A Contraction Theory Approach to Stochastic Incremental Stability

Quang-Cuong Pham, Nicolas Tabareau and Jean-Jacques Slotine

Abstract

We investigate the incremental stability properties of Itô stochastic dynamical systems. Specifically, we derive a stochastic version of nonlinear contraction theory that provides a bound on the mean square distance between any two trajectories of a stochastically contracting system. This bound can be expressed as a function of the noise intensity and the contraction rate of the noise-free system. We illustrate these results in the contexts of nonlinear observers design and stochastic synchronization.

Index Terms

Stochastic stability, incremental stability, nonlinear contraction theory

I. Introduction

Nonlinear stability properties are often considered with respect to an equilibrium point or to a nominal system trajectory (see e.g. [1]). By contrast, *incremental* stability is concerned with the behavior of system trajectories *with respect to each other*. From the triangle inequality, global exponential incremental stability (any two trajectories tend to each other exponentially) is a stronger property than global exponential convergence to a single trajectory.

Historically, work on deterministic incremental stability can be traced back to the 1950's [2; 3; 4] (see e.g. [5; 6] for a more extensive list and historical discussion of related references). More recently, and largely independently of these earlier studies, a number of works have put incremental stability on a broader theoretical basis and have clarified the relations with more traditional stability approaches [7; 8; 9; 10]. Furthermore, it has been shown that incremental stability is especially relevant in the study of such problems as observer design or synchronization analysis.

While the above references are mostly concerned with *deterministic* stability notions, stability theory has also been extended to *stochastic* dynamical systems, see for instance [11; 12]. This includes important recent developments in Lyapunov-like approaches [13; 14], as well as applications to standard problems in systems and control [15; 16; 17]. However, stochastic versions of incremental stability have not yet been systematically investigated.

The goal of this paper is to extend some concepts and results in incremental stability to stochastic dynamical systems. More specifically, we derive a stochastic version of contraction analysis in the specialized context of state-independent metrics.

We prove in section II that the mean square distance between any two trajectories of a stochastically contracting system is upper-bounded by a constant after exponential transients. In contrast with previous works on incremental stochastic stability [18], we consider the case when the two trajectories are affected by *distinct* and independent noises, as detailed in section II-B. This specificity enables our theory to have a number of new and practically important applications. However, the fact that the noise does not vanish as two trajectories get very close to each other will prevent us from obtaining asymptotic almost-sure stability results (see section III-B). In section III-D, we show that results on combinations of

This work was supported in parts by the European Community projects BACS (FP6-IST-027140) and Neurobotics (FP6-IST-001917).
Quang-Cuong Pham (corresponding author) and Nicolas Tabareau are with Laboratoire de Physiologie de la Perception et de l'Action (LPPA), UMR 7152 Collège de France - CNRS, Paris, France (e-mail: cuong.pham@normalesup.org;tabareau.nicolas@gmail.com).
Jean-Jacques Slotine is with the Nonlinear Systems Laboratory, MIT, Cambridge, MA 02139, USA (e-mail: jjs@mit.edu).

deterministic contracting systems have simple analogues in the stochastic case. Finally, as illustrations of our results, we study in section IV the convergence of contracting observers with noisy measurements, and the synchronization of noisy FitzHugh-Nagumo oscillators.

II. The stochastic contraction theorem

A. Background: nonlinear contraction theory

Nonlinear contraction theory [8] provides a set of tools to analyze the incremental exponential stability of nonlinear systems, and has been applied notably to observer design [19; 20], synchronization analysis [21; 22] and systems neuroscience modelling [23]. Nonlinear contracting systems enjoy desirable aggregation properties, in that contraction is preserved under many types of system combinations given suitable simple conditions [8].

While we shall derive global properties of nonlinear systems, many of our results can be expressed in terms of eigenvalues of symmetric matrices [24]. Given a square matrix $\mathbf{A}$, the symmetric part of $\mathbf{A}$ is denoted by $\mathbf{A}_s$. The smallest and largest eigenvalues of $\mathbf{A}_s$ are denoted by $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$. Given these notations, a matrix $\mathbf{A}$ is *positive definite* (denoted $\mathbf{A} > \mathbf{0}$) if $\lambda_{\min}(\mathbf{A}) > 0$. Finally, a time- and state-dependent matrix $\mathbf{A}(\mathbf{x}, t)$ is *uniformly* positive definite if

$$\exists \beta > 0 \quad \forall \mathbf{x}, t \quad \lambda_{\min}(\mathbf{A}(\mathbf{x}, t)) \geq \beta$$

The basic theorem of contraction analysis, derived in [8], can be stated as follows

Theorem 1 (Deterministic contraction): Consider, in $\mathbb{R}^n$, the deterministic system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{II.1}$$

where $\mathbf{f}$ is a smooth nonlinear function satifying standard conditions for the global existence and uniqueness of solutions (for instance: for all $T \in [0, \infty)$, there are constants M and L such that $\forall t \in [0, T],\ \forall \mathbf{x} \in \mathbb{R}^n :\ \|\mathbf{f}(\mathbf{x}, t)\| \leq M + L\|\mathbf{x}\|$ [4]).

