history with contributions from the late 1950s and early 1960s [1]–[3] and is an especially active research area currently [4]–[9]. The applications of stochastic stability theory are numerous, and we list just a few here: trajectory tracking despite modeling errors [10, p. 426], ensuring that state estimation error stays bounded [11], target tracking for partially observable systems [12], and establishing the convergence of sampling-based optimization algorithms [13], [14].

Stochastic stability theory typically assesses system behavior by evaluating the magnitude $|\cdot|$, e.g., Euclidean norm, of the state $x_t$ over time in probability or in expectation. Evaluating the probability of a harmful event, e.g., $|x_t|$ exceeds a desired threshold, provides one approach to analyze stability. To analyze stability from a different perspective, one may evaluate the moments of $|x_t|$. For example, the uniform exponential $\rho$-stability property (informally) means that the expectation of $|x_t|^\rho$ decays exponentially at a rate independent of the initial condition. However, instead of merely analyzing $E(|x_t|^\rho)$, we may wish to analyze additional characteristics of the distribution of the state energy $|x_t|^\rho$ over time. For instance, the state energy may represent parameter estimation error in stochastic gradient descent, state estimation error in a Kalman filter, or trajectory tracking error of a robotic system. Rarer larger realizations of the state energy may arise in practice but may be concealed at the analysis stage by restricting one’s attention to the expectation of $|x_t|^\rho$. There may be incomplete knowledge about the system model, suggesting the importance of analyzing the state energy under distributional ambiguity. Also, the classical $\rho$-stability property ignores the variability of the state energy with respect to its mean, whereas it may be desirable to stabilize such variability in practice.

Ultimately, we wish to design control systems whose energy dynamics enjoy specific desirable distributional characteristics. For instance, we may wish the state to decay in expectation conditioned on a given fraction of the worst cases. In other settings, we may wish the variance of the state energy to decay, or we may wish the expected state energy to decay in a distributionally robust sense. To design such control systems in the future, we require techniques to analyze state energy distributions from more diverse perspectives.

In the literature, the term stability refers to different properties, for example, uniform boundedness of the state trajectory in probability [15, (D4), p. 31], driving the state to the origin within a given time period (i.e., prescribed-time stabilization) [8], and input-to-state and noise-to-state stability concepts. Since Sontag’s pioneering work in the setting of deterministic systems in the late 1980s [16], stochastic input-to-state and noise-to-state stability notions have been developed, e.g., see [6], [7], [17]–[22]. In particular, deterministic input-to-state stability was extended to stochastic noise-to-state stability by Krstić, Deng, and colleagues in the late 1990s and early 2000s [17], [19]. In the current work, we emphasize uniform stability notions of the form

$$\rho(\psi(x_t)) \leq a\lambda^t\psi(x) + b, \quad t \in \{1, 2, \ldots\},$$

where $\rho$ is a map from random variables to the extended real line, $\psi$ is a state energy function, $x \in \mathbb{R}^n$ is an initial condition, and $\lambda \in [0, 1)$, $a \geq 0$, and $b \in \mathbb{R}$ are parameters. If $b$ depends on a disturbance process, e.g., $b = \sup_{t \in [0,1]} E(|w_t|^2) < \infty$, then the above stability notion describes a noise-to-state stability property.

In the current work, we develop a uniform stability framework for the analysis of diverse characteristics of the state energy’s distribution by permitting a fairly general map $\rho$ in (1) called a risk functional. A risk functional (i.e., risk measure) is a map from a space of random variables representing costs to the extended real line [23, p. 261]. Expo-
nential utility is the classical risk functional in the control systems literature pioneered by Whittle in the 1980s and 1990s [24–26]. Other seminal works about exponential utility include those by Howard and Matheson [27] and by Jacobson [28] in the early 1970s. The exponential utility $\rho_{e,\theta}(Z) := \frac{1}{\theta} \log E(\exp(\theta Z))$ of a nonnegative random variable $Z$ with $\theta \neq 0$ is interpreted as a mean-variance approximation, i.e., if the magnitude of $\sqrt{\text{var}}(Z)$ is small enough, then $\rho_{e,\theta}(Z) \approx E(Z) - \frac{\var(Z)}{\theta}$ [24, p. 765]. From a viewpoint based on expected utility theory [29], $\rho_{e,\theta}$ assumes that an exponential function represents the user’s preferences. Such assumptions need not be appropriate for every application, motivating investigations of additional risk-aware criteria.

Building on the above contributions, additional types of risk functionals have been developed [23], [30]–[32]. A real-valued coherent risk functional satisfies four useful properties: convexity, monotonicity, translation equivariance $\rho(Z + a) = \rho(Z) + a$ if $a \in \mathbb{R}$, and positive homogeneity $\rho(\lambda Z) = \lambda \rho(Z)$ if $\lambda > 0$ [30], [23, Ch. 6.3]. Such a functional also enjoys a distributionally robust representation [23, Th. 6.6]. A common example is the conditional value-at-risk of an integrable random variable $Z$, which represents the expectation of $Z$ in a given fraction of the worst cases [23, Th. 6.2]. A mean-dispersion risk functional takes the form $\rho(Z) = E(Z) + \lambda D(Z)$, where $\lambda \geq 0$ and $D$ is a measure of dispersion relative to $E(Z)$. Examples of $D$ include variance, standard deviation, and upper-semideviation, i.e., $E(\max\{Z - E(Z), 0\})$. A recursive risk functional of $Z = Z_0 + Z_1 + Z_2$ takes the form $\rho(Z) = \rho_1(Z_0) + \rho_1(Z_1 + \rho_1(Z_2))$, where $\rho_i$ is a map between spaces of random variables. While the structure of a recursive risk functional $\rho$ facilitates the development of risk-aware algorithms, it can be difficult to interpret $\rho$, except in special cases, such as when $\rho$ is the expectation. We will further describe specific risk functionals in Section II.

To meet application-specific needs and preferences about managing uncertainty, it is desirable to have at hand a broad collection of maps that assess different distributional characteristics. This is one motivation for the theory of risk functionals and the pursuit of additional research in the intersection of risk functionals and control systems. Research in this intersection has been gaining much momentum in recent years. There are new contributions, for example, in risk-aware mean-field games [33], [34], model predictive control [35], [36], temporal logic [37], [38], barrier functions [39], and optimal-control-based safety analysis [40], [41]. We refer the reader to our recent survey article about risk-aware optimal control theory [42] for additional literature. We will focus the remainder of the literature review on existing work related to risk-aware stability theory, which has been less studied.

A. Literature related to risk-aware stability theory

First, we will describe connections between the exponential utility functional and stability theory, which have been known for the past several decades. Then, we will describe the closest related works to the current work: Singh et al. [35], Sopasakis et al. [36], Kishida [43], and Tsianis et al. [44].

In 1988, Glover and Doyle established connections between the linear time-invariant controller that minimizes the infinite-time exponential utility cost functional and the class of stabilizing controllers that satisfy an $H_\infty$-norm bound in a linear-quadratic setting [45]. In 1995, James and Baras developed a solution approach for the robust $H_\infty$ output feedback nonlinear control problem, which includes an asymptotic stability specification, using techniques from an earlier study of a partially observable exponential utility control problem [46], [47]. In 1999, Pan and Başar provided conditions that guarantee global probabilistic asymptotic stability for nonlinear continuous-time stochastic systems in the context of a long-term exponential utility cost functional [48]. In 2000, Dupuis et al. derived an upper bound for the average output power in terms of the input power and the exponential utility functional, which led to a stochastic small gain theorem [49].

Singh et al. [35] and Sopasakis et al. [36] both studied risk-averse model predictive control with risk-aware stability properties that are defined using a recursive risk functional; see [35, Def. V.1] and [36, Def. 4], respectively. Singh et al. considered a linear system with multiplicative noise, i.e., $x_{t+1} = A(w_t)x_t + B(w_t)u_t$, where there are finitely many realizations of $w_t$ [35], Sopasakis et al. considered a nonlinear generalization with joint state-input constraints [36]. Lyapunov stability conditions are provided by [35, Lemma VI.1] and [36, Lemma 5], respectively. In particular, the lemma [35, Lemma VI.1] follows directly from a well-established sufficient condition for a classical exponential stability property [50, Def. 1, Lemma 1], which builds on techniques from [51]. In contrast to the above works, we consider linear models with additive noise, which yields noise-to-state stability properties, and we emphasize nonrecursive risk functionals and systems with continuous disturbance spaces.

Kishida developed stability theory for linear systems in the context of a specific coherent risk functional, that is, a distributionally robust conditional value-at-risk (CVaR) functional [43]. The system is linear with additive noise, and the ambiguity set is the family of disturbance distributions with zero mean and known time-invariant covariance [43, Eq. (2)]. Sufficient conditions for the robust CVaR stability property include the maximum singular value $\sigma_{\max}(A)$ of the dynamics matrix $A$ being strictly less than one [43, Lemma 3.2]. These efforts have inspired us to investigate stability conditions in the context of a broad family of coherent risk functionals for linear systems (Theorem 2). Our linear system model in Theorem 2 does not require zero-mean or identically distributed disturbances, and we circumvent the condition of $\sigma_{\max}(A) < 1$ (instead, our analysis relies on Schur stable matrices). We study connections between stability definitions that apply to nonlinear systems (Theorem 1), and we develop stability theory in the context of a particular noncoherent risk functional as well (Theorem 3).

In a linear-quadratic optimal control setting, Tsianis et al. proposed a constraint on a conditional variance, which admits a quadratic form [44]. We propose a risk functional that is a
weighted sum of the mean and a conditional variance (Section II Example 4). Then, we incorporate this functional into our risk-aware stability framework, deriving sufficient conditions for stability (Theorem 3). The idea of a mean-conditional-variance risk functional comes from the work [44]. However, we require different techniques that are useful for risk-aware stability theory, whereas the focus of [44] is risk-aware linear-quadratic optimal control theory.

Overall, in contrast to the above works, we advocate for a generalized risk-aware stability viewpoint, in which one can evaluate the state energy in terms of any risk functional. We see value in this degree of generality to accommodate the potential needs of diverse applications, in which different characterizations of state energy distributions may be useful.

B. Contributions

We propose a uniform stability analysis framework that characterizes diverse features of the state energy’s distribution by permitting \( p \) in (1) to be a general risk functional. In contrast, the standard paradigm characterizes the average energy or the probability that the state trajectory is close to the origin. Using a generalized risk-aware (uniform) stability property for nonlinear systems (Definition 3), our first contribution is to draw connections between different instances of this property for real-valued coherent risk functionals and other common risk functionals (Theorem 1). Then, specializing to linear systems, our second contribution is to derive sufficient conditions under which a risk-aware stability property holds, where the property is defined using a real-valued coherent risk functional (Theorem 2). These sufficient conditions also provide a distributionally robust risk-neutral stability property using the ambiguity set corresponding to the coherent risk functional of interest. Our final contribution is to derive sufficient conditions for stability in the context of a mean-conditional-variance functional (Theorem 3). The last two theorems uncover novel risk-aware noise-to-state stability properties, which have not been reported or studied in the literature to our knowledge.

C. Organization, notation, and terminology

Section II studies risk-aware (uniform) stability properties in the context of various risk functionals and discrete-time nonlinear systems. Section III specializes to linear systems and derives sufficient conditions under which risk-aware stability properties are guaranteed for some classes of risk functionals. Section IV offers illustrative examples. In particular, we require different techniques that are useful for risk-aware stability theory, whereas the focus of [44] is risk-aware linear-quadratic optimal control theory.

II. Generalized risk-aware stability

Consider a fully observable stochastic discrete-time system

\[
x_{t+1} = f(x_t, w_t), \quad t = 0, 1, \ldots,
\]

where \( f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^n \) is a Borel-measurable function, \( n \) and \( d \) are natural numbers, \((x_0, x_1, \ldots)\) is an \( \mathbb{R}^n \)-valued stochastic process, and \((w_0, w_1, \ldots)\) is an \( \mathbb{R}^d \)-valued independent stochastic process. The processes are defined on an arbitrary probability space \((\Omega, F, P)\). We assume that the initial state is fixed at an arbitrary vector \( x \in \mathbb{R}^n \).

There are many types of energy-like functions in control theory, e.g., class-\( K \) functions, locally positive definite functions, and decrescent functions [53, Sec. 5.3.1]. We will find the following energy-like function useful in this work.

Definition 1 (State energy function \( \psi \)): A state energy function \( \psi : \mathbb{R}^n \rightarrow \mathbb{R} \) is Borel-measurable, nonnegative, and satisfies \( \psi(x_0) = 0 \) if and only if \( x = 0_n \).

It is natural to choose \( \psi \) so that \( \psi(x) \) increases as \( |x| \) increases. However, we do not need this condition in the current section. In this section, we will study the relationships between different types of risk-aware stability notions for any \( \psi \) satisfying Definition 1. In contrast, in Section III we will specialize to \( \psi(x) = x^\top Rx \) with \( R \in \mathbb{S}^n_{++} \).

