



# Remote state estimation problem: Towards the data-rate limit along the avenue of the second Lyapunov method<sup>☆</sup>

Christoph Kawan<sup>a</sup>, Alexey S. Matveev<sup>b,d</sup>, Alexander Yu. Pogromsky<sup>c,d,\*</sup>

<sup>a</sup> Institute of Informatics, LMU Munich, Germany

<sup>b</sup> Department of Mathematics and Mechanics, Saint Petersburg University, St. Petersburg, Russia

<sup>c</sup> Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

<sup>d</sup> Department of Control Systems and Industrial Robotics, Saint-Petersburg National Research University of Information Technologies Mechanics and Optics (ITMO), Russia

## ARTICLE INFO

### Article history:

Received 5 May 2020

Received in revised form 19 September 2020

Accepted 14 December 2020

Available online 12 January 2021

### Keywords:

Entropy

Nonlinear systems

First and second Lyapunov methods

Remote state estimation

Finite bit-rates

## ABSTRACT

In the context of control and estimation under information constraints, restoration entropy measures the minimal required data rate above which the state of a system can be estimated so that the estimation quality does not degrade over time and, conversely, can be improved. The remote observer here is assumed to receive its data through a communication channel of finite bit-rate capacity. In this paper, we provide a new characterization of the restoration entropy which does not require to compute any temporal limit, i.e., an asymptotic quantity. Our new formula is based on the idea of finding a specific Riemannian metric on the state space which makes the metric-dependent upper estimate of the restoration entropy as tight as one wishes.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Recent decades have witnessed a substantially growing attention to networked control systems (Heemels, Teel, Van De Wouw, & Nešić, 2010; Hespanha, Naghshtabrizi, & Yung, 2007) and related problems of control and/or state estimation via communication channels with constrained bit-rates; for extended surveys of this area, we refer the reader to Andrievsky, Matveev, and Fradkov (2010), Matveev and Savkin (2009), Yüksel and Başar (2013) and references therein. One of the fundamental concerns in this context is to find a minimal data rate between the communication peers above which remote state estimation (Wong & Brockett, 1997, see also Fradkov, Andrievsky, & Ananyevskiy, 2015 for a related problem) is feasible. In other words, the receiver can reconstruct the current state of the remote system in the real-time regime. The receiver is updated by means of a bit flow with a limited bit-rate. A minimal threshold of this rate, which still ensures that the remote observer is able to keep track

of the system, is the quantity one likes to evaluate in a constructive manner. Loosely speaking, the communication rate between the system and the observer has to exceed the rate at which the system “generates information”, while the latter concept is classically formalized in a form of entropy-like characteristic of the dynamical system at hands. The related mathematical results are usually referred to as Data Rate Theorems (see, e.g. Matveev & Pogromsky, 2019; Matveev & Savkin, 2009; Nair, Fagnani, Zampieri, & Evans, 2007 and references therein) - their various versions coexist to handle various kinds of observability and models of both the plant and the constrained communication channel.

Those results deliver a consistent message that the concept of the topological entropy (TE) of the system and its recent offshoots provide the figure-of-merit needed to evaluate the channel capacity for control applications; the mentioned modifications of TE are partly aimed to properly respond to miscellaneous phenomena crucial for control problems, like uncertainties in the observed system (Kawan & Yüksel, 2017, 2018; Savkin, 2006), implications of control actions (Colonius & Kawan, 2009; Colonius, Kawan, & Nair, 2013; Hagihara & Nair, 2013; Rungger & Zamani, 2017), the decay rate of the estimation error (Liberzon & Mitra, 2018), or Lipschitz-like relations between the exactness of estimation and the initial state uncertainty (Matveev & Pogromsky, 2019). Keeping in mind the relevance of communication constraints in modern control engineering, constructive methods to compute or finely estimate those entropy-like characteristics take on crucial

<sup>☆</sup> The material in this paper was partially presented at the 21st IFAC World Congress (IFAC 2020), July 12–17, 2020, Berlin, Germany. This paper was recommended for publication in revised form by Associate Editor Romain Postoyan under the direction of Editor Daniel Liberzon.

\* Corresponding author at: Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.