Denote the Jacobian matrix of $\mathbf{f}$ with respect to its first variable by $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$. If there exists a square matrix $\mathbf{\Theta}(\mathbf{x}, t)$ such that $\mathbf{M} = \mathbf{\Theta}^T\mathbf{\Theta}$ is uniformly positive definite and $\mathbf{F} = \left(\dot{\mathbf{\Theta}} + \mathbf{\Theta}\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right)\mathbf{\Theta}^{-1}$ is uniformly negative definite, then all system trajectories converge exponentially to a single trajectory, with convergence rate $\sup_{\mathbf{x},t} |\lambda_{\max}(\mathbf{F})| = \lambda > 0$. The system is said to be *contracting*, $\mathbf{F}$ is called its *generalized Jacobian*, $\mathbf{M}$ its contraction *metric* and λ its contraction *rate*.

B. Settings

Consider a noisy system described by an Itô stochastic differential equation

$$\begin{cases} d\mathbf{a} = \mathbf{f}(\mathbf{a}, t)dt + \sigma(\mathbf{a}, t)dW^d \\ \mathbf{a}(0) = \xi \end{cases} \tag{II.2}$$

where $\mathbf{f}$ is a $\mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^n$ function, σ is a $\mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^{nd}$ matrix-valued function, W^d is a standard d-dimensional Wiener process and ξ is a random variable independent of the noise W^d. To ensure existence and uniqueness of solutions to equation (II.2), we assume that for all $T \in [0, \infty)$

(Lipschitz condition) there exists a constant $K_1 > 0$ such that $\forall t \in [0, T],\ \forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^n$

$$\|\mathbf{f}(\mathbf{a}, t) - \mathbf{f}(\mathbf{b}, t)| + \|\sigma(\mathbf{a}, t) - \sigma(\mathbf{b}, t)\| \leq K_1\|\mathbf{a} - \mathbf{b}\|$$

(restriction on growth) there exists a constant $K_2 > 0$ such that $\forall t \in [0, T],\ \forall \mathbf{a} \in \mathbb{R}^n$

$$\|\mathbf{f}(\mathbf{a}, t)\|^2 + \|\sigma(\mathbf{a}, t)\|^2 \leq K_2(1 + \|\mathbf{a}\|^2)$$

Under these conditions, one can show ([25], p. 105) that equation (II.2) has on $[0, \infty)$ a unique $\mathbb{R}^n$-valued solution $\mathbf{a}(t)$, which is continuous with probability one.

In order to investigate the incremental stability properties of system (II.2), consider now two system trajectories $\mathbf{a}(t)$ and $\mathbf{b}(t)$. Our goal will consist of studying the trajectories $\mathbf{a}(t)$ and $\mathbf{b}(t)$ with respect to each other. For this, we consider the *augmented* system $\mathbf{x}(t) = (\mathbf{a}(t), \mathbf{b}(t))^T$, which follows the equation

$$\left\{ \begin{array}{rcl} d\mathbf{x} & = & \begin{pmatrix} \mathbf{f}(\mathbf{a},t) \\ \mathbf{f}(\mathbf{b},t) \end{pmatrix} dt + \begin{pmatrix} \sigma(\mathbf{a},t) & 0 \\ 0 & \sigma(\mathbf{b},t) \end{pmatrix} \begin{pmatrix} dW_1^d \\ dW_2^d \end{pmatrix} \\ & = & \widehat{\mathbf{f}}(\mathbf{x},t)dt + \widehat{\sigma}(\mathbf{x},t)dW^{2d} \\ \mathbf{x}(0) & = & (\mathbf{a}(0), \mathbf{b}(0)) = (\xi_1, \xi_2) \end{array} \right. \tag{II.3}$$

Important remark As stated in the introduction, the systems $\mathbf{a}$ and $\mathbf{b}$ are driven by *distinct* and independent Wiener processes W_1^d and W_2^d. This makes our approach considerably different from [18], where the authors studied two trajectories driven by *the same* Wiener process.

Our approach enables us to study the stability of the system with respect to differences in initial conditions *and* to random perturbations: indeed, two trajectories of any real-life system are typically affected by *distinct realizations* of the noise. In the deterministic domain, incremental stability with respect to different initial conditions *and* different *deterministic inputs* (incremental Input-to-State Stability or δISS) has been studied in [9; 10; 26]. Besides, it should be noted that our approach leads very naturally to nice results on the comparison of noisy and noise-free trajectories (cf. section III-C), which are particularly useful in applications (cf. section IV).

However, because of the very fact that the two trajectories are driven by distinct Wiener processes, one cannot expect the influence of the noise to vanish when the two trajectories get very close to each other. This constrasts with [18], and more generally, with standard stochastic stability approaches, where the noise is assumed to vanish near the origin. The consequences of this will be discussed in detail in section III-B.

C. Statement and proof of the theorem

We first recall a Gronwall-type lemma

Lemma 1: Let $g : [0, \infty) \to \mathbb{R}$ be a continuous function, C a real number and λ a *strictly positive* real number. Assume that

$$\forall u, t \quad 0 \le u \le t \quad g(t) - g(u) \le \int_u^t -\lambda g(s) + C ds \tag{II.4}$$

Then

$$\forall t \ge 0 \quad g(t) \le \frac{C}{\lambda} + \left[g(0) - \frac{C}{\lambda} \right]^+ e^{-\lambda t} \tag{II.5}$$

where $[\cdot]^+ = \max(0, \cdot)$.