The next definition is a non-risk-aware (i.e., risk-neutral) uniform stability property. Variations of the definition to follow are available from [51, Def. 1], [50, Def. 1], and [10, Def. 7.3.8], for example.

More precisely, we assume that the distribution of \( x_0 \) is the Dirac measure concentrated at \( x \), where \( x \) can be any vector in \( \mathbb{R}^n \).
\textbf{Definition 2 (Risk-neutral stability):} Let a state energy function $\psi$ and a subset $S \subseteq \mathbb{R}^n$ be given. The system is uniformly exponentially stable with an offset with respect to the mean in region $S$ if and only if there exist $\lambda \in [0, 1)$, $a \in [0, \infty)$, and $b \in \mathbb{R}$ such that
\begin{equation}
E(\psi(x_t)) \leq \alpha \lambda \psi(x) + b
\end{equation}
for every time $t \in \mathbb{N}$ and initial condition $x \in S$. For brevity, we often write "stability with respect to the mean" instead of "uniform exponential stability with an offset with respect to the mean in region $S$." We refer to $\lambda$ as a rate parameter, $a$ as a scale parameter, and $b$ as an offset parameter.

Definition describes a uniform property because $(\lambda, a, b)$ does not depend on $x$ or $t$. While one can also consider a definition in which $(\lambda, a, b)$ can depend on $x$ or $t$, we focus on the uniform case in this work. Definition describes a local property if $S$ is a strictly proper subset of $\mathbb{R}^n$ and a global property otherwise. If $\psi(x) = |x|^2$ and $b = 0$, then Definition describes uniform exponential mean-square stability. If $b = \text{sup}_{\xi \in \mathbb{R}^n} E(\psi(\xi))^q < \infty$ for some $c \in (0, \infty)$ and $M \in S^+_n$, then Definition describes a risk-neutral noise-to-state stability property. (Additional risk-neutral noise-to-state stability definitions have been introduced in the literature; see [17, Def. 4.1] and [22, Def. 2] for two examples.)

As described in the Introduction, we are interested in analyzing the state energy over time, which is a random process $(\psi(x_1), \psi(x_2), \ldots)$. If we only consider the average state energy, as in Definition then we ignore other characteristics of the distribution of the state energy, including the dispersion with respect to the mean (e.g., variance) and the shape of the upper tail (e.g., the mean in a fraction of the worst cases). Hence, we will develop a risk-aware stability concept to facilitate the analysis of a broader range of distributional characteristics of the state energy.

To generalize Definition the key observation is that the expectation is a map from a family of random variables to $\mathbb{R}$. Hence, we can consider maps that characterize additional features of the distribution of a random variable. Let $Z$ be a family of random variables defined on $(\Omega, \mathcal{F}, P)$, and let $\rho$ be a map from $Z$ to $\mathbb{R}$. $\rho$ is called a risk functional (i.e., risk measure), and several examples will follow. The precise definition of $\rho$ will depend on the risk functional of interest. While Examples can be found in [23], they are necessary to present to keep the current work self-contained.

\textbf{Example 1 (Value-at-risk):} Let $Z$ be the entire family of random variables on $(\Omega, \mathcal{F}, P)$. The value-at-risk of $Z \in Z$ at level $\alpha \in (0, 1]$ is defined by
\begin{equation}
\text{VaR}_\alpha(Z) := \inf\{z \in \mathbb{R} : F_Z(z) \geq 1 - \alpha\},
\end{equation}
where $F_Z(z) := P(\{Z \leq z\})$ is the distribution function of $Z$. The map $(1 - \alpha) \mapsto \text{VaR}_\alpha(Z)$ is the generalized inverse of $F_Z$, i.e., the quantile function of $F_Z$ [54, p. 304].

\text{Example 2 (Conditional value-at-risk):} The conditional value-at-risk of $Z \in Z = L^1$ at level $\alpha \in (0, 1]$ is defined by
\begin{equation}
\text{CVaR}_\alpha(Z) := \inf_{s \in \mathbb{R}} \left(s + \frac{1}{\alpha} E(\max\{Z - s, 0\})\right).
\end{equation}

CVaR$_1(Z)$ equals $E(Z)$, and the limit of CVaR$_\alpha(Z)$ as $\alpha \to 0$ from above equals the essential supremum of $Z$ [55]. The name conditional value-at-risk comes from the following: If $\alpha \in (0, 1)$ and $F_Z$ is continuous at $\text{VaR}_\alpha(Z)$, then $\text{CVaR}_\alpha(Z) = E(Z)1_{\{Z \geq \text{VaR}_\alpha(Z)\})} [23, Th. 6.2]$. Synonyms include average value-at-risk and expected shortfall.

The value-at-risk and the conditional value-at-risk assess risk in terms of quantiles. The following functionals assess risk in terms of dispersions relative to the mean.

\textbf{Example 3 (Mean-deviation, mean-upper-semideviation):} Let $q \in [1, \infty)$, $\beta \in [0, \infty)$, and $Z \in Z = L^q$ be given. The mean-deviation of $Z$ is defined by
\begin{equation}
\text{MD}_{q,\beta}(Z) := E(Z) + \beta \|Z - E(Z)\|_q,
\end{equation}
while the mean-upper-semideviation of $Z$ is defined by
\begin{equation}
\text{MUS}_{q,\beta}(Z) := E(Z) + \beta \max\{Z - E(Z), 0\}\|_q.
\end{equation}

The second term in penalizes realizations of $Z$ above the mean but not realizations of $Z$ below the mean. However, the second term in does not distinguish between the two cases.

Conditional value-at-risk and the mean-dispersion functionals of Example for particular choices of $q$ and $\beta$ belong to the class of real-valued coherent risk functionals [23, Ch. 6.3]. As described previously, such a functional satisfies four desirable properties and also admits a dual representation as a distributionally robust expectation (to be presented). Let $\varrho$ be a real-valued coherent risk functional defined on $L^q$ with $q \in [1, \infty)$. Then, there is a bounded family $A \subset L^{q\ast}$ of densities such that
\begin{equation}
\varrho(Z) = \sup_{\xi \in A} E(Z(\xi)), \quad Z \in L^q,
\end{equation}
by one direction of [23, Th. 6.6] in particular, every $\xi \in A$ satisfies $\xi \in L^{q\ast}$, $E(\xi) = 1$, and $\xi \geq 0$ a.e. [23, Eq. (6.38)]. Using (9), one can show that if $Z = 0$ a.e., then $\varrho(Z) = 0$. We will find the exact forms of $A$ for Examples and useful for Theorem to come.

\textbf{Remark 1 (Special cases of $A$):} For conditional value-at-risk at level $\alpha \in (0, 1]$, $A$ is given by [23, Eq. (6.70)]
\begin{equation}
A = \{\xi \in L^{\infty} : 0 \leq \xi \leq \frac{1}{\alpha} \text{ a.e.}, \; E(\xi) = 1\}.
\end{equation}

Note that $\xi \in A$ implies that $|\xi| \leq \frac{1}{\alpha}$ a.e. Here, the upper bound is hyperbolic in the parameter $\alpha$. In the following two cases, however, we will present an upper bound that is affine in the corresponding parameter.

For $q = 1$ and $\beta \in [0, \frac{1}{2}]$, mean-deviation is real-valued and coherent [23, Ex. 6.19], and $A$ is given by [23, Eq. (6.90)]
\begin{equation}
A = \{\xi \in L^{\infty} : \xi = 1 + \eta - E(\eta), \; \eta\|_{\infty} \leq \beta\}.
\end{equation}
If $\xi \in \mathcal{A}$, then $0 \leq \xi \leq 1 + 2\beta$ a.e. The upper bound $1 + 2\beta$ is affine in $\beta \in [0, \frac{1}{2}]$.

For $q \in [1, \infty)$ and $\beta \in [0, 1]$, mean-upper-semideviation is real-valued and coherent [23, Ex. 6.20]. The corresponding family $\mathcal{A}$ of densities is given by [23, Eq. (6.96)]

$$\mathcal{A} = \{ \xi \in L^q : \xi = 1 + \eta - E(\eta), \|\eta\|_{L^q} \leq \beta, \eta \geq 0 \text{ a.e.} \}.$$  

(12)

In particular, if $q = 1$ and $\xi \in \mathcal{A}$, then $0 \leq \xi \leq 1 + 2\beta$ a.e. The upper bound is affine in $\beta \in [0, 1]$.

The next risk functional $\rho_\nu$ to be described, which takes inspiration from [44], is related to Example 3 because it also assesses risk in terms of a dispersion relative to the mean. While $\rho_\nu$ is noncoherent in general, it enjoys a special meaning in dynamical systems applications, as it depends on a sub $\sigma$-algebra, which can encode the history of a stochastic process.

Example 4 (Mean-conditional-variance $\rho_\nu$): Given $\nu \in [0, \infty)$ and a sub $\sigma$-algebra $\mathcal{F}_t$ of $\mathcal{F}$, the mean-conditional-variance of $Z \in \mathcal{Z} = L^1$ is defined by

$$\rho_\nu(Z) := E(Z) + \nu E(\Delta_t^2),$$  

(13)

where $\Delta_t$ is a real-valued prediction error defined by

$$\Delta_t := E(Z|\mathcal{F}_t) - Z,$$  

(14)

$$F_t := \{ \omega \in \Omega : E(Z|\mathcal{F}_t)(\omega) \in \mathbb{R} \}. \quad (15)$$

Note that $\rho_\nu(Z)$ can be infinite, $E(\Delta_t^2)$ is called a conditional variance because $\Delta_t = E(Z|\mathcal{F}_t) - Z$ a.e.

Any risk functional, e.g., see Examples 1-4, can be incorporated into the following definition.

Definition 3 (Risk-aware stability): Let a risk functional $\rho : \mathcal{Z} \rightarrow \mathbb{R}$, a state energy function $\psi$, and a subset $S \subseteq \mathbb{R}^n$ be given. The system (5) is uniformly exponentially stable with an offset with respect to region $S$ if and only if $\psi(x_t) \in S$ for every $t \in \mathbb{N}$ and there exist $\lambda \in [0, 1), a \in [0, \infty)$, and $b \in \mathbb{R}$ such that

$$\rho(\psi(x_t)) \leq a\lambda^t \psi(x) + b$$  

(16)

for every time $t \in \mathbb{N}$ and initial condition $x \in S$. For brevity, we often write “stability with respect to $\rho$” instead of “uniform exponential stability with an offset with respect to $\rho$ in region $S$.” As before, we refer to $\lambda$ as a rate parameter, $a$ as a scale parameter, and $b$ as an offset parameter.

In Section III we will see how $b$ can depend on the disturbance process and the chosen measure of risk, leading to a risk-aware noise-to-state stability property. In Definition 3 the condition $\psi(x_t) \in S$ for every $t \in \mathbb{N}$ is required because the domain of $\rho$ need not be the entire family of random variables on $(\Omega, \mathcal{F}, P)$, e.g., see Examples 2-4. In the special case of $\rho$ being the expectation and $S$ being the entire family of random variables on $(\Omega, \mathcal{F}, P)$, Definition 3 reduces to Definition 2. While Definition 3 is inspired by [35, Def. V.1], [36, Def. 4], and [43, Def. 3.1], Definition 3 invokes a generalized viewpoint because $\rho$ can be any risk functional. In contrast, a recursive risk functional was considered in [35, Def. V.1] and [36, Def. 4], and a distributionally robust CVaR functional was considered in [43, Def. 3.1]. The next theorem will demonstrate connections between several instances of Definition 3.

Theorem 1 (Analysis of Definition 3): Consider the discrete-time nonlinear system $\beta$. Let a state energy function $\psi$, a subset $S \subseteq \mathbb{R}^n$, scalars $\alpha \in (0, 1)$ and $q \in [1, \infty)$, and a real-valued coherent risk functional $\varrho$ on $L^q$ be given. Then, the following statements hold:

1) Stability w.r.t. $\text{CVaR}_\alpha$ in region $S$ with parameters $(\lambda, a, b)$ implies the probabilistic stability property $P\{\psi(x_t) \leq a\lambda^t \psi(x) + b\} \geq 1 - \alpha$ for every $t \in \mathbb{N}$ and $x \in S$.

2) Stability w.r.t. the mean with parameters $(\lambda, a, b)$ implies stability w.r.t. $\text{CVaR}_\alpha$ with parameters $(\lambda, a, b)$.

3) Stability w.r.t. the mean with parameters $(\lambda, a, b)$ implies stability w.r.t. the mean-deviation on $L^1$ and the mean-upper-semideviation on $L^1$, both with parameters $(\lambda, a(1 + 2\beta), b(1 + 2\beta))$. In the case of the mean-deviation, $\beta \in [0, \frac{1}{2}]$. In the case of the mean-upper-semideviation, $\beta \in [0, 1]$.