E-mail addresses: [christoph.kawan@lmu.de](mailto:christoph.kawan@lmu.de) (C. Kawan), [almat1712@yahoo.com](mailto:almat1712@yahoo.com) (A. Matveev), [A.Pogromsky@tue.nl](mailto:A.Pogromsky@tue.nl) (A. Pogromsky).

not only theoretical but also practical importance. Several steps have been done in this direction in [Hafstein and Kawan \(2019\)](#), [Matveev and Pogromsky \(2019\)](#) and [Pogromsky and Matveev \(2011\)](#), where corresponding upper estimates were found by following up the ideas of the second Lyapunov method. Moreover, it was shown that for some particular prototypical chaotic systems of low dimensions, these upper estimates are exact in the sense that they coincide with the true value of the estimated quantity.

Whether these inspiring samples of precise calculations are mere incidents, or, conversely, are particular manifestations of a comprehensive capacity inherent in the employed approach? Confidence in the last option would constitute a rationale for undertaking special efforts aimed to fully unleash the potential of this approach via its further elaboration.

The primary goal of the current paper is to answer the posed question; we show that among the above options, the last one is the true one. This is accomplished via a sort of a converse result, which is similar in spirit to the celebrated converse Lyapunov theorems. Among various descendants of TE, we pick the so-called *restoration entropy* (RE) ([Matveev & Pogromsky, 2019](#)) to deal with. In the previous work ([Matveev & Pogromsky, 2019](#)), it has been shown that an upper estimate of RE can be derived in terms of singular values of the derivative of the system flow, calculated with respect to some metric. Any metric involved in such calculations will result in a valid upper estimate. The main contribution of this paper is a result showing that one can find a metric that lets the corresponding upper estimate get arbitrarily close to the true value of RE.

The tractability of the developed approach is confirmed by a closed-form computation of the restoration entropy for the celebrated Lanford system (see, e.g., [Belozyorov, 2015](#)). Meanwhile, computation or even fine estimation of TE and the likes has earned the reputation of an extremely complicated matter ([Koiraan, 2001](#)).

The paper is organized as follows. Section 2 offers the remote state estimation problem statement and presents the main assumptions. Section 3 contains the main results, which are illustrated by an example in Section 4. Section 5 summarizes the main results of the paper and outlines some open questions. The technical proofs of the main results are collected in the [Appendices](#).

The following notations are adopted in this paper:  $\log$  – logarithm base 2,  $Df$  – Jacobian matrix of function  $f$ ,  $:=$  – “is defined to be”,  $B_\delta(x)$  – open ball of radius  $\delta$  centered at  $x$ ,  $C^0(X, Y)$  – set of all continuous functions from  $X$  to  $Y$ ,  $P^t$  with  $t \in \mathbb{R}$  –  $t$ th power of a symmetric positive definite matrix  $P$ . Given a matrix  $A$ , its singular values repeated in accordance to their multiplicities and ordered in the nonincreasing order, are denoted by  $\alpha_i(A)$ . If the matrix  $A$  is parameterized:  $A = A(x)$  or  $A = A(t, x)$ , then for the sake of brevity, the corresponding singular values will be denoted as  $\alpha_i(x)$ , or  $\alpha_i(t, x)$ , provided the choice of the matrix  $A$  is clear from the context.

## 2. State estimation via limited bit-rate communication and restoration entropy

The objective of this section is to outline basic points of recent results that motivate the research reported in this paper.

We consider time-invariant dynamical systems of the form

$$x^\nabla(t) = \varphi[x(t)], \quad t \in \mathfrak{T}_+, \quad x(0) \in K \subset \mathbb{R}^n \quad (1)$$

in the following two cases:

- (c-t) Either  $\mathfrak{T}_+ = [0, \infty) \subset \mathbb{R}$ , and the symbol  $x^\nabla$  stands for the derivative of the function:  $x^\nabla(t) := \dot{x}(t)$ ;
- (d-t) Or  $\mathfrak{T}_+$  is the set of integers  $t \geq 0$ , and  $x^\nabla$  denotes the forward time-shift by one step:  $x^\nabla(t) := x(t+1)$ .

In (1),  $\varphi(\cdot)$  is of class  $C^1$  and  $K$  is a given compact set of initial states that are of our interest; in the case (d-t), all considered time variables assume integer values by default.