Proof See [27] □

We now introduce two hypotheses

(H1) There exists a state-independent, uniformly positive definite metric $\mathbf{M}(t) = \mathbf{\Theta}(t)^T \mathbf{\Theta}(t)$, with the lower-bound $\beta > 0$ (i.e. $\forall \mathbf{x}, t \ \ \mathbf{x}^T \mathbf{M}(t) \mathbf{x} \ge \beta \|\mathbf{x}\|^2$) and $\mathbf{f}$ is contracting in that metric, with contraction rate λ, i.e. uniformly,

$$\lambda_{\max} \left(\left(\frac{d}{dt} \mathbf{\Theta}(t) + \mathbf{\Theta}(t) \frac{\partial \mathbf{f}}{\partial \mathbf{a}} \right) \mathbf{\Theta}^{-1}(t) \right) \le -\lambda$$

or equivalently, uniformly,

$$\mathbf{M}(t) \frac{\partial \mathbf{f}}{\partial \mathbf{a}} + \left(\frac{\partial \mathbf{f}}{\partial \mathbf{a}} \right)^T \mathbf{M}(t) + \frac{d}{dt} \mathbf{M}(t) \le -2\lambda \mathbf{M}(t)$$

(H2) $\operatorname{tr}\left(\sigma(\mathbf{a},t)^T \mathbf{M}(t) \sigma(\mathbf{a},t) \right)$ is uniformly upper-bounded by a constant C

Definition 1: A system that verifies **(H1)** and **(H2)** is said to be *stochastically contracting* in the metric $\mathbf{M}(t)$, with rate λ and bound C.

Consider the Lyapunov-like function $V(\mathbf{x},t) = (\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{a}-\mathbf{b})$. Using **(H1)** and **(H2)**, we derive below an inequality on $\mathscr{L}V(\mathbf{x},t)$ where $\mathscr{L}$ denotes the differential generator of the Itô process $\mathbf{x}(t)$ ([11], p. 15).

Lemma 2: Under **(H1)** and **(H2)**, one has

$$\forall \mathbf{x}, t \quad \mathscr{L}V(\mathbf{x},t) \leq -2\lambda V(\mathbf{x},t) + 2C \tag{II.6}$$

Proof Let us compute first $\mathscr{L}V$

$$\begin{aligned}
\mathscr{L}V(\mathbf{x},t) &= \frac{\partial V}{\partial t} + \frac{\partial V}{\partial \mathbf{x}}\widehat{\mathbf{f}}(\mathbf{x},t) + \frac{1}{2}\mathrm{tr}\left(\widehat{\sigma}(\mathbf{x},t)^T \frac{\partial^2 V}{\partial \mathbf{x}^2}\widehat{\sigma}(\mathbf{x},t)\right)\\
&= (\mathbf{a}-\mathbf{b})^T\left(\frac{d}{dt}\mathbf{M}(t)\right)(\mathbf{a}-\mathbf{b})\\
&+ 2(\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{f}(\mathbf{a},t)-\mathbf{f}(\mathbf{b},t))\\
&+ \mathrm{tr}(\sigma(\mathbf{a},t)^T\mathbf{M}(t)\sigma(\mathbf{a},t)) + \mathrm{tr}(\sigma(\mathbf{b},t)^T\mathbf{M}(t)\sigma(\mathbf{b},t))
\end{aligned}$$

Fix $t>0$, then, according to [28], there exists $\mathbf{c}\in[\mathbf{a},\mathbf{b}]$ such that

$$\begin{aligned}
&(\mathbf{a}-\mathbf{b})^T\left(\tfrac{d}{dt}\mathbf{M}(t)\right)(\mathbf{a}-\mathbf{b}) + 2(\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{f}(\mathbf{a})-\mathbf{f}(\mathbf{b}))\\
&= (\mathbf{a}-\mathbf{b})^T\left(\tfrac{d}{dt}\mathbf{M}(t) + \mathbf{M}(t)\tfrac{\partial \mathbf{f}}{\partial \mathbf{a}}(\mathbf{c},t) + \tfrac{\partial \mathbf{f}}{\partial \mathbf{a}}(\mathbf{c},t)^T\mathbf{M}(t)\right)(\mathbf{a}-\mathbf{b})\\
&\leq -2\lambda(\mathbf{a}-\mathbf{b})^T\mathbf{M}(t)(\mathbf{a}-\mathbf{b}) = -2\lambda V(\mathbf{x})
\end{aligned} \tag{II.7}$$

where the inequality is obtained by using **(H1)**.