4) Stability w.r.t. the mean-deviation on $L^q$ implies stability w.r.t. the mean-upper-semideviation on $L^q$ with the same parameters.

5) Stability w.r.t. $\varrho$ in region $S$ with parameters $(\lambda, a, b)$ implies the distributionally robust stability property w.r.t. the mean $E(\psi(x_t) \xi) \leq a\lambda^t \psi(x) + b$, $\xi \in \mathcal{A}$, $t \in \mathbb{N}$, $x \in S$, where $\mathcal{A}$ is the family of densities in the dual representation of $\varrho$.

Remark 2 (Interpretation of Theorem 1): The first item of Theorem 1 states that a CVaR stability property guarantees a probabilistic stability property. While connections between CVaR and probabilities are well-established, the first item of Theorem 1 is illuminating in light of the extensive history of probabilistic stability theory, e.g., see [10], [15], [56]. The second and third items indicate that stability with respect to the mean ensures stability with respect to some common real-valued coherent risk functionals with transformed scale and offset parameters. The nature of the transformation depends on the specific family of densities in the dual representation. While the mean-deviation and the mean-upper-semideviation are similar functionals, the fourth item indicates that stability with respect to the mean-deviation is a stronger property. The last item states that stability with respect to a real-valued coherent risk functional on $L^q$ guarantees a distributionally robust risk-neutral stability property, which we will see again in the final part of Theorem 2 to be presented in Section III.

The proof of Theorem 1 has five parts, one for each item. Each part is based on the properties enjoyed by the risk functional at hand.

Proof: Part 1: Let $t \in \mathbb{N}$ and $x \in S$ be given. Since $\alpha \in (0, 1)$ and $\psi(x_t) \in L^1$, a minimizer of the right side of (6) with $Z = \psi(x_t)$ is $\text{VaR}_\alpha(\psi(x_t))$ [23, p. 258], and consequently, $\text{VaR}_\alpha(\psi(x_t)) \leq \text{CVaR}_\alpha(\psi(x_t))$. Hence, $\text{CVaR}_\alpha(\psi(x_t)) \leq a\lambda^t \psi(x) + b$ implies that

$$\text{VaR}_\alpha(\psi(x_t)) \leq a\lambda^t \psi(x) + b.$$  

(17)
Since $\alpha \in (0, 1)$ and $z = a\lambda^t \psi(x) + b \in \mathbb{R}$,

$$\text{VaR}_\alpha(\psi(x_t)) \leq z \iff P(\{\psi(x_t) \leq z\}) \geq 1 - \alpha \quad (18)$$

by [54, Lemma 21.1 (i)] and the definition of $\text{VaR}_\alpha$.  

The previous statement applies to both Parts 2 and 3, in which stability with respect to the mean is assumed. By Definition 2 and nonnegativity of $\psi$, it holds that $0 \leq E(\psi(x_t)) \leq a\lambda^t \psi(x) + b$ for every $t \in \mathbb{N}$ and $x \in S$. In particular, $\psi(x_t) \in L^1$ for every $t \in \mathbb{N}$.

Part 2: Let $t \in \mathbb{N}$ and $x \in S$ be given. Since $\psi(x_t) \in L^1$ and $\psi(x_t)$ is nonnegative, CVaR$_\alpha(\psi(x_t)) \leq \frac{1}{\alpha} E(\psi(x_t))$ [40, Lemma 2.1]. Thus, CVaR$_\alpha(\psi(x_t)) \leq \frac{1}{\alpha} (a\lambda^t \psi(x) + b)$.

Part 3: For $\beta \in (0, \frac{1}{2}]$, the mean-deviation on $L^1$ is real-valued and coherent. Denoting the associated family of densities by $A$, every $\xi \in A$ satisfies $0 \leq \xi \leq 1 + 2\beta$ a.e. (Remark 1). Let $t \in \mathbb{N}$ and $\xi \in A$ be given. Since $\psi(x_t) \geq 0$ and $\xi \in [0, 1 + 2\beta]$ a.e., $0 \leq E(\psi(x_t)\xi) \leq E(\psi(x_t))(1 + 2\beta)$.

Using $\psi(x_t) \in L^1$ and taking the supremum over $A$, we find that $\text{MD}_{1,\beta}(\psi(x_t)) \leq E(\psi(x_t))(1 + 2\beta)$. Then, the result follows from $0 \leq E(\psi(x_t)) \leq a\lambda^t \psi(x) + b$. The derivation in the case of the mean-upper-semideviation is analogous.

Part 4: Let $\beta \in [0, \infty)$ be given. By assumption, $\psi(x_t) \in L^q$ for every $t \in \mathbb{N}$, and there exist $\lambda \in [0, 1)$, $a \in [0, \infty[$, and $b \in \mathbb{R}$ such that

$$\text{MD}_{q,\beta}(\psi(x_t)) \leq a\lambda^t \psi(x) + b$$

for every $t \in \mathbb{N}$ and $x \in S$. Now, let $t \in \mathbb{N}$ and $x \in S$ be given. Since $q \in [1, \infty)$, $\psi(x_t) \in L^q$, and $P(\Omega) = 1$, $E(\psi(x_t))$ and $\|\psi(x_t) - E(\psi(x_t))\|_q$ are finite. Also, since $0 \leq \max\{g, 0\} \leq |g|$ for any $g : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $y \rightarrow y^\gamma$ is nondecreasing on $[0, \infty)$, for any $\gamma \in (0, \infty)$,

$$\max\{\psi(x_t) - E(\psi(x_t)), 0\} \leq \|\psi(x_t) - E(\psi(x_t))\|_q.$$

Hence, MUS$_{q,\beta}(\psi(x_t)) \leq \text{MD}_{q,\beta}(\psi(x_t)) \leq a\lambda^t \psi(x) + b$.

Part 5: By assumption, $\psi(x_t) \in L^q$ for every $t \in \mathbb{N}$, and there exist $\lambda \in [0, 1)$, $a \in [0, \infty[$, and $b \in \mathbb{R}$ such that $\psi(\psi(x_t)) \leq a\lambda^t \psi(x) + b$ for every $t \in \mathbb{N}$ and $x \in S$. Using the dual representation 3 of $\psi$ with the corresponding family $A$ of densities and $\psi(x_t) \in L^q$ for every $t \in \mathbb{N}$, $\xi \in A$,

$$E(\psi(x_t)\xi) \leq \sup_{\xi \in A} E(\psi(x_t)\xi) \leq a\lambda^t \psi(x) + b$$

for every $\xi \in A$, $t \in \mathbb{N}$, and $x \in S$.

The previous theorem provides connections between some different instances of Definition 3. In the next section, we will focus on linear systems and derive sufficient conditions for stability with respect to any real-valued coherent risk functional on $L^q$ (Theorem 3). Then, we will provide a stability result in the context of the mean-conditional-variance functional on $L^1$ (Theorem 5). The techniques will involve a combination of the properties of the risk functionals of interest and Lyapunov stability theory for discrete-time linear systems.

### III. Risk-aware Stability Conditions for Stochastic Linear Systems

Now, we consider a stochastic linear time-invariant system, which is a special case of 3, of the form

$$x_{t+1} = Ax_t + w_t, \quad t = 0, 1, \ldots$$

where $A \in \mathbb{R}^{n \times n}$ is deterministic, $(x_0, x_1, \ldots)$ is an $\mathbb{R}^n$-valued stochastic process, and $(w_0, w_1, \ldots)$ is an $\mathbb{R}^n$-valued independent stochastic process. The processes are defined on $(\Omega, \mathcal{F}, P)$. The initial state is fixed at an arbitrary vector $x \in \mathbb{R}^n$. For every $t \in \mathbb{N}_0$, we define the random object $h_t := (x_0, x_1, \ldots, x_t)$ and the $\sigma$-algebra $\mathcal{F}_t := \sigma(h_t)$ induced by $h_t$. Thus, $x_t : (\Omega, \mathcal{F}_t) \rightarrow (\mathbb{R}^n, \mathcal{B}_\mathbb{R})$. As well as $(w_0, w_1, \ldots)$ being independent, we assume that $w_t$ and $h_t$ are independent and $E(w_t w_{t+1}) \in \mathbb{R}^{n \times n}$ for every $t \in \mathbb{N}_0$. The above conditions are standard and ensure that $E(\|x_t\|^2)$ is finite for every $t \in \mathbb{N}_0$. Observe that $(w_0, w_1, \ldots)$ need not be zero-mean, and the disturbance process need not be identically distributed. When we consider the linear system 10, we implicitly assume the conditions stated in this paragraph.

The first result of the section will develop sufficient conditions for stability of the linear system 10 with respect to any real-valued coherent risk functional on $L^q$. Then, we will examine how the conditions are related to those that guarantee risk-neutral stability of 10.

Theorem 2 (Coherent stability): Consider the linear system 19 and the quadratic state energy function $\psi(x) := x^T R x$, where $R \in \mathbb{R}^{n \times n}_+$ is given. Let $\bar{q}$ be a real-valued coherent risk functional on $L^q$ with $q \in [1, \infty)$, where $A$ is the family of densities in the associated dual representation. We make the following assumptions:

1) $|w_t|_2 \in L^q$ for every $t \in \mathbb{N}_0$, $\sup_{t \in \mathbb{N}_0} \|\|w_t\|^2\|_q$ is finite, and

2) there exists an $H \in S^+_{\mathbb{R}}$ such that $H - A^T H R A \in S^+_{\mathbb{R}}$. (Recall that $H : (R^+)^* \rightarrow H R^+$.) Then, the system 19 is stable with respect to $\bar{q}$ in $\mathbb{R}^n$ (Definition 3). Define $\eta := \frac{\lambda_{\min}(H - A^T H R A)}{\lambda_{\min}(H)}$, and it follows that $\eta \in (0, 1]$. In the case of $\eta < 1$, for any fixed $\bar{q} \in (0, 1)$, one can choose

$$\lambda = 1 - \kappa \eta, \quad a = \frac{\lambda_{\min}(H)}{\lambda_{\min}(H)} b = c , \quad b = \bar{c}$$

where $\lambda \in (0, 1)$, $c := \frac{\lambda}{\lambda_{\min}(H)}(1 - \lambda)(1 - \eta^2)$, and

$$b' := \sup_{t \in \mathbb{N}_0} \|w_t^T H R w_t\|.$$  

In the case of $\eta = 1$, one can choose $\lambda = a = 0$ and $b = \bar{b}$, where $b'$ is defined by 21 and $\bar{c} := 1 / \lambda_{\min}(H)$. Moreover, a distributionally robust risk-neutral stability property holds:

$$E(\psi(x_t)\xi) \leq a\lambda^t \psi(x) + b$$

for every density $\xi \in A$, time $t \in \mathbb{N}$, and initial condition $x \in \mathbb{R}^n$.

Remark 3 (Discussion of Theorem 2): Theorem 2 specifies sufficient conditions that guarantee

$$\bar{q}(x_t^T R x_t) \leq a\lambda^t \psi(x) + b'$$

where $\lambda \in [0, 1)$, and $x_t$ is defined in Theorem 2. Moreover, since $\lambda \in [0, 1)$, we derive the following risk-aware stability bound:

$$\lim \sup_{t \rightarrow \infty} \bar{q}(x_t^T R x_t) \leq \bar{c} \sup_{t \in \mathbb{N}_0} \|w_t^T H R w_t\|.$$  

Now, let us discuss the assumptions of Theorem 2. In the case of $q = 1$, $\sup_{t \in \mathbb{N}_0} \|w_t^2\|_1 = \sup_{t \in \mathbb{N}_0} E(\|w_t^2\|)$ is finite, for example, if there is a matrix $\Sigma \in \mathbb{S}_n$ such that $\Sigma - E(w_tw_t^T) \in \mathcal{S}_n$ for every $t \in \mathbb{N}_0$. If $E(w_tw_t^T)$ is time-invariant, then such a $\Sigma$ exists immediately. (Recall that $E(w_tw_t^T) \in \mathbb{R}^{n \times n}$ in the model (19).) More generally, the suitability of the first assumption is problem-dependent, and the assumption holds in particular if $w_t$ is uniformly bounded almost everywhere. Since $R \in \mathcal{S}_n^+$, the second assumption is equivalent to $A$ being Schur stable; i.e., every eigenvalue of $A$ has magnitude strictly less than one (Lemma 1 Appendix).

We have developed our proof of Theorem 2 from first principles using several techniques, including properties of $\mathcal{S}_n$ from [57], some basic manipulations from [51, Lemma 1], and properties of real-valued coherent risk functionals from [23].