We deal with a situation where at a remote site  $S_{\text{est}}$ , direct observation of the time-varying state  $x(t)$  is impossible but a reliable estimate  $\hat{x}(t)$  of  $x(t)$  is needed at any time  $t \in \mathfrak{T}_+$ . Meanwhile, data may be communicated to  $S_{\text{est}}$  from another site, where  $x(t)$  is fully measured at time  $t$ . The main issue of our interest arises from the fact that the channel of communication between these sites allows to transmit only finitely many bits per unit time. *How large their number must be so that a good estimate can be generated?*

We clarify this issue only in general terms by following [Matveev and Pogromsky \(2016, 2019\)](#) and [Matveev and Savkin \(2009, Sec. 3.4\)](#), and refer the reader there for full details. The channel is a communication facility that transmits finite-bit data packets; transmission consumes time (during which the channel cannot process new packets). The time  $t_*$  when the transfer of a packet is started, and its bit-size and contents are somewhat manipulable; the channel is used repeatedly, thus transmitting a flow of messages. Despite all variability in the ways of channel usage, the channel imposes an upper bound  $b_+(r)$  on the total number of bits that can be transferred within any interval of duration  $r$ . On the positive side, there is a way to deliver no less than  $b_-(r)$  bits. We consider the case where the two associated averaged per-unit-time amounts of bits converge to a common value  $c = \lim_{r \rightarrow \infty} b_\pm(r)/r$  as the duration  $r$  grows without limits; this value  $c$  is called the *capacity* of the channel.

In the situation at hands, the *observer* is composed of a *coder* and a *decoder*. The coder is located at the measurement site; its function is to generate the departure times  $t_*$  and the messages to  $S_{\text{est}}$  based on the preceding measurements, as well as on the initial estimate  $\hat{x}(0)$  and its accuracy  $\delta > 0$ .

$$\|x(0) - \hat{x}(0)\| < \delta, \quad x(0), \hat{x}(0) \in K. \quad (2)$$

The *decoder* is built at the site  $S_{\text{est}}$  and has access to  $\hat{x}(0)$  and  $\delta > 0$ ; its duty is to generate an estimate  $\hat{x}(t)$  of the current state  $x(t)$  at the current time  $t$  based on the messages fully received through the channel prior to this time. Both the coder and decoder are aware of  $\varphi(\cdot)$  and  $K$  from (1).

Given the plant (1), the possibility of building a reliable observer is dependent on the capacity  $c$  of the employed channel ([Matveev & Pogromsky, 2016, 2019](#)). Those values of  $c$  that make reliable observation possible are in our focus. Their minimum is defined by the system itself, is called the *observability rate*  $\mathfrak{R}(\varphi, K)$ , is measured in bits per unit time, and depends on the concept of observation reliability, as is specified in the following.

**Definition 1** ([Matveev & Pogromsky, 2016, 2019](#)). An observer is said

- (i) *to observe the system (1)* if for any  $\varepsilon > 0$  there is  $\delta(\varepsilon) > 0$  such that whenever (2) holds with  $\delta = \delta(\varepsilon)$ , the estimation error never exceeds  $\varepsilon$ :

$$\|x(t) - \hat{x}(t)\| \leq \varepsilon \quad \forall t \in \mathfrak{T}_+;$$

- (ii) *to regularly observe the system (1)* if the anytime error is uniformly proportional to the initial error: there exist  $\delta_* > 0$  and  $G > 0$  such that whenever  $\delta \leq \delta_*$  in (2),

$$\|x(t) - \hat{x}(t)\| \leq G\delta \quad \forall t \in \mathfrak{T}_+;$$

- (iii) *to finely observe the system (1)* if in addition to (ii) the observation error exponentially decays to zero as time progresses: there exist  $\delta_* > 0$ ,  $G > 0$ , and  $\eta > 0$  such that whenever  $\delta \leq \delta_*$  in (2), the following holds:

$$\|x(t) - \hat{x}(t)\| \leq G\delta e^{-\eta t} \quad \forall t \in \mathfrak{T}_+.$$

The system (1) is said to be (i) *observable*, (ii) *regularly observable*, or (iii) *finely observable* via a given communication channel if there is an observer that firstly, operates through this channel and secondly, observes, regularly observes, or finely observes the system, respectively.

The infimum value of the rates  $c$  such that (1) is observable in the selected (from (i)–(iii)) sense via any channel with capacity  $c$  is denoted by  $\mathcal{R}(f, K)$ , where the index  $o, \text{reg}$ , and  $\text{fine}$  is attached to  $\mathcal{R}$  in the cases (i), (ii), and (iii), respectively.