Finally, combining equation (II.7) with **(H2)** allows to obtain the desired result □

We can now state the stochastic contraction theorem

Theorem 2 (Stochastic contraction): Assume that system (II.2) verifies **(H1)** and **(H2)**. Let $\mathbf{a}(t)$ and $\mathbf{b}(t)$ be two trajectories whose initial conditions are independent of the noise and given by a probability distribution $p(\xi_1,\xi_2)$. Then

$$\begin{aligned}
&\forall t\geq 0 \quad \mathbb{E}\left((\mathbf{a}(t)-\mathbf{b}(t))^T\mathbf{M}(t)(\mathbf{a}(t)-\mathbf{b}(t))\right) \leq\\
&\frac{C}{\lambda} + e^{-2\lambda t}\int\left[(\mathbf{a}_0-\mathbf{b}_0)^T\mathbf{M}(0)(\mathbf{a}_0-\mathbf{b}_0) - \frac{C}{\lambda}\right]^+ dp(\mathbf{a}_0,\mathbf{b}_0)
\end{aligned} \tag{II.8}$$

In particular, $\forall t\geq 0$

$$\mathbb{E}\left(\|\mathbf{a}(t)-\mathbf{b}(t)\|^2\right) \leq \frac{1}{\beta}\left(\frac{C}{\lambda} + \mathbb{E}\left((\xi_1-\xi_2)^T\mathbf{M}(0)(\xi_1-\xi_2)\right)e^{-2\lambda t}\right) \tag{II.9}$$

Proof Let $\mathbf{x}_0 = (\mathbf{a}_0,\mathbf{b}_0)\in\mathbb{R}^{2n}$. By Dynkin's formula ([11], p. 10)

$$\mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(t),t) - V(\mathbf{x}_0,0) = \mathbb{E}_{\mathbf{x}_0}\int_0^t \mathscr{L}V(\mathbf{x}(s),s)ds$$

Thus one has $\forall u,t \quad 0\leq u\leq t<\infty$

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(t),t) - \mathbb{E}_{\mathbf{x}_0}V(\mathbf{x}(u),u) &= \mathbb{E}_{\mathbf{x}_0}\int_u^t \mathscr{L}V(\mathbf{x}(s),s)ds\\
&\leq \mathbb{E}_{\mathbf{x}_0}\int_u^t(-2\lambda V(\mathbf{x}(s),s)+2C)ds
\end{aligned} \tag{II.10}$$

$$= \int_u^t (-2\lambda \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(s),s) + 2C) ds \quad \text{(II.11)}$$

where inequality (II.10) is obtained by using lemma 2 and equality (II.11) by using Fubini's theorem (since $s \mapsto \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(s),s)$ is continuous on $[u,t]$, one has $\int_u^t |-2\lambda \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(s),s) + 2C| ds < \infty$).

Denote by $g(t)$ the *deterministic* quantity $\mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(t))$. As remarked above, $g(t)$ is a continuous function of t. It then satisfies the conditions of the Gronwall-type lemma 1, and as a consequence

$$\forall t \geq 0 \quad \mathbb{E}_{\mathbf{x}_0} V(\mathbf{x}(t),t) \leq \frac{C}{\lambda} + \left[V(\mathbf{x}_0,0) - \frac{C}{\lambda}\right]^+ e^{-2\lambda t}$$

which leads to (II.8) by integrating with respect to $(\mathbf{a}_0,\mathbf{b}_0)$. Next, (II.9) follows from (II.8) by observing that

$$\int \left[(\mathbf{a}_0 - \mathbf{b}_0)^T \mathbf{M}(0)(\mathbf{a}_0 - \mathbf{b}_0) - \frac{C}{\lambda}\right]^+ dp(\mathbf{a}_0,\mathbf{b}_0)$$

$$\begin{aligned} &\leq \int (\mathbf{a}_0 - \mathbf{b}_0)^T \mathbf{M}(0)(\mathbf{a}_0 - \mathbf{b}_0) dp(\mathbf{a}_0,\mathbf{b}_0) \\ &= \mathbb{E}\left((\xi_1 - \xi_2)^T \mathbf{M}(0)(\xi_1 - \xi_2)\right) \end{aligned}$$

and

$$\|\mathbf{a}(t) - \mathbf{b}(t)\|^2 \leq \frac{1}{\beta}(\mathbf{a}(t) - \mathbf{b}(t))^T \mathbf{M}(t)(\mathbf{a}(t) - \mathbf{b}(t)) \quad \square$$

III. REMARKS

A. "Optimality" of the mean square bound

Consider the following linear dynamical system, known as the Ornstein-Uhlenbeck (colored noise) process

$$da = -\lambda a dt + \sigma dW \quad \text{(III.1)}$$

Clearly, the noise-free system is contracting with rate λ and the trace of the noise matrix is upper-bounded by σ^2. Let $a(t)$ and $b(t)$ be two system trajectories starting respectively at a_0 and b_0 (deterministic initial conditions). Then by theorem 2, we have

$$\forall t \geq 0 \quad \mathbb{E}\left((a(t) - b(t))^2\right) \leq \frac{\sigma^2}{\lambda} + \left[(a_0 - b_0)^2 - \frac{\sigma^2}{\lambda}\right]^+ e^{-2\lambda t} \quad \text{(III.2)}$$

Let us assess the quality of this bound by solving directly equation (III.1). The solution of equation (III.1) is ([25], p. 134)

$$a(t) = a_0 e^{-\lambda t} + \sigma \int_0^t e^{\lambda(s-t)} dW(s) \quad \text{(III.3)}$$