Proof: The property $\eta \in (0, 1]$ holds by Lemma 2 (Appendix). We shall prove the theorem in the case of $\eta < 1$ and leave the case of $\eta = 1$ to the reader, as the techniques are similar. First, we will verify some properties of the parameters in (20). The properties $\eta \in (0, 1)$ and $\kappa \in (0, 1)$ imply that $\lambda \in (0, 1)$. The property $a \in (0, \infty)$ holds because $H \in \mathcal{S}_n^+$. The property $c \in (0, \infty)$ is true because $\lambda \in (0, 1)$, $\lambda > 1 - \eta$, and $H \in \mathcal{S}_n^+$. For every $\xi \in A$ and $t \in \mathbb{N}_0$,

$$0 \leq E(w_t^T H_{Rw_t} \xi) \leq \lambda_{\max}(H_R)E(\|w_t^2\|_2^2)$$

(24) because $H_{R} \in \mathcal{S}_n^+$, $w_t$ is $\mathbb{R}^n$-valued, and $\xi \geq 0$ a.e. Also,

$$E(\|w_t^2\|_2^2) \leq \|w_t^2\|_q \|\xi\|_q, \quad \xi \in A, \quad t \in \mathbb{N}_0,$$

(25) by applying [58, Hölder’s Inequality, Th. 6.8 (a)]. Since $g$ is a real-valued coherent risk functional on $\mathcal{L}^q$, $A$ is a bounded subset of $\mathcal{L}^q$ (23, Th. 6.6). The property $b \in [0, \infty)$ follows from (24)–(25), boundedness of $A$ in $\|\cdot\|_q$, the assumptions of the theorem, the dual representation of $g$ (7), and $c \in (0, \infty)$.

Second, we will verify properties of some expectations. For any density $\xi \in A$, initial condition $x \in \mathbb{R}^n$, and function $g : (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$,

$$E(g(x_0)\xi) = g(x)$$

(26) because $E(\xi) = 1$ and the distribution of $x_0$ is the Dirac measure concentrated on $x$. Let $t \in \mathbb{N}_0$, $\xi \in A$, and $M \in \mathcal{S}_n$ be given. Then, $x_{t+1} = Mx_t \in \mathcal{L}^q$ and $x_t \in \mathcal{L}^1$, and these functions are a.e.-nonnegative (Lemma 4 Appendix). In particular, since $R \in \mathcal{S}_n^+$ and $H_{R} \in \mathcal{S}_n^+$, $\psi(x_t) = x_t^T H_{R} x_t \in \mathcal{L}^q$ and $E(x_t^T H_{R} x_t) \in (0, \infty)$.

Having verified the above inequalities, we are ready to show that $g(\psi(x_t)) \leq a \lambda^t \psi(x) + b$ for every time $t \in \mathbb{N}$ and initial condition $x \in \mathbb{R}^n$. Define $v : \mathbb{R}^n \to \mathbb{R}$ by $v(z) := z^T H_{R} z$, which will serve as a Lyapunov function. Recall the definition of $b'$ from (21). We claim it suffices to prove that, for every $\xi \in A$ and $t \in \mathbb{N}_0$,

$$E(v(x_{t+1})\xi) \leq \lambda E(v(x_t)\xi) + \frac{\lambda^t}{(1 - \lambda)(\lambda - 1 - \eta)} b'.$$

(27) Indeed, the statement (27) and $E(v(x_t)\xi) \in (0, \infty)$ for every $t \in \mathbb{N}_0$ imply

$$0 \leq E(v(x_t)\xi) \leq \lambda^t E(v(x_0)\xi) + \frac{\lambda^t}{(1 - \lambda)(\lambda - 1 - \eta)}$$

(28) for every $t \in \mathbb{N}$, where we use $\lambda \in (0, 1)$, $\lambda > 1 - \eta$, the geometric series formula, and $b' \in [0, \infty)$ (Lemma 5 Appendix). Since $H \in \mathcal{S}_n^+$, $R \in \mathcal{S}_n^+$, and every $\xi \in A$ is nonnegative a.e., we have

$$\lambda_{\min}(H)E(\psi(x_t)\xi) \leq E(v(x_t)\xi) \leq \lambda_{\max}(H)E(\psi(x_t)\xi)$$

(29) for every $\xi \in A$ and $t \in \mathbb{N}_0$ (Lemma 6 Appendix). We divide by $\lambda_{\min}(H) \in (0, \infty)$ to find

$$0 \leq E(\psi(x_t)\xi) \leq \frac{1}{\lambda_{\min}(H)} E(v(x_t)\xi).$$

(30) Using (29) with $t = 0$ and (26) with $g = \psi$ lead to

$$0 \leq E(v(x_0)\xi) \leq \lambda_{\max}(H)E(\psi(x_0)\xi) = \lambda_{\max}(H)\psi(x)$$

(31) for every density $\xi \in A$ and initial condition $x \in \mathbb{R}^n$. The statements (28) and (31) imply

$$0 \leq E(v(x_t)\xi) \leq \lambda^t \lambda_{\max}(H)\psi(x) + \frac{\lambda^t}{(1 - \lambda)(\lambda - 1 - \eta)}$$

(32) for every $\xi \in A$, $t \in \mathbb{N}$, and $x \in \mathbb{R}^n$. We divide (32) by $\lambda_{\min}(H) \in (0, \infty)$ and substitute the expressions for $a$ and $b$ from (20) to find

$$0 \leq \frac{1}{\lambda_{\min}(H)} E(v(x_t)\xi) \leq a \lambda^t \psi(x) + b$$

(33) for every $\xi \in A$, $t \in \mathbb{N}$, and $x \in \mathbb{R}^n$. Then, taking the supremum over $A$ and using the dual representation (9) of $g$ lead to

$$g(\psi(x_t)) \leq a \lambda^t \psi(x) + b, \quad t \in \mathbb{N}, \quad x \in \mathbb{R}^n.$$ (35)

Hence, it suffices to show (27), which we prove in the Appendix (see Lemma 3 and Lemma 7).

Theorem 2 provides sufficient interpretable conditions for global stability of the linear system (19) with respect to any real-valued coherent risk functional $g$ on $\mathcal{L}^q$ (Definition 3). We have presented some common examples of $g$ in Section 1 which assess different distributional characteristics of the state energy. For instance, $g(\psi(x_t)) = \text{CVar}_{\gamma}(\psi(x_t))$ is the average energy in the $\alpha$-fraction of the largest energies, if $\psi(x_t)$ is a continuous random variable. For another example, $g(\psi(x_t)) = \text{MUS}_{\infty}(\psi(x_t))$ is a weighted sum of the average energy $\mu_t := E(\psi(x_t))$ and the order-$q$ upper-semideviation $\|\max\{\psi(x_t) - \mu_t, 0\}\|_q$ of the energy. Theorem 2 facilitates the analysis of various characteristics of the distribution of the state energy and offers a distributionally robust stability guarantee (see also Part 5 of Theorem 1).

It is quite interesting that the assumptions of Theorem 2 in the case of $q = 1$ ensure risk-neutral stability of the system (19), i.e., the existence of $\lambda_N \in (0, 1)$, $a_N \in [0, \infty)$, and $b_N \in \mathbb{R}$ such that

$$E(\psi(x_t)) \leq a_N \lambda_N \psi(x) + b_N, \quad t \in \mathbb{N}, \quad x \in \mathbb{R}^n.$$ (36)

While the proof of this risk-neutral result is nontrivial, it is a simpler version of the proof of Theorem 2 (consider $\xi = 1$ a.e. and $q = 1$). The values of $\lambda_N$ and $a_N$ can be chosen to be the values of $\lambda$ and $a$, respectively, provided by Theorem 2. However, the offset term $b_N$ in the risk-neutral result and the
offset term $b$ in the risk-aware result (Theorem 2) are distinct.
In the case of $\eta < 1$, one can choose
\[ b_N = c \sup_{t \in \mathbb{N}} E(w_t^\top H_RW_t), \]  
(37)
while in the case of $\eta = 1$, one can choose $b_N = \bar{c} \sup_{t \in \mathbb{N}} E(w_t^\top H_RW_t)$, where $c$ and $\bar{c}$ are specified by Theorem 3.

Let us further discuss the different offset terms. The offset term $b_N$ in (37) is proportional to the supremum of the mean of $w_t^\top H_RW_t$. In contrast, the offset term $b$ in Theorem 2 is proportional to the supremum of the risk of $w_t^\top H_RW_t$, where the meaning of risk is specific to the functional $\varphi$ of interest. As well as permitting a risk-aware perception of $w_t^\top H_RW_t$ that is specific to the primal interpretation of $\varphi$, Theorem 2 permits distributional ambiguity in $w_t^\top H_RW_t$ due to the dual representation of $\varphi$. This attribute is useful in settings with distributional modeling uncertainty, which we will illustrate by using the mean-upper-semideviation as an example (Section IV: Illustration 1).

Theorem 3 to follow will use a mean-conditional-variance functional $\rho_\nu$ (Example 4) and additional assumptions about the distribution of $w_t$, which are inspired by [44]. Theorem 3 will not provide a distributional robustness property, which is not surprising because $\rho_\nu$ is noncoherent in general. However, Theorem 3 will provide a versatile risk-aware noise-to-state stability property that depends on second-, third-, and fourth-order centered noise statistics, which is relevant for controller design (to be explored in Illustration 3 Section IV). Prior to presenting Theorem 3 we will present some useful notation.

Remark 4 (Notation for Theorem 3): As in Theorem 3, Theorem 3 will also involve the linear system (19) and a quadratic state energy function $\psi(x) := x^\top Rx$, where $R \in \mathcal{S}^+_N$ is given. Since $\psi(x_t) \in L^1$ for any $t \in \mathbb{N}$, the mean-conditional-variance of $\psi(x_t)$ is well-defined (Example 4). For every $t \in \mathbb{N}$, $\rho_\nu(\psi(x_t))$ denotes the mean-conditional-variance of $\psi(x_t)$, where $F_{t-1} = \sigma(x_0,x_1,\ldots,x_{t-1})$ is the sub $\sigma$-algebra of $\mathcal{F}$ of interest. For every $t \in \mathbb{N}$, we define $\bar{w}_{t-1} := E(w_{t-1}, d_t := w_{t-1} - \bar{w}_{t-1}$, and $\Sigma_t := E(d_t d_t^\top)$.

Theorem 3 (Mean-cond.-variance stability): Consider the stochastic linear system (19), and let $R \in \mathcal{S}^+_N$ and $\nu \in [0, \infty)$ be given. Consider the quadratic state energy function $\psi(x) := x^\top Rx$. We make the following assumptions:

1) $E(|w_t|^4)$ is finite for every $t \in \mathbb{N}_0$, and there exists a matrix $\Sigma_\nu \in \mathcal{S}_N$ such that $\Sigma_\nu - \Sigma_1 \in \mathcal{S}_N$ for every $t \in \mathbb{N}$.
2) Define $R_\nu := R + 4\nu R\Sigma_\nu R$. There exists a matrix $H_\nu \in \mathcal{S}^+_N$ such that $H_\nu = A^\top H_\nu A \in \mathcal{S}^+_N$.
3) The statistics $\bar{w}_t, \gamma_t$, and $\delta_t$ defined by
   \[ \bar{w}_t := (I_n - A^{t+1})^{-1} \sum_{i=0}^{t} A^i \bar{w}_{t-i}, \quad t \in \mathbb{N}_0, \]
   \[ \gamma_t := E(d_t d_t^\top R d_t), \quad t \in \mathbb{N}, \]
   \[ \delta_t := E((d_t^\top R d_t - \text{tr}(\Sigma_t R))^2), \quad t \in \mathbb{N}, \]
   satisfy $\sup_{t \in \mathbb{N}_0} |\bar{w}_t|^2 < \infty$, $\sup_{t \in \mathbb{N}} |\gamma_t|^2 < \infty$, and $\sup_{t \in \mathbb{N}} |\delta_t|^2 < \infty$, respectively.