Compute next the mean square distance between the two trajectories $a(t)$ and $b(t)$

$$\begin{aligned} \mathbb{E}((a(t) - b(t))^2) &= (a_0 - b_0)^2 e^{-2\lambda t} \\ &+ \sigma^2 \mathbb{E}\left(\left(\int_0^t e^{\lambda(s-t)} dW_1(s)\right)^2\right) \\ &+ \sigma^2 \mathbb{E}\left(\left(\int_0^t e^{\lambda(u-t)} dW_2(u)\right)^2\right) \\ &= (a_0 - b_0)^2 e^{-2\lambda t} + \frac{\sigma^2}{\lambda}(1 - e^{-2\lambda t}) \end{aligned}$$

$$\leq \quad \frac{\sigma^2}{\lambda} + \left[(a_0 - b_0)^2 - \frac{\sigma^2}{\lambda}\right]^+ e^{-2\lambda t}$$

The last inequality is in fact an equality when $(a_0 - b_0)^2 \geq \frac{\sigma^2}{\lambda}$. Thus, this calculation shows that the upper-bound (III.2) given by theorem 2 is optimal, in the sense that it can be attained.

B. No asymptotic almost-sure stability

From the explicit form (III.3) of the solutions, one can deduce that the distributions of $a(t)$ and $b(t)$ converge to the normal distribution $\mathscr{N}\left(0, \frac{\sigma^2}{2\lambda}\right)$ ([25], p. 135). Since $a(t)$ and $b(t)$ are independent, the distribution of the difference $a(t) - b(t)$ will then converge to $\mathscr{N}\left(0, \frac{\sigma^2}{\lambda}\right)$. The last observation shows that one cannot – in general – obtain *almost-sure* stability results.

Indeed, the main difference with the approaches in [16; 17; 18] lies in the term $2C$. This extra term comes from the fact that the influence of the noise does not vanish when two trajectories get very close to each other (cf. section II-B). It prevents $\mathscr{L}V(\mathbf{x}(t))$ from being always non-positive, and as a result, $V(\mathbf{x}(t))$ is not always *non-increasing*. Thus, $V(\mathbf{x}(t))$ is not – in general – a supermartingale, and one cannot then use the supermartingale inequality (or its variations) to obtain asymptotic almost-sure bounds, as in ([11], pp. 47-48) or in [16; 17; 18].

However, if one is interested in *finite time* bounds then the supermartingale inequality is still applicable, see ([11], p. 86) for details.

C. Noisy and noise-free trajectories

Consider the following augmented system

$$d\mathbf{x} = \begin{pmatrix} \mathbf{f}(\mathbf{a}, t) \\ \mathbf{f}(\mathbf{b}, t) \end{pmatrix} dt + \begin{pmatrix} 0 & 0 \\ 0 & \sigma(\mathbf{b}, t) \end{pmatrix} \begin{pmatrix} dW_d^1 \\ dW_d^2 \end{pmatrix} =$$

$$\widehat{\mathbf{f}}(\mathbf{x}, t)dt + \widehat{\sigma}(\mathbf{x}, t)dW_{2d} \tag{III.4}$$

This equation is the same as equation (II.3) except that the $\mathbf{a}$-system is not perturbed by noise. Thus $V(\mathbf{x}) = \|\mathbf{a} - \mathbf{b}\|^2$ represents the distance between a noise-free trajectory and a noisy one. All the calculations are the same as in section II-C, with C being replaced by $C/2$. One can then derive the following corollary (for simplicity, we consider the case of identity metric; the general case can be easily adapted)

Corollary 1: Assume that system (II.2) verifies **(H1)** and **(H2)** with $\mathbf{M} = \mathbf{I}$. Let $\mathbf{a}(t)$ be a *noise-free* trajectory starting at $\mathbf{a}_0$ and $\mathbf{b}(t)$ a *noisy* trajectory whose initial condition is independent of the noise and given by a probability distribution $p(\xi_2)$. Then $\forall t \geq 0$

$$\mathbb{E}\left(\|\mathbf{a}(t) - \mathbf{b}(t)\|^2\right) \leq \frac{C}{2\lambda} + \mathbb{E}\left(\|\mathbf{a}_0 - \xi_2\|^2\right) e^{-2\lambda t} \tag{III.5}$$

Remarks

- One can note here that the derivation of corollary 1 is only permitted by our initial choice of considering *distinct* driving Wiener process for the $\mathbf{a}$- and $\mathbf{b}$-systems (cf. section II-B).
- Corollary 1 provides a robustness result for contracting systems, in the sense that any contracting system is *automatically* protected against noise, as quantified by (III.5). This robustness could be related to the exponential nature of contraction stability.

D. Combination properties

Stochastic contraction inherits naturally from deterministic contraction [8] its convenient combination properties. Because contraction is a state-space concept, such properties can be expressed in more general forms than input-output analogues such as passivity-based combinations [29].

It should be noted that, in the deterministic domain, combination properties have been obtained for δISS systems [10; 26] (for the definition of δISS, see section II-B).

Consider two connected systems

$$\left\{ \begin{array}{l} d\mathbf{x}_1 = \mathbf{f}_1(\mathbf{x}_1,\mathbf{x}_2,t)dt + \sigma_1(\mathbf{x}_1,t)dW_1 \\ d\mathbf{x}_2 = \mathbf{f}_2(\mathbf{x}_1,\mathbf{x}_2,t)dt + \sigma_2(\mathbf{x}_2,t)dW_2 \end{array} \right.$$

where system i ($i=1,2$) is stochastically contracting with respect to $\mathbf{M}_i = \mathbf{\Theta}_i^T\mathbf{\Theta}_i$, with rate λ_i and bound C_i (here, $\mathbf{M}_i$ and $\mathbf{\Theta}_i$ are set to be constant matrices for simplicity; the case of time-varying metrics can be easily adapted).

Assume that these systems are connected by *negative feedback* [30], i.e. the Jacobian of their coupling matrices verify $\mathbf{\Theta}_1\mathbf{J}_{12}\mathbf{\Theta}_2^{-1} = -k\mathbf{\Theta}_2\mathbf{J}_{21}^T\mathbf{\Theta}_1^{-1}$, with k a positive constant. The Jacobian matrix of the augmented noise-free system is given then by

$$\mathbf{J} = \left(\begin{array}{cc} \mathbf{J}_1 & -k\mathbf{\Theta}_1^{-1}\mathbf{\Theta}_2\mathbf{J}_{21}^T\mathbf{\Theta}_1^{-1}\mathbf{\Theta}_2 \\ \mathbf{J}_{21} & \mathbf{J}_2 \end{array} \right)$$

Consider the coordinate transform $\mathbf{\Theta} = \left(\begin{array}{cc} \mathbf{\Theta}_1 & \mathbf{0} \\ \mathbf{0} & \sqrt{k}\mathbf{\Theta}_2 \end{array} \right)$ associated with the metric $\mathbf{M} = \mathbf{\Theta}^T\mathbf{\Theta} > \mathbf{0}$. After some calculations, one has

$$\begin{array}{rcl} \left(\mathbf{\Theta}\mathbf{J}\mathbf{\Theta}^{-1}\right)_s & = & \left(\begin{array}{cc} \left(\mathbf{\Theta}_1\mathbf{J}_1\mathbf{\Theta}_1^{-1}\right)_s & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{\Theta}_2\mathbf{J}_2\mathbf{\Theta}_2^{-1}\right)_s \end{array} \right) \\ & \leq & \max(-\lambda_1,-\lambda_2)\mathbf{I} \quad \text{uniformly} \end{array} \tag{III.6}$$

The augmented system is thus stochastically contracting in the metric $\mathbf{M}$, with rate $\min(\lambda_1,\lambda_2)$ and bound $C_1 + kC_2$.

Similarly, one can show that (with $\mathrm{sing}(\mathbf{A})$ denoting the largest singular value of $\mathbf{A}$)

- **Hierarchical combination:** If $\mathbf{J}_{12} = \mathbf{0}$ and $\mathrm{sing}^2(\mathbf{\Theta}_2\mathbf{J}_{21}\mathbf{\Theta}_1^{-1}) \leq K$, then the augmented system is stochastically contracting in the metric $\mathbf{M}_\epsilon$, with rate $\frac{1}{2}(\lambda_1+\lambda_2-\sqrt{\lambda_1^2+\lambda_2^2})$ and bound $C_1 + \frac{2C_2\lambda_1\lambda_2}{K}$, where $\epsilon = \sqrt{\frac{2\lambda_1\lambda_2}{K}}$.
- **Small gains:** Define $\mathbf{B}_\gamma = \frac{1}{2}\left(\sqrt{\gamma}\mathbf{\Theta}_2\mathbf{J}_{21}\mathbf{\Theta}_1^{-1} + \frac{1}{\sqrt{\gamma}}\left(\mathbf{\Theta}_1\mathbf{J}_{12}\mathbf{\Theta}_2^{-1}\right)^T\right)$. If there exists $\gamma > 0$ such that $\mathrm{sing}^2(\mathbf{B}_\gamma) < \lambda_1\lambda_2$ then the augmented system is stochastically contracting in the metric $\mathbf{M}_\gamma$, with bound $C_1 + \gamma C_2$ and rate λ verifying

$$\lambda \geq \frac{\lambda_1+\lambda_2}{2} - \sqrt{\left(\frac{\lambda_1-\lambda_2}{2}\right)^2 + \mathrm{sing}^2(\mathbf{B}_\gamma)} \tag{III.7}$$

Taken together, the combination properties presented above allow one to build by recursion stochastically contracting systems of arbitrary size.

IV. Some examples

A. Effect of measurement noise on contracting observers

Consider a nonlinear dynamical system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},t) \tag{IV.1}$$

If a measurement $\mathbf{y} = \mathbf{y}(\mathbf{x})$ is available, then it may be possible to choose an output injection matrix $\mathbf{K}(t)$ such that the dynamics

$$\dot{\hat{\mathbf{x}}} = \mathbf{f}(\hat{\mathbf{x}},t) + \mathbf{K}(t)(\hat{\mathbf{y}} - \mathbf{y}) \tag{IV.2}$$

is contracting, with $\hat{\mathbf{y}} = \mathbf{y}(\hat{\mathbf{x}})$. Since the actual state $\mathbf{x}$ is a particular solution of (IV.2), any solution $\hat{\mathbf{x}}$ of (IV.2) will then converge towards $\mathbf{x}$ exponentially.