Then, there exist $\{\lambda_\nu, \lambda_0\} \subset [0, 1)$, $\{a_\nu, a_0\} \subset [0, \infty)$, and $b_\nu \in \mathbb{R}$ such that
\[ \rho_\nu(\psi(x_t)) \leq a_\nu \lambda_\nu \psi(\bar{w}_t) + 4\nu |\gamma_t|^3 \sqrt{a_0 \lambda_0 \max(R) \psi(\lambda_\nu \bar{w}_t)} + b_\nu \]  
(39)
Exponentially decreasing terms
for every time $t \in \mathbb{N}$ and initial condition $x \in \mathbb{R}^n$, where we define $\psi_\nu(x) := 4\nu \psi(x) + 4\lambda_\max(R) \sup_{t \in \mathbb{N}_0} |\bar{w}_t|^2$ and $\psi_\nu(x) := x^\top R_\nu x$. In particular, one can choose

\[ \lambda_\nu = 1 - \frac{\lambda_{\min}(H_{R_\nu} - A^\top H_{R_\nu} A)}{\lambda_{\min}(H_{R_\nu}^\nu)}, \quad a_\nu = \frac{\lambda_{\max}(H_{R_\nu}^\nu)}{\lambda_{\min}(H_{R_\nu}^\nu)}, \]  
(40)
and
\[ b_\nu = \sup_{t \in \mathbb{N}} \left\{ \rho_\nu(\psi(d_t)) + b_{\nu,t} \right\}, \]  
(41)
where, for every $t \in \mathbb{N}$,
\[ b_{\nu,t} := c_\nu \left( \sup_{t \in \mathbb{N}} E(d_t^\top R_{\nu,t} d_t) - E(d_t^\top R_{\nu,t} d_t) \right) \]
\[ + 2\bar{w}_{t-1}^\top R_{\nu,t} \bar{w}_{t-1} + 4\nu |\gamma_t|^3 R^\top \bar{w}_{t-1}, \]
\[ \text{Risk-neutral zero-mean noise bias} \]
\[ \text{Noise mean and skewness terms} \]
\[ R_{\nu,t} := R + 4\nu R\Sigma_\nu R, \quad c_\nu := \frac{2 \lambda_{\min}(R)^2 (1 - \lambda_\nu)}{\lambda_{\min}(H_{\nu}^\nu)}, \]

Remark 5 (Discussion of Theorem 3): Theorem 3 establishes a global stability property with respect to $\rho_\nu$, closely resembling Definition 3 where a sum of exponentially decreasing terms appears on the right side of (39). Similar to Theorem 2, Theorem 3 also uncovers a risk-aware noise-to-state stability property. By denoting the sum of exponentially decreasing terms in (39) by $g_{\nu,x,t}$ and substituting the expression for $b_\nu$ from (41), we can express (39) as follows:
\[ \rho_\nu(\psi(x_t)) \leq g_{\nu,x,t} + \sup_{t \in \mathbb{N}} \left\{ \rho_\nu(\psi(d_t)) + b_{\nu,t} \right\} \]
(43)
for every time $t \in \mathbb{N}$ and initial condition $x \in \mathbb{R}^n$. The risk functional $\rho_\nu$ evaluates the state energy on the left side of (43), and $\rho_\nu$ evaluates the energy of the centered noise $d_t$ on the right side of (43). Under the assumptions of Theorem 3, $g_{\nu,x,t}$ decays to zero as $t \to \infty$, and therefore, Theorem 3 provides the following risk-aware stability bound:
\[ \lim_{t \to \infty} \sup_{t \in \mathbb{N}} \rho_\nu(\psi(x_t)) \leq \sup_{t \in \mathbb{N}} \left\{ \rho_\nu(\psi(d_t)) + b_{\nu,t} \right\}. \]
(44)

Now, let us discuss the assumptions of Theorem 3. For any random variable $Y$ on $(\Omega, \mathcal{F}, P)$, $E(Y^4)$ is finite in many cases, including when $Y$ has a Gaussian distribution, a distribution with bounded support, a log-normal distribution (which is heavy-tailed), or a mixture of these distributions. Using $R_\nu \in \mathcal{S}^+_N$, the second assumption of Theorem 3 is equivalent to $A$ being Schur stable (Lemma 1 Appendix). Hence, one is not an eigenvalue of $A$, which ensures that $I_n - A^{t+1}$ is invertible for every $t \in \mathbb{N}_0$. In the case of $\bar{w}_t = \bar{w}$ for every $t \in \mathbb{N}_0$, it follows that $\bar{w}_t = (I_n - A)^{-1}\bar{w}$, and so $\sup_{t \in \mathbb{N}_0} |\bar{w}_t|^2$ is finite (more details to be provided in the proof). We will prove Theorem 3 next.

Proof: Let $\nu \in [0, \infty)$ be given, and let $x \in \mathbb{R}^n$ be an arbitrary initial condition of the system (19). Unrolling the
recursion provides \( x_{t+1} = A^{t+1}x_0 + \sum_{i=0}^{t} A^i w_{t-i} \) for every \( t \in \mathbb{N}_0 \), and thus,

\[
x_{t+1} = A^{t+1}x_0 + \sum_{i=0}^{t} A^i \bar{w}_{t-i} + \sum_{i=0}^{t} A^i (w_{t-i} - \bar{w}_{t-i})
\]

\[
= A^{t+1}x_0 + (I_n - A^{t+1})\bar{w}_t + \sum_{i=0}^{t} A^i d_{t-i+1},
\]

where the first line of (38) specifies the definition of \( \bar{w}_t \). That is, for every \( t \in \mathbb{N}_0 \), we have the “centered” solution

\[
x_{t+1} - \bar{w}_t = A^{t+1}(x_0 - \bar{w}_t) + \sum_{i=0}^{t} A^i d_{t-i+1}.
\]

(45)

It is convenient to define the “centered” state

\[
\hat{x}_{t+1} := x_{t+1} - \bar{w}_t \iff x_{t+1} = \hat{x}_{t+1} + \bar{w}_t
\]

(47)

for every \( t \in \mathbb{N}_0 \). Additionally, the mean of \( x_{t+1} \) is given by

\[
E(x_{t+1}) = A^{t+1}x + \sum_{i=0}^{t} A^i \bar{w}_{t-i} = A^{t+1}(x - \bar{w}_t) + \bar{w}_t
\]

for every \( t \in \mathbb{N}_0 \). As a special case, in the setting of mean-stationary noise, i.e., \( \bar{w}_t = \bar{w} \) for every \( t \in \mathbb{N}_0 \), we obtain

\[
E(x_{t+1}) = A^{t+1}x + \sum_{i=0}^{t} A^i \bar{w}
\]

\[
= A^{t+1}x + (I_n - A^{t+1})(I_n - A)^{-1}\bar{w},
\]

(49)

and so \( \bar{w}_t = (I_n - A)^{-1}\bar{w} \). (\( A \) being Schur stable implies \( \sum_{i=0}^{t} A^i = (I_n - A^{t+1})(I_n - A)^{-1} \) [57, Ex. 5.6.P26].) This case may be analyzed as well and provide improved constants, but we do not pursue it further for the sake of generality.

Now, let us study the stability of the system with respect to \( \rho_x \). For convenience, we define \( R_x := R + 4\nu RSuR \) and \( C_t \) for every \( t \in \mathbb{N} \) by

\[
C_t := \nu \delta_t - 4\nu \text{var}(\Sigma_t R)^2 - 4\nu E(d_t^T R \Sigma_t R d_t),
\]

(50)

where \( \text{var} \) denotes variance. For every \( t \in \mathbb{N} \), we have (see Lemma 8 in the Appendix for the first line)

\[
\rho_x(\psi(x_t)) = E(x_t^T (R + 4\nu R \Sigma_u R) x_t + 4\nu x_t^T R \gamma_t) + C_t
\]

\[
= E(x_t^T (R + 4\nu R \Sigma_u R) x_t + 4\nu \gamma_t^T R E(x_t) + C_t
\]

\[
\leq E(x_t^T (R + 4\nu R \Sigma_u R) x_t + 4\nu \gamma_t^T R E(x_t) + C_t
\]

\[
= E(x_t^T \bar{R}_x x_t) + 4\nu \gamma_t^T R E(x_t) + C_t,
\]

(51)

where we have used the assumption that \( \Sigma_u \in \mathcal{S}_n \) satisfies \( \Sigma_u - \Sigma_t \in \mathcal{S}_n \) for every \( t \in \mathbb{N} \).

The risk \( \rho_x(\psi(x_t)) \) has two distinct dynamical terms whose stability requires analysis, i.e., the quadratic term \( E(x_t^T \bar{R}_x x_t) \) and the cross-term \( \gamma_t^T R E(x_t) \). For the quadratic term, we use (46) to write, for every \( t \in \mathbb{N} \),

\[
E(x_t^T \bar{R}_x x_t) \leq 2E(x_t^T \bar{R}_x \hat{x}_t) + 2\bar{w}_{t-1}^T R_x \bar{w}_{t-1}.
\]

(52)

For the expectation on the right side of (52), we apply (46) to derive (note that the noise process \( d_t \) is zero-mean)

\[
E(\hat{x}_t^T R_x \hat{x}_t)
\]

\[
= E \left( (x_0 - \bar{w}_{t-1})^T [A^t]^T R_x A^t (x_0 - \bar{w}_{t-1}) \right)
\]

\[
+ \left[ \sum_{i=0}^{t-1} A^i d_{t-i} \right]^T R_x \left[ \sum_{i=0}^{t-1} A^{i-1} d_{t-i} \right]
\]

\[
\leq E \left( 2x_0^T [A^t]^T R_x A^t x_0 + 2\bar{w}_{t-1}^T [A^t]^T R_x A^t \bar{w}_{t-1} \right)
\]

\[
+ \left[ \sum_{i=0}^{t-1} A^i d_{t-i} \right]^T R_x \left[ \sum_{i=0}^{t-1} A^{i-1} d_{t-i} \right]
\]

\[
= E(\hat{x}_t^T R_x \hat{x}_t) + 2\bar{w}_{t-1}^T [A^t]^T R_x A^t \bar{w}_{t-1},
\]

(53)

for every \( t \in \mathbb{N} \), where the process \( (\hat{x}_1, \hat{x}_2, \ldots) \) is defined by

\[
\hat{x}_{t+1} = A\hat{x}_t + d_{t+1}, \quad t \in \mathbb{N}_0,
\]

(55)

where we define \( \hat{x}_0 := \sqrt{2}x_0 \). Now, consider a Schur stable deterministic system defined by \( z_{t+1} = A z_t \) for every \( t \in \mathbb{N}_0 \) with an arbitrary initial state \( z_0 = z \in \mathbb{R}^n \). It holds that

\[
0 \leq \bar{w}_{t-1}^T [A^t]^T R_x A^t \bar{w}_{t-1} \leq |\bar{w}_{t-1}|^2 \sup_{|z|=1} z^T R_x z_t
\]

(57)

for every \( t \in \mathbb{N} \). Consequently, we have shown that

\[
E(\hat{x}_t^T R_x \hat{x}_t) \leq E(\hat{x}_t^T R_x \hat{x}_t) + 2|\bar{w}_{t-1}|^2 \sup_{|z|=1} z^T R_x z_t
\]

(58)

for every \( t \in \mathbb{N} \). Next, let us study the cross term \( \gamma_t^T R E(x_t) \).

\[\text{For any } t \in \mathbb{N}, \text{ it holds that}\]

\[
\bar{w}_{t-1}^T [A^t]^T R_x A^t \bar{w}_{t-1} = |R_x^{1/2} A^t \bar{w}_{t-1}|^2
\]

\[
\leq \left( |R_x^{1/2} A^t \bar{w}_{t-1}|^2 \right)^2
\]

\[
= |\bar{w}_{t-1}|^2 \left( \sup_{|z|=1} |R_x^{1/2} A^t z|^2 \right)^2
\]

\[
= |\bar{w}_{t-1}|^2 \sup_{|z|=1} |R_x^{1/2} A^t z|^2
\]

\[
= |\bar{w}_{t-1}|^2 \sup_{|z|=1} z^T [A^t]^T R_x A^t z
\]

\[
= |\bar{w}_{t-1}|^2 \sup_{|z|=1} z^T R_x z_t
\]

(56)
in (51). For any $t \in \mathbb{N}$, we use (48) to find
\[
\gamma_t^\top R E(x_t) = \gamma_t^\top R A^\top (x - \bar{w}_{t-1}) + \gamma_t^\top R \bar{w}_{t-1}
\leq |R|^{1/2} \gamma_t |A|^{1/2} (x - \bar{w}_{t-1}) + \gamma_t^\top R \bar{w}_{t-1}
= |R|^{1/2} \gamma_t \sqrt{|A|} |\bar{w}_{t-1}|^2 + |R|^{1/2} \gamma_t \bar{w}_{t-1},
\]
where $x_t^d = A x_t^d$ for every $t \in \mathbb{N}_0$ with initialization $x_0^d = \sqrt{2} x$ is another Schur stable deterministic system. Overall, for every $t \in \mathbb{N}$, we have shown that
\[
\rho_\nu(\psi(x_t)) \leq 2 E(\tilde{x}_t^\top R \bar{w}_t x_t) + 4 |\bar{w}_{t-1}|^2 \sup_{|z|=1} z_t^\top R z_t
+ 4 |R|^{1/2} \gamma_t \left| \sqrt{|A|} \bar{w}_{t-1}^2 + 2 |\bar{w}_{t-1}|^2 \sup_{|z|=1} z_t^\top R z_t
+ 2 \bar{w}_{t-1}^\top R x_{t-1} \bar{w}_{t-1} + 4 \gamma_t^\top R \bar{w}_{t-1} + C_t.
\]
(60)
To complete the proof, we analyze the terms in the right side of (60). For convenience, we define
\[
s_\nu := \sup_{t \in \mathbb{N}} E(d_t^\top H_{R,\nu}^\top d_t), \quad \eta_\nu := \frac{\lambda_{\min}(H_{R,\nu}^\top A^\top H_{R,\nu})}{\lambda_{\max}(H_{R,\nu}^\top A^\top)}.
\]
(61)
and $\lambda_\nu := 1 - \eta_\nu$, and thus, $s_\nu \in [0, \infty)$ and $\lambda_\nu \in [0, 1)$ using the assumptions of Theorem 3 (also see Lemma 2 Appendix). Moreover, we define $\psi_\nu(y) := y^\top R_{\nu} y$ for every $y \in \mathbb{R}^n$ and $a_\nu := \lambda_{\max}(H_{\nu}^\top)/\lambda_{\min}(H_{\nu}^\top)$; note that $a_\nu \in (0, \infty)$ by the second assumption. Using $R \in S^+_n$ and $R_{\nu} \in S^+_n$, the second assumption implies the existence of a matrix $H_{\nu} \in S^+_n$ such that $H_{\nu}^R = A^T H_{\nu} R = A \in S^+_n$ (see Lemma 1 Appendix). We define $R_0 := R$, $a_0 := \lambda_{\max}(H_0^\top)/\lambda_{\min}(H_0^\top)$, $\eta_0$ by the analogous expression in (61) with $\nu = 0$, $\lambda_0 := 1 - \eta_0$, and $\psi_0(y) := y^\top R_{0} y$ for every $y \in \mathbb{R}^n$. By applying the second assumption in particular, the following statements about the terms in (60) hold:

- **The term $\mathbb{E}(\tilde{x}_t^\top R_{\nu} \tilde{x}_t)$:**
\[
\mathbb{E}(\tilde{x}_t^\top R_{\nu} \tilde{x}_t) \leq 2 a_\nu \lambda_\nu^2 \psi_\nu(x) + \frac{s_\nu}{\lambda_{\min}(H_{\nu}^\top)(1 - \lambda_\nu)}, \quad t \in \mathbb{N}.
\]
(62)

- **The term $z_t^\top R_\nu z_t$ for every $\nu \geq 0$:**
\[
z_t^\top R_\nu z_t \leq a_\nu \lambda_\nu \psi_\nu(z), \quad t \in \mathbb{N}, \quad z \in \mathbb{R}^n,
\]
(63)
and therefore, for every $t \in \mathbb{N}$,
\[
\sup_{|z|=1} z_t^\top R_\nu z_t \leq a_\nu \lambda_\nu \sup_{|z|=1} \psi_\nu(z) = a_\nu \lambda_{\min}(R_\nu) \lambda_\nu.
\]
(64)

- **The term $[x_t^d]^\top R x_t^d$:**
\[
[x_t^d]^\top R x_t^d \leq 2 a_0 \lambda_0^2 \psi_0(x), \quad t \in \mathbb{N}.
\]
(65)

By defining
\[
c_\nu := \frac{2}{\lambda_{\min}(H_{\nu}^\top)(1 - \lambda_\nu)},
\]
(66)
\[
\tilde{\psi}_\nu(x) := 4 \psi_\nu(x) + 4 \lambda_{\max}(R_\nu) \sup_{t \in \mathbb{N}} |\bar{w}_{t-1}|^2,
\]
(67)
and $\tilde{\psi}_\nu(x)$ by (67) with $\nu = 0$, we consolidate our previous results as follows:
\[
2 E(\tilde{x}_t^\top R_{\nu} \tilde{x}_t) + 4 |\bar{w}_{t-1}|^2 \sup_{|z|=1} z_t^\top R_\nu z_t \leq a_\nu \lambda_\nu^2 \tilde{\psi}_\nu(x) + c_\nu s_\nu
\]
(68)
and
\[
[x_t^d]^\top R x_t^d + 2 |\bar{w}_{t-1}|^2 \sup_{|z|=1} z_t^\top R_\nu z_t \leq a_0 \lambda_0^2 \tilde{\psi}_0(x)
\]
(69)
for every $t \in \mathbb{N}$. The statements (60), (63), and (69) provide the following (almost final) bound for any $t \in \mathbb{N}$:
\[
\rho_\nu(\psi(x_t)) \leq a_\nu \lambda_\nu^2 \tilde{\psi}_\nu(x) + 4\nu |R|^{1/2} \gamma_t \sqrt{\lambda_{\min}(R_\nu)}
\]
(70)
Exponentially decreasing terms
\[
+ c_\nu s_\nu + 2 |\bar{w}_{t-1}|^2 R_\nu \bar{w}_{t-1} + 4 \gamma_t^\top R \bar{w}_{t-1} + C_t.
\]
Bias terms
Next, we use (50), the definition $R_{\nu,t} := R + 4 \nu R_{\nu} \bar{w}_{\nu,t}$, and the independence of $d_t$ and $h_{\nu,t}$ for any $t \in \mathbb{N}$ to derive
\[
C_t = \nu \text{var}(d_t R_t d_t) + \text{E}(d_t R_t d_t) - \text{E}(d_t^\top R_t d_t)
- 4 \nu \text{E}(d_t^\top R_\nu \bar{w}_{\nu,t} d_t)
= \text{E}(d_t R_t d_t) + \nu \text{var}(d_t R_t d_t) - \text{E}(d_t^\top R_{\nu,t} d_t)
\]
Noise mean-variance
\[
= \rho_\nu(\psi(d_t)) - \text{E}(d_t^\top R_{\nu,t} d_t),
\]
(71)
Finally, we use $|R|^{1/2} \gamma_t \leq \sqrt{\lambda_{\max}(R)} \gamma_t$, (70), and (71) to derive
\[
\rho_\nu(\psi(x_t)) \leq a_\nu \lambda_\nu^2 \tilde{\psi}_\nu(x) + 4 \nu |R|^{1/2} \gamma_t \sqrt{\lambda_{\min}(R_\nu)} \tilde{\psi}_0(x)
\]
Exponentially decreasing terms
\[
+ c_\nu \left( \text{E}(d_t^\top H_{R,\nu}^\top d_t) - \text{E}(d_t^\top R_{\nu,t} d_t) \right)
\]
Risk neutral zero-mean noise bias
\[
+ \rho_\nu(\psi(d_t)) + 2 |\bar{w}_{t-1}|^2 R_{\nu,t} \bar{w}_{\nu,t} + 4 \gamma_t^\top R \bar{w}_{t-1}
\]
Noise mean and skewness terms
(72)
for every $t \in \mathbb{N}$, completing the proof.

Similarities and differences between Theorem 2 and Theorem 3 deserve some discussion. Both theorems show how the risk of the state energy and a risk-aware perception of a disturbance process can be decoupled, although the theorems quantify risk in distinct ways; compare (22) and (43). Theorem 2 applies to any real-valued coherent risk functional on $\mathcal{L}^q$, while Theorem 3 applies to a mean-conditional-variance functional on $\mathcal{L}^1$. The offset parameters from the two theorems each reflect a particular perception of noise, which is induced by the chosen measure of risk; indeed, compare (21) and (41). Theorem 3 provides a rate parameter $\lambda_\nu$, which depends on $\nu$, where $\nu$ is specific to the risk functional $\rho_\nu$. However, the
rate parameter from Theorem 2 is equivalent to the one in the risk-neutral setting. Later on, we will see that Theorem 3 provides insights about the behavior of a simple risk-aware myopic controller, which mitigates extreme peaks of the state energy in simulation (Section IV Illustration 3). In contrast, how to develop a simple risk-aware controller in the setting of Theorem 2 is an open research question.

Ideally, we would like to optimize the parameters provided by Theorem 2 or Theorem 3 by choosing a suitable state-feedback matrix $K$. For example, one can consider the problem of minimizing the offset parameter $b = \sup_{t \in N_0} \phi(w^T_t H_R w_t)$ subject to the assumptions of Theorem 2 where $A = \bar{A} - \bar{B} K$, $\bar{A}$ and $\bar{B}$ are given matrices, and $K$ is to be chosen. This is a difficult risk-aware optimization problem to solve exactly, where the matrix variables $H$ and $K$ are coupled through $A^T H_R A = (\bar{A} - \bar{B} K)^T H_R (\bar{A} - \bar{B} K)$. Moreover, the rate parameter $\lambda$ should be minimized simultaneously, adding to the difficulty. Hence, we reserve investigations of such optimization problems for future work. The next section will provide examples to illuminate aspects of our theoretical developments.

IV. ILLUSTRATIVE EXAMPLES

Illustration 1 (Risk of noise energy): Theorem 2 indicates that the risk of the state energy $x^T t R x_t$ of the system under study is of the order of the corresponding risk of the noise energy $w^T_t H_R w_t$ under some assumptions. (The term noise energy refers to any quadratic form $w^T_t M w_t$ with $M \in S^n_+$.) We consider for simplicity a one-dimensional, Schur stable linear system driven by independent and identically distributed noise $w_t$ (i.e., a stable order-1 autoregressive process). This illustration will examine the risk of $w^2_t$, which arises in the right side of $b'$ in a special case, compared to realizations of $w^2_t$ to demonstrate the need for and the usefulness of evaluating the noise energy from a risk-aware perspective.

Let us select the mean-upper-semideviation of order-$q$ for this illustration. Note that whenever $\beta \in [0,1]$ we obtain a coherent risk functional for every value of $q \geq 1$, with its dual representation being true with the uncertainty set $A$ being given by $\mathcal{D}_0$.

While we may assume a generic nominal noise model, e.g., $w_t \sim \mathcal{U}[-1,1]$, at the same time we may be uncertain about it. In such a case, we would like to obtain a robust estimate of the average energy of the noise with respect to potential distributional modeling uncertainty. This is precisely achieved by using a (coherent) risk functional to evaluate $w_t$ under its nominal model (which arises from characterizing system stability through Theorem 2). To illustrate this, let us consider two alternative realities for $w^2_t$ defined by

$$w^2_{t,i} := w^2_t \xi_{t,i}, \quad i \in \{1,2\}, \quad t \in N_0,$$

where $\xi_{t,i}$ for $i \in \{1,2\}$ are multiplicative random perturbation processes of the form

$$\xi_{t,i} = 1 + \eta_{t,i} - E(\eta_{t,i}), \quad i \in \{1,2\},$$

where, for every $t \in N_0$,

$$\eta_{t,1} \sim \mathcal{U}[0,1] \quad \text{and} \quad \eta_{2,t} = z_t I_{[0,1]}(z_t) + I_{>1}(z_t),$$

with $z_t = -w_t + 1/4$. In other words, each $\xi_{t,i}$ for $i \in \{1,2\}$ belongs to the uncertainty set $\mathcal{D}_0$ associated with the order-$q$ mean-upper-semideviation risk functional for every $q \geq 1$. Observe that $\xi_{t,1}$ is a uniform perturbation and independent of $w_t$, whereas $\xi_{t,2}$ is nonuniform and highly dependent on $w_t$. Still, both $\xi_{t,1}$ and $\xi_{t,2}$ are independent over different values of $t$ (and thus the same happens for $w^2_{t,i}$, $i \in \{1,2\}$, which is helpful in numerical simulation). It follows that, for $\beta \in [0,1]$ and $q \geq 1$,

$$E(w^2_{t,i}) \leq \sup_{\xi \in \mathcal{A}} E(w^2_t \xi) = \text{MUS}_{q,\beta}(w^2_t), \quad i \in \{1,2\},$$

illustrating that the mean-upper-semideviation risk functional (which is just one example) can provide worst-case risk-neutral estimates of the noise energy over a particular but rich class of possible realities, including the ones described above.

Figure 1 illuminates the above facts by depicting the time trajectories and corresponding empirical CDFs of the noise.
energy in all three cases. Figure 1 also shows the risk-neutral estimate ($\beta = 0$) and two risk-aware estimates ($\beta = 1$, $q = 2$ and $\beta = 1$, $q = 12$). We have computed the risk-aware estimates empirically by using the primal representation of mean-upper-semideviation under the nominal noise model (that is, no information about any possible alternative reality is needed). As anticipated, the risk-neutral estimate (expectation) masks the potential dispersion of the noise energy (in fact, in every case), with the issue being more pronounced for the alternative reality, and especially for the alternative reality whose noise distribution exhibits a fatter tail. In contrast, the risk-aware estimates are biased toward more extreme noise energy realizations (especially in the case of $q = 12$), which equivalently (via risk duality) provide uniformly cautious estimates for a variety of possible realities, including $w^2_{t,1}$ and $w^2_{t,2}$, due to the richness of the uncertainty set (12). These facts are readily apparent by observing the noise energy trajectories (along with the horizontal lines) and the empirical CDFs (along with the vertical lines) in Figure 1.