Assume now that the measurements are corrupted by additive "white noise". In the case of *linear* measurement, the measurement equation becomes $\mathbf{y} = \mathbf{H}(t)\mathbf{x} + \Sigma(t)\eta(t)$ where $\eta(t)$ is a multidimensional "white noise" and $\Sigma(t)$ is the matrix of measurement noise intensities.

The observer equation is now given by the following Itô stochastic differential equation (using the formal rule $dW = \eta dt$)

$$d\hat{\mathbf{x}} = (\mathbf{f}(\hat{\mathbf{x}}, t) + \mathbf{K}(t)(\mathbf{H}(t)\mathbf{x} - \mathbf{H}(t)\hat{\mathbf{x}}))dt + \mathbf{K}(t)\Sigma(t)dW \tag{IV.3}$$

Next, remark that the solution $\mathbf{x}$ of system (IV.1) is a also a solution of the noise-free version of system (IV.3). By corollary 1, one then has, for any solution $\hat{\mathbf{x}}$ of system (IV.3)

$$\forall t \geq 0 \quad \mathbb{E}\left(\|\hat{\mathbf{x}}(t) - \mathbf{x}(t)\|^2\right) \leq \frac{C}{2\lambda} + \|\hat{\mathbf{x}}_0 - \mathbf{x}_0\|^2 e^{-2\lambda t} \tag{IV.4}$$

where

$$\lambda = \inf_{\mathbf{x},t} \left| \lambda_{\max}\left(\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial \mathbf{x}} - \mathbf{K}(t)\mathbf{H}(t)\right)\right|$$

$$C = \sup_{t \geq 0} \operatorname{tr}\left(\Sigma(t)^T \mathbf{K}(t)^T \mathbf{K}(t)\Sigma(t)\right)$$

Remark The choice of the injection gain $\mathbf{K}(t)$ is governed by a trade-off between convergence speed (λ) and noise sensitivity (C/λ) as quantified by (IV.4). More generally, the explicit computation of the bound on the expected quadratic estimation error given by (IV.4) may open the possibility of *measurement selection* in a way similar to the linear case. If several possible measurements or sets of measurements can be performed, one may try at each instant (or at each step, in a discrete version) to select the most relevant, i.e., the measurement or set of measurements which will best contribute to improving the state estimate. Similarly to the Kalman filters used in [31] for linear systems, this can be achieved by computing, along with the state estimate itself, the corresponding bounds on the expected quadratic estimation error, and then selecting accordingly the measurement which will minimize it.

B. Synchronization of noisy FitzHugh-Nagumo oscillators

We analyze in this section the synchronization of two noisy FitzHugh-Nagumo oscillators (see [21] for the references). The interested reader is referred to [32] for a more complete study.

The dynamics of two diffusively-coupled noisy FitzHugh-Nagumo oscillators is given by

$$\begin{cases} dv_i = (c(v_i + w_i - \frac{1}{3}v_i^3 + I_i) + k(v_0 - v_i))dt + \sigma dW_i \\ dw_i = -\frac{1}{c}(v_i - a + bw_i)dt \end{cases}$$

where $i = 1, 2$. Let $\mathbf{x} = (v_1, w_1, v_2, w_2)^T$ and $\mathbf{V} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$. The Jacobian matrix of the projected noise-free system is then given by

$$\begin{pmatrix} c - \frac{c(v_1^2 + v_2^2)}{2} - k & c \\ -1/c & -b/c \end{pmatrix}$$

Thus, if the coupling strength verifies $k > c$ then the projected system will be stochastically contracting in the diagonal metric $\mathbf{M} = \operatorname{diag}(1, c)$ with rate $\min(k - c, b/c)$ and bound σ^2. Hence, the average absolute difference between the two membrane potentials $|v_1 - v_2|$ will be upper-bounded by $\sigma/\sqrt{\min(1, c)\min(k - c, b/c)}$ after exponential transients (see Fig. 1 for a numerical simulation).

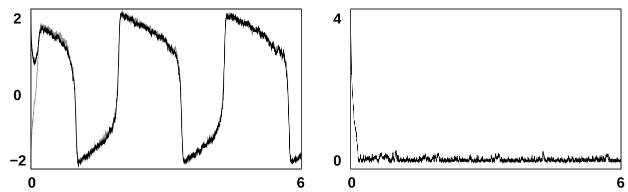

Fig. 1. Synchronization of two noisy FitzHugh-Nagumo oscillators. Left plot: membrane potentials of two coupled noisy FN oscillators. Right plot: absolute difference between the two membrane potentials.

ACKNOWLEDGMENTS

We are grateful to Dr S. Darses, Prof D. Bennequin and Prof M. Yor for stimulating discussions, and to the Editors and the reviewers for their helpful comments.