Illustration 2 (Trade-offs related to $\kappa$): Here, we will illustrate trade-offs related to the parameter $\kappa$ from Theorem 2. Let us consider the case of $\eta < 1$, where the matrices $R$, $A$, and $H$ have been chosen. The parameters $\lambda$ and $c$ depend on an adjustable parameter $\kappa \in (0, 1)$. Since $\lambda = 1 - \kappa\eta$ (20), where $\eta \in (0, 1)$ is determined by $R$, $A$, and $H$, choosing $\kappa$ to be closer to 1 would provide faster decay. However, we would also like $c$ to be small to reduce noise effects. (Recall that $b = cb'$ (20), where $b'$ (21) is the supremum of the risk of the noise energy $w^t H w_t$.) There is a nonlinear relationship between $c$ and $\kappa$, that is, by direct substitution, we obtain

$$c = \frac{1 - \kappa\eta}{\lambda_{\text{min}}(H) \kappa\eta (\eta - \kappa\eta)}.$$  

(78)

Figure 3 depicts $c \lambda_{\text{min}}(H)$ versus $\kappa$ for several values of $\eta \in (0, 1)$, illustrating the effect of $\eta$ on the minimum value of $c \lambda_{\text{min}}(H)$ with respect to $\kappa$.

Illustration 3 (Insights about controllers using Theorem 3): An exciting potential use case of the risk-aware stability bounds we have developed is to inform the design of risk-aware controllers. The mean-conditional-variance functional studied in Theorem 3 turns out to be particularly useful for this purpose, mainly due to its quadratic form. To demonstrate this, suppose we are given a linear system of the form

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad t \in \mathbb{N}_0,$$

(79)
where, for simplicity, we assume that the noise process \( \tilde{w}_t \) is stationary but with possibly nontrivial dispersive behavior (to be described). For each \( t \in \mathbb{N}_0 \), we consider the myopic (one-step-ahead) regularized optimal control problem

\[
\min_{u_t \in \mathbb{R}^{m}} \quad \rho_\nu(\psi(x_{t+1})|x_t) + \mu u_t^\top u_t, \quad \mu \geq 0,
\]

where \( \rho_\nu(\cdot|\cdot) \) denotes the conditional version of \( \rho_\nu \) (note that \( \rho_\nu \) admits an expectation representation, see the first line of (51)), when the state \( x_t \) at time \( t \) is given. This problem is equivalent to the quadratic program

\[
\min_{u_t \in \mathbb{R}^{m}} \quad E(x_{t+1}^\top R \nu x_{t+1} + 4\nu x_{t+1}^\top R \gamma(x_t) + \mu u_t^\top u_t),
\]

where \( \gamma_t = \gamma \) and \( \Sigma_t = \Sigma_u \) for every \( t \). Provided that \( \mu I_n + \hat{B}^\top R \nu \hat{B} \) is invertible, the solution is

\[
u_t^{\kappa,\nu} = -K_\nu^\kappa x_t - T_\nu^\kappa \gamma_t, \quad t \in \mathbb{N}_0,
\]

where the inflated gain \( K_\nu^\kappa \) and tail bias \( T_\nu^\kappa \) are given by

\[
K_\nu^\kappa = (\mu I_n + \hat{B}^\top R \nu \hat{B})^{-1} \hat{B}^\top R \nu \hat{A} \quad \text{and}
\]

\[
T_\nu^\kappa = 2\nu(\mu I_n + \hat{B}^\top R \nu \hat{B})^{-1} \hat{B}^\top R.
\]

Observe that, when \( \nu = 0 \), we obtain the standard myopic linear-quadratic-regulator controller \( u_t^{\kappa,\nu} = -K_0^\kappa x_t \). Also, note that the risk-aware controller \( u_t^{\kappa,\nu} \) is distinct from the controller recently developed in [44], the latter being nonmyopic and derived via stochastic dynamic programming.

By substituting the stationary controller (82) into (79), we obtain the closed-loop system

\[
x_{t+1} = (\hat{A} - \hat{B} K_\nu^\kappa) x_t + (\tilde{w}_t - \hat{B} T_\nu^\kappa \gamma_t)
\]

\[
= A x_t + \tilde{w}_t, \quad t \in \mathbb{N}_0,
\]

where, in contrast to the risk-neutral case (\( \nu = 0 \)), the risk-aware controller not only regulates the state (in a risk-aware sense through the inflated gain \( K_\nu^\kappa \), driving the state away from directions with high second-order variability, as captured by \( R_\nu \)), but also shifts the noise by a quantity proportional to its third-order moment behavior, as captured by the statistic \( \gamma \). Essentially, the risk-aware controller optimally "de-biases" the skewed behavior of \( \tilde{w}_t \) by simply shifting its mean. This implies that the shifted noise \( \tilde{w}_t \) has the same central statistics (in particular, \( \gamma, \delta, \) and covariance) as the original noise \( \tilde{w}_t \).

Figure 2 shows the time trajectories and empirical CDFs of the state energy and the respective control effort (where applicable) corresponding to no controller (\( u_t \) is zero), the risk-neutral controller for \( \mu = 0.25 \), and the risk-aware controller for \( \mu = 0.25 \) and \( \nu = 10 \), for a Schur stable system of the form of (79) by choosing

\[
\hat{A} = \begin{bmatrix} 0.8 & 0.4 \\ 0 & -0.8 \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \text{and} \quad R = I_2,
\]

and where \( \tilde{w}_t \) follows a Gaussian mixture distribution

\[
\tilde{w}_t \overset{\text{iid}}{\sim} 0.7 \times \mathcal{N}(0, I_2) + 0.3 \times \mathcal{N}(\begin{bmatrix} 2 & 15 \end{bmatrix}^\top, 10 I_2), \quad t \in \mathbb{N}_0.
\]

In other words, for 70% of the time, the additive disturbance to the system is standard Gaussian, while for the remaining 30% of the time, the system exhibits abrupt Gaussian shocks with large mean and variance in both state coordinates.

We observe that, for the same level of control regularization (\( \mu = 0.25 \)), the risk-aware controller offers a dramatic improvement on the state energy \( |x_t|^2 \) in terms of stabilizing its statistical variability (i.e., risky behavior) as compared to its risk-neutral counterpart, and it is particularly effective in mitigating the (more infrequent) shocks due to the highly dispersive noise \( \tilde{w}_t \). Additionally, the control effort \( |u_t|^2 \) of the risk-aware controller is significantly smaller in magnitude, however less sparse and more persistent as compared to the risk-neutral controller. This is explained due to the strategically designed affine form of the risk-aware controller, which provides increased degrees of freedom, allowing more effective state regulation.

Interestingly, the risk-aware controller \( u_t^{\kappa,\nu} \) exhibits disturbance attenuation behavior in the sense of the risk functional \( \rho_\nu \). Indeed, such an effect can be readily observed through an empirical application of Theorem 3, which also reveals the usefulness of Theorem 3 for informing and evaluating different controller designs. To demonstrate this, in Figure

\[\text{Fig. 3. Plots of } c \lambda_{\text{min}}(H) \text{ vs } \kappa \text{ for } \eta \in (0,1) \text{ (see Theorem 2 in the case of } \eta < 1)\].
we report, for each value of $\nu$, the ratios of the biases $b_{\nu}$ (respectively, decay rates $\lambda_{\nu}$) achieved by using the risk-aware controller $u_{\nu}^{\text{rw}}$ over the risk-neutral controller $u_{\nu}^{\text{rn}}$. From Figure 4 (bottom), we observe that, as $\nu$ increases, the corresponding bias-ratio consolidates sharply to a value roughly equal to 0.6. In the context of Theorem 3, this finding implies a reduction in the bias term $b_{\nu}$ of our stability bound as much as 40% (roughly achieved for $\nu = 10$) compared to the bias achieved by the standard risk-neutral controller. Such a reduction demonstrates a drastic disturbance attenuation effect exhibited by the risk-aware controller $u_{\nu}^{\text{rw}}$. Lastly, we observe a similar decreasing trend for the corresponding rate-ratio (see Figure 4 (top)). Hence, the risk-aware controller also improves (albeit slightly) the decay rate of the exponentially decreasing terms of the bound of Theorem 3.

V. Conclusion

By investigating a generalized risk-aware stability viewpoint, we have discovered conditions that guarantee new risk-aware noise-to-state stability properties for linear systems. In the case of any real-valued coherent risk functional on $\mathcal{L}^q$, the long-term risk of the state energy is on the same order as the supremum of the risk of the noise energy. In the case of a mean-conditional-variance functional, a similar relationship appears in [44], and additionally, the noise-dependent bias term $b_{\nu}$ can be attenuated by a simple risk-aware controller (Illustration 3). In the future, we plan to investigate extensions to nonlinear systems, including systems with incomplete or misspecified models using statistical learning techniques. We are particularly excited about extending the risk-aware stability theory developed here to enrich the analysis and design of stochastic gradient descent algorithms.

Appendix

The following lemma can be viewed as a corollary of the classical discrete-time Lyapunov stability theorem for deterministic linear systems [59, Th. 7.3.2].

Lemma 1: Let $R \in \mathbb{S}_n^+$ and $A \in \mathbb{R}^{n \times n}$ be given. The existence of an $H \in \mathbb{S}_n^+$ such that $H - A^T H R A \in \mathbb{S}_n^+$ is equivalent to $A$ being Schur stable.

Proof: We will show one direction. (The omitted direction uses $H \in \mathbb{S}_n^+$ and involves left-multiplying by $v^*$ and right-multiplying by $v$, where $v \in \mathbb{C}^n$ is an eigenvector of $A$.) Now, $A$ is Schur stable if and only if, for any $Q \in \mathbb{S}_n^+$, there is a unique $X \in \mathbb{S}_n^+$ such that $X = -A^T X A = Q$ [59, Th. 7.3.2]. Assume that $A$ is Schur stable, and let $Q \in \mathbb{S}_n^+$ be given. Since $R^\dagger$ is nonsingular, consider $H := (-X^\dagger)^T X R^\dagger$. Since $R^\dagger$ is nonsingular and $X \in \mathbb{S}_n^+$, $H \in \mathbb{S}_n^+$ holds as well. Moreover, since $H R = X$, we have shown the existence of an $H \in \mathbb{S}_n^+$ such that $H - A^T H R A \in \mathbb{S}_n^+$.

Lemma 2: Let $R \in \mathbb{S}_n^+$ and $A \in \mathbb{R}^{n \times n}$ be given. Suppose that there is an $H \in \mathbb{S}_n^+$ such that $H - A^T H R A \in \mathbb{S}_n^+$. Then, $\eta := \lambda_{\max}(H - A^T H R A) / \lambda_{\min}(H + R A) \in (0,1]$.

Proof: The property $\eta \in (0,\infty)$ holds because $H - A^T H R A \in \mathbb{S}_n^+$ and $H \in \mathbb{S}_n^+$. If $\lambda_{\max}(H - A^T H R A) \leq \lambda_{\max}(H R)$, then $\eta \leq 1$, so it suffices to show (88). Since $A^T H R A \in \mathbb{S}_n^+$, the inequality (88) follows from the eigenvalue monotonicity theorem [57, Cor. 4.3.12].

Lemma 3: For any $M \in \mathbb{S}_n$, $y \in \mathbb{R}^n$, and $z \in \mathbb{R}^n$, and for every $\epsilon \in (0,\infty)$, it is true that

$$(y + z)^T M (y + z) \leq (1 + \epsilon) y^T M y + (1 + \frac{\epsilon}{2}) z^T M z. \quad (89)$$

Lemma 3 is standard, and so we omit the proof.

Lemma 4: Let $\mathcal{A}$ be the family of densities in the dual representation (9) of a real-valued coherent risk functional $g$ on $\mathcal{L}^q$ with $q \in [1,\infty)$. Consider the linear system (19) described in the first paragraph of Section III and assume that $|w_t|^2 \in \mathcal{L}^q$ for every $t \in \mathbb{N}_0$.

1) For any $t \in \mathbb{N}_0$, $\xi \in \mathcal{A}$, and $M \in \mathbb{S}_n$, $x_t^T M x_t \in \mathcal{L}^q$ and $x_t^T M x_t \xi \in \mathcal{L}^q$, and these functions are a.e.-nonnegative.

2) Given $H \in \mathbb{S}_n^+$ and $R \in \mathbb{S}_n^+$, define $\psi(z) := z^T R z$ and $\nu(x) := z^T H R z$ for every $z \in \mathbb{R}^n$. For every $\xi \in \mathcal{A}$ and $t \in \mathbb{N}_0$, the statement (29) holds.

Proof: Part 1: $x_t^T M x_t$ is a sum of finitely many $(\mathcal{F},\mathcal{B}_R)$-measurable functions, and so it is $(\mathcal{F},\mathcal{B}_R)$-measurable. The property of $|x_t|^2 \in \mathcal{L}^q$ for every $t \in \mathbb{N}_0$ holds by induction, where we use $|w_t|^2 \in \mathcal{L}^q$ and $x_{t+1} = Ax_t + w_t$ with $x_0$ being fixed. In addition, we use

$$0 \leq |x_{t+1}(\omega)|^2 \leq 2\lambda_{\max}(A^T A)|x_t(\omega)|^2 + 2|w_t(\omega)|^2 \quad (90)$$

for every $\omega \in \Omega$ (see Lemma 3) and $y \mapsto y^\top y$ being nondecreasing on $[0,\infty)$ for any $\gamma \in (0,\infty)$ to find

$$0 \leq |x_{t+1}|^2 \leq 0 \leq 2\lambda_{\max}(A^T A)|x_t|^2 + 2|w_t|^2 \quad (91)$$

The property of $x_t^T M x_t \in \mathcal{L}^q$ and $x_t^T M x_t \xi \in \mathcal{L}^q$, $y \mapsto y^\top y$ being nondecreasing on $[0,\infty)$ for any $\gamma \in (0,\infty)$, $|x_t|^2 \in \mathcal{L}^q$, and $\lambda_{\max}(M)|x_t|^2 \xi \in \mathcal{L}^q$.

Now, $\xi \in \mathcal{A}$ implies that $\xi \geq 0$ a.e. and $\xi \in \mathcal{L}^q$. Hence, $x_t^T M x_t \xi \geq 0$ a.e. and $|x_t^T M x_t |\xi |1 \leq |x_t^T M x_t |\xi |\xi |_{\mathcal{N}^q} < \infty$ [58, Hölder’s Inequality, Th. 6.8 (a)].

Part 2: Since $H \in \mathbb{S}_n^+$,

$$0 \leq \lambda_{\min}(H)|z|^2 \leq z^T H z \leq \lambda_{\max}(H)|z|^2, \quad z \in \mathbb{R}^n. \quad (92)$$

Given $y \in \mathbb{R}^n$, consider $z = R^\dagger y$ in (92) to find

$$0 \leq \lambda_{\min}(H)y^T R y \leq y^T H R y \leq \lambda_{\max}(H)y^T R y, \quad (93)$$

where we use $R = (R^\dagger)^T R^\dagger$ and $H R = (R^\dagger)^T H R^\dagger$. We substitute $\psi(y) = y^T R y$ and $\nu(y) = y^T H R y$ to find

$$0 \leq \lambda_{\min}(H)\psi(y) \leq \nu(y) \leq \lambda_{\max}(H)\psi(y). \quad (94)$$

Now, let $\xi \in \mathcal{A}$ and $t \in \mathbb{N}_0$ be given. Since (29) holds for every $y \in \mathbb{R}^n$ and since $\xi \geq 0$ a.e.,

$$0 \leq \lambda_{\min}(H)\psi(x_t)\xi \leq \nu(x_t)\xi \leq \lambda_{\max}(H)\psi(x_t)\xi \quad \text{a.e.} \quad (95)$$

The statement (29) follows from (95) and basic integration properties [52, Th. 1.5.9, see also p. 47].

Lemma 5: Suppose that $\lambda \in (0,1)$, $\nu \in [0,\infty)$, and $(s_t)_{t \in \mathbb{N}_0}$ is a sequence in $[0,\infty)$ such that $s_{t+1} \leq \lambda s_t + \nu$ for every $t \in \mathbb{N}_0$. Then, $0 \leq s_t \leq \lambda^t s_0 + \nu \sum_{i=0}^{t-1} \lambda^i$ for every $t \in \mathbb{N}$.

Proof: By induction, $0 \leq s_t \leq \lambda^t s_0 + \nu \sum_{i=0}^{t-1} \lambda^i$ for every $t \in \mathbb{N}$. Since $\lambda \in (0,1)$ and $\nu \in [0,\infty)$, we use the geometric series formula to find that $\nu \sum_{i=0}^{t-1} \lambda^i \leq \frac{\nu}{1-\lambda}$. \hfill ■
Lemma 6: Let $R \in S^n_+$ and $A \in \mathbb{R}^{n \times n}$ be given. Suppose that there is an $H \in S^n_+$ such that $H R - A^T H R A \in S_n$. Recall that $\eta := \lambda_{\min}(HR - A^T H R A)/\lambda_{\max}(HR)$. Then, $\eta \in [0, 1]$ and $0 \leq z^T H R z \leq (1 - \eta) z^T R z$ for every $z \in \mathbb{R}^n$.

Proof: Since $H R - A^T H R A \in S_n$ and $A^T H R A \in S_n$, we have

$$0 \leq \lambda_{\min}(HR - A^T H R A) \leq \lambda_{\max}(HR)$$

by applying the eigenvalue monotonicity theorem [57, Cor. 4.3.12]. Since $H R \in S^n_+$, we divide by $\lambda_{\max}(HR) \in (0, \infty)$ to find $0 \leq \eta \leq 1$. Now, let $z \in \mathbb{R}^n$ be given. We multiply the inequalities $0 \leq z^T H R z \leq \lambda_{\max}(HR) |z|^2$ by $\eta$ to find

$$0 \leq \eta z^T H R z \leq \lambda_{\min}(HR - A^T H R A)|z|^2.$$  

We use this and $H R - A^T H R A \in S_n$ to derive

$$\eta z^T H R z \leq z^T (HR - A^T H R A) z.$$  

Then, we use $A^T H R A \in S_n$, and we add and subtract by $z^T H R z$ to find

$$0 \leq z^T A^T H R A z = z^T (H R - A^T H R A) z + z^T H R z \leq -\eta z^T H R z + z^T H R z.$$  

We have applied this in the last line to complete the proof.

Lemma 7: Let the assumptions of Theorem 2 hold, and suppose that $\eta < 1$. Then, $E(v(x_{t+1})) \leq \lambda E(v(x_t)) + \frac{\lambda \eta}{1 - \eta} v^T$ for every $x \in A$ and $t \in \mathbb{N}_0$.

Proof: Let $x \in A$ and $t \in \mathbb{N}_0$ be given, and recall that $\xi \geq 0$ almost everywhere. We use $v(x_{t+1}) = v(x_t) + H R x_t$, $H R \in S^n_+$, $x_{t+1} = x_t + w_t$, and Lemma 5 to derive

$$v(x_{t+1}) \leq (1 + \epsilon) v(x_t) + \frac{1 + \epsilon}{\lambda} w^T H R w$$

for any $\epsilon \in (0, \infty)$. Multiplying (100) by $\xi$, taking expectations, and using that $E(w^T H R w) \leq b'$ from (21) lead to

$$E(v(x_{t+1})) \leq (1 + \epsilon) E(v(x_t) A^T H R A x_t) + \left(1 + \frac{\epsilon}{\lambda}\right) b'.$$

Lemma 3 helps circumvent the issue of the cross term $E(x_t^T A^T H R A x_t)$ need not being zero. Due to (101) and $\lambda = 1 - \kappa \eta$ (20) for any fixed $\kappa \in (0, 1)$, it suffices to show that $E(x_t^T A^T H R A x_t) \leq (1 - \eta) E(v(x_t))$, which readily follows in particular from Lemma 6 and $\xi$ being nonnegative almost everywhere. To see this, note that

$$(1 + \epsilon)(1 - \eta) < 1 \iff \epsilon < \frac{\eta}{1 - \eta}.$$  

By choosing $\epsilon = (1 - \kappa) \eta/(1 - \eta)$, we obtain

$$\lambda \left(1 + (1 - \kappa) \frac{\eta}{1 - \eta}\right)(1 - \eta) = 1 - \kappa \eta.$$  

Using the same value for $\epsilon$ produces the quantity multiplying $b'$ in the statement of Lemma 7.

Lemma 8: Consider the linear system (19), and let $R \in S^n_+$ and $\nu \in [0, \infty)$ be given. Consider the quadratic state function $\psi(x) := x^T Rx$. Under the assumptions of Theorem 3 for every $x \in \mathbb{N}$, the following equality holds

$$\rho_v(\psi(x_t)) = E(x_t^T (R + 4\nu R \Sigma_t R) x_t + 4\nu x_t^T R \gamma_t) + C_t,$$

where $C_t := \nu \delta_t - 4\nu \text{tr}(\Sigma_t R)^2$, and $\rho_v(\psi(x_t))$ is finite.

Proof: The proof is inspired by the proof of [44, Prop. 1] with additional steps to circumvent expressions of the form $\infty - \infty$. Let $t \in \mathbb{N}$ be given. The assumptions of Theorem 3 imply that $E(w_{t-1}) = \bar{w}_{t-1} \in \mathbb{R}$, $E(d_t^T) = \gamma_t \in S_n$, $\gamma_t \in \mathbb{R}^n$, and $\delta_t \in [0, \infty)$.

We use standard measure-theoretic principles, including H"older’s inequality [52, Th. 2.4.5] and Minkowski’s inequality [52, Th. 2.4.7]. We do not use sup_{t \in \mathbb{N}} |w| < \infty, sup_{t \in \mathbb{N}} |\gamma| < \infty, or sup_{t \in \mathbb{N}} \delta_t < \infty. The conditional expectation $\hat{\chi}_t := E(x_t | F_{t-1})$ takes values in $\mathbb{R}^n$ and is $(\mathcal{F}_{t-1}, \mathbb{B}_{\mathbb{R}^n})$-measurable [52, Th. 6.4.3, Th. 1.5.8]. One should be cautious about using $\hat{\chi}_t R \hat{\chi}_t$ because it may take the ill-defined form of $\infty - \infty$. Hence, we will alter $\hat{\chi}_t$ on a set of measure one. Since $E(x_t) \in \mathbb{R}^n$, $\hat{\chi}_t$ is $\mathbb{R}^n$-valued almost everywhere [52, Th. 6.5.4 (a), Th. 1.6.6 (a)), and thus, the set $B_t := \{ \omega \in \Omega : \hat{\chi}_t(\omega) \in \mathbb{R}^n \}$ satisfies $P(B_t) = 1$. Moreover, $B_t \in \mathcal{F}_{t-1}$ because $\hat{\chi}_t$ is $(\mathcal{F}_{t-1}, \mathbb{B}_{\mathbb{R}^n})$-measurable. We define $\tilde{z}_t := I_{B_t} \hat{\chi}_t$. Hence, $\tilde{z}_t$ is $(\mathcal{F}_{t-1}, \mathbb{B}_{\mathbb{R}^n})$-measurable, $\tilde{z}_t(\omega) \in \mathbb{R}^n$ for every $\omega \in \Omega$, and $\tilde{z}_t \equiv \hat{\chi}_t$ almost everywhere. We use properties of conditional expectations [52, Ch. 6.5], $(\mathcal{F}_{t-1}, \mathbb{B}_{\mathbb{R}^n})$-measurability of $x_{t-1}$, $x_t = A x_{t-1} + w_{t-1}$, and independence of $w_{t-1}$ and $h_{t-1}$ to find that $\tilde{z}_t = x_t - d_t$ almost everywhere. Moreover, we derive that $E(x_t^T R x_t | F_{t-1}) = \tilde{z}_t^T R \tilde{z}_t + tr(\Sigma_t R)$ almost everywhere.

Next, we apply the previous two statements, the definition of $\Delta_t$ (14) with $Z = x_t^T R x_t$ and $F_t = F_{t-1}$, and $\tilde{z}_t^T R \tilde{z}_t$ being finite to find that $\Delta_t = \Delta_t a.e.$, where $\Delta_t$ is defined by

$$\Delta_t := tr(\Sigma_t R) - 2\tilde{z}_t R d_t - d_t^T R d_t.$$  

In addition, all terms of $\Delta_t^2$ are integrable. In particular, note that $\tilde{z}_t$ and $d_t$ are independent and $E(d_t) = 0_a$. Then, we use $E(\Delta_t^2) = E(\tilde{z}_t^2)$ and $\tilde{z}_t = x_t - d_t$ a.e. to find

$$E(\Delta_t^2) = 4E(x_t^T R \Sigma_t R x_t) + 4E(x_t^T R \gamma_t) + \delta_t - 4tr(\Sigma_t R)^2,$$

which is finite. Finally, using $\rho_v(\tilde{z}_t^T R x_t) = E(x_t^T R x_t) + \nu E(\Delta_t^2)$ completes the proof.

Acknowledgement

M.P.C. would like to gratefully acknowledge Erick Mejia Uzeda for discussions.

References

[1] J. E. Bertram and P. E. Sarachik, “Stability of circuits with randomly time-varying parameters,” IEEE Transactions on Circuit Theory, vol. 6, pp. 260–270, 1959.

[2] I. Y. Kats and N. N. Krasovskii, “On the stability of systems with random attributes,” Journal of Applied Mathematics and Mechanics, vol. 24, pp. 1225–1246, 1960, (translated).

[3] N. N. Krasovskii and E. A. Lidskii, “Analytical design of controllers in systems with random attributes I, II, III,” Automation and Remote Control, vol. 22, no. 9, pp. 1021–1025, no. 10, pp. 1141–1146, no. 11, pp. 1289–1294, 1961, (translated).

[4] H. Min, S. Xu, and Z. Zhang, “Adaptive finite-time stabilization of stochastic nonlinear systems subject to full-state constraints and input saturation,” IEEE Transactions on Automatic Control, vol. 66, no. 3, pp. 1306–1313, 2020.

[5] Y. Qin, M. Cao, and B. D. Anderson, “Lyapunov criterion for stochastic systems and its applications in distributed computation,” IEEE Transactions on Automatic Control, vol. 65, no. 2, pp. 546–560, 2020.