REFERENCES

[1] J.-J. Slotine and W. Li, *Applied Nonlinear Control*. Prentice-Hall, 1991.
[2] D. Lewis, “Metric properties of differential equations,” *American Journal of Mathematics*, vol. 71, pp. 294–312, 1949.
[3] B. Demidovich, “Dissipativity of a nonlinear system of differential equations,” *Ser. Mat. Mekh.*, 1961.
[4] P. Hartmann, *Ordinary differential equations*. Wiley, 1964.
[5] W. Lohmiller and J.-J. Slotine, “Contraction analysis of nonlinear distributed systems,” *International Journal of Control*, vol. 78, 2005.
[6] J. Jouffroy, “Some ancestors of contraction analysis,” in *Proceedings of the IEEE Conference on Decision and Control*, 2005.
[7] E. Sontag and Y. Wang, “Output-to-state stability and detectability of nonlinear systems,” *Systems and Control Letters*, vol. 29, pp. 279–290, 1997.
[8] W. Lohmiller and J.-J. Slotine, “On contraction analysis for nonlinear systems,” *Automatica*, vol. 34, pp. 671–682, 1998.
[9] V. Fromion, G. Scorletti, and G. Ferreres, “Nonlinear performance of a PI controlled missile: an explanation,” *Int. J. Robust Nonlinear Control*, vol. 9, pp. 485–518, 1999.
[10] D. Angeli, “A Lyapunov approach to incremental stability properties,” *IEEE Transactions on Automatic Control*, vol. 47, pp. 410–422, 2002.
[11] H. Kushner, *Stochastic Stability and Control*. Academic Press, 1967.
[12] R. Has’minskii, *Stochastic Stability of Differential Equations*. Rockville: Sijthoff and Nordhoff, 1980.
[13] P. Florchinger, “Lyapunov-like techniques for stochastic stability,” *SIAM Journal of Control and Optimization*, vol. 33, pp. 1151–1169, 1995.
[14] X. Mao, *Stability of Stochastic Differential Equations with Respect to Semimartingales*. White Plains, NY: Longman, 1991.
[15] P. Florchinger, “Feedback stabilization of affine in the control stochastic differential systems by the control Lyapunov function method,” *SIAM Journal of Control and Optimization*, vol. 35, 1997.
[16] J. Tsinias, “The concept of “exponential ISS” for stochastic systems and applications to feedback stabilization,” *Systems and Control Letters*, vol. 36, pp. 221–229, 1999.
[17] H. Deng, M. Krstic, and R. Williams, “Stabilization of stochastic nonlinear systems driven by noise of unknown covariance,” *IEEE Transactions on Automatic Control*, vol. 46, 2001.
[18] T. Caraballo, P. Kloeden, and B. Schmalfuss, “Exponentially stable stationary solutions for stochastic evolution equations and their pertubation,” *Applied Mathematics and Optimization*, vol. 50, pp. 183–207, 2004.
[19] N. Aghannan and P. Rouchon, “An intrinsic observer for a class of Lagrangian systems,” *IEEE Transactions on Automatic Control*, vol. 48, 2003.

[20] J. Jouffroy and T. Fossen, “On the combination of nonlinear contracting observers and UGES controllers for output feedback,” in *Proceedings of the IEEE Conference on Decision and Control*, 2004.
[21] W. Wang and J.-J. E. Slotine, “On partial contraction analysis for coupled nonlinear oscillators.” *Biol Cybern*, vol. 92, no. 1, pp. 38–53, Jan. 2005.
[22] Q.-C. Pham and J.-J. Slotine, “Stable concurrent synchronization in dynamic system networks.” *Neural Networks*, vol. 20, no. 1, pp. 62–77, Jan. 2007.
[23] B. Girard, N. Tabareau, Q.-C. Pham, A. Berthoz, and J.-J. Slotine, “Where neuroscience and dynamic system theory meet autonomous robotics: a contracting basal ganglia model for action selection,” *Neural Networks*, vol. 21, pp. 628–641, 2008.
[24] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
[25] L. Arnold, *Stochastic Differential Equations : Theory and Applications*. Wiley, 1974.
[26] A. Pavlov, N. van de Wouw, and H. Nijmeijer, *Uniform Output Regulation of Nonlinear Systems*. Birkhauser, 2006.
[27] Q.-C. Pham. (2007) A variation of gronwall’s lemma. [Online]. Available: http://arxiv.org/abs/0704.0922
[28] A. Pavlov, A. Pogromsky, N. van de Wouw, and H. Nijmeijer, “Convergent dynamics, a tribute to Boris Pavlovich Demidovich,” *Systems and Control Letters*, vol. 52, pp. 257–261, July 2004.
[29] V. Popov, *Hyperstability of Control Systems*. Springer-Verlag, 1973.
[30] N. Tabareau and J.-J. Slotine, “Notes on contraction theory,” MIT NSL Report, Tech. Rep., 2005.
[31] E. Dickmanns, *Dynamic Vision for Intelligent Vehicles*. Course Notes, MIT EECS dept, 1998.
[32] N. Tabareau, J.-J. Slotine, and Q.-C. Pham. (2008) How synchronization helps reduce the influence of noise in neuronal networks. [Online]. Available: http://arxiv.org/abs/0801.0011