The rates  $\mathcal{R}_o(\varphi, K)$ ,  $\mathcal{R}_{\text{reg}}(\varphi, K)$ ,  $\mathcal{R}_{\text{fine}}(\varphi, K)$  could be viewed as an answer to the key question posed at the start of the section, but only if a method to effectively compute them be available. Their definitions do not directly suggest such a method since they refer to complementing the system with an a priori uncertain object – an observer. So the first step is to get rid of this uncertainty and to fully express those rates via features of the system only.

Under certain technical assumptions, this feature is the *topological entropy*  $H_{\text{top}}(\varphi, K)$  of the plant (1) in the case of  $\mathcal{R}_o(\varphi, K)$ , which is simply equal to  $H_{\text{top}}(\varphi, K)$ ; we refer the reader to [Dow-narowicz \(2011\)](#) for the definition of  $H_{\text{top}}(\varphi, K)$ . In the case of both  $\mathcal{R}_{\text{reg}}(\varphi, K)$  and  $\mathcal{R}_{\text{fine}}(\varphi, K)$ , this is another characteristic of the system. It is introduced in [Matveev and Pogromsky \(2019\)](#) for the continuous-time case, extended to the discrete-time one in [Kawan \(2019\)](#), and is called the *restoration entropy*  $H_{\text{res}}(\varphi, K)$ .

To recall its definition, we emphasize once more the following.

**Assumption 2.** The map  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is of class  $C^1$  and the set  $K \subset \mathbb{R}^n$  is compact and forward-invariant.

Further,  $x(t, a)$  denotes the solution of (1) that starts with  $x(0) = a$ , and  $\varphi^t$  stands for the time- $t$  map of the system,  $a \mapsto x(t, a)$ .

Let a duration  $\tau \in \mathbb{T}_+$ , a state  $a \in \mathbb{R}^n$ , and an “error tolerance level”  $\delta > 0$  be given. We denote by  $N_{\varphi, K}(\tau, a, \delta)$  the minimal number of open  $\delta$ -balls needed to cover the image  $\varphi^\tau[B_\delta(a) \cap K]$ . Then

$$H_{\text{res}}(\varphi, K) := \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \limsup_{\delta \rightarrow 0} \sup_{a \in K} \log N_{\varphi, K}(\tau, a, \delta). \quad (3)$$

Here  $\lim_{\tau \rightarrow \infty}$  exists by Fekete’s lemma (see [Colonius et al., 2013](#), Lem. 2.1 for a proof) and is due to the subadditivity of the concerned quantity in  $\tau$ . The interest in (3) is caused by the fact that ([Matveev & Pogromsky, 2019](#))

$$H_{\text{res}}(\varphi, K) = \mathcal{R}_{\text{reg}}(\varphi, K) = \mathcal{R}_{\text{fine}}(\varphi, K).$$

Also,  $H_{\text{res}}(\varphi, K) \geq H_{\text{top}}(\varphi, K)$  ([Kawan, 2019](#); [Matveev & Pogromsky, 2019](#)). Meanwhile, the concerned two concepts are not identical:  $H_{\text{top}} < H_{\text{res}}$  for the logistic map ([Pogromsky & Matveev, 2016a](#), Ex. 5.1), whereas ([Kawan, 2019](#)) offers a characterization of the hyperbolic systems with  $H_{\text{top}} = H_{\text{res}}$  and suggests that  $H_{\text{top}} = H_{\text{res}}$  is a relatively rare occurrence.

The restoration entropy can be also linked to the *finite-time Lyapunov exponents*  $t^{-1} \ln \alpha_i(t, x)$ , where  $\alpha_1(t, x) \geq \dots \geq \alpha_n(t, x)$  are the singular values of the Jacobian matrix  $\mathbf{D}\varphi^t(x)$ . It is convenient for us to divide these exponents by  $\ln 2$ , which results in the change of the logarithm base:

$$\Lambda_i(t, x) := t^{-1} \log \alpha_i(t, x).$$

As is shown in [Matveev and Pogromsky \(2019, Thm. 11\)](#),

$$\begin{aligned} H_{\text{res}}(\varphi, K) &= \max_{x \in K} \lim_{t \rightarrow \infty} \sum_{i=1}^n \max\{0, \Lambda_i(t, x)\} \\ &= \lim_{t \rightarrow \infty} \max_{x \in K} \sum_{i=1}^n \max\{0, \Lambda_i(t, x)\} \end{aligned} \quad (4)$$

provided that the following holds.

**Assumption 3.** The set  $K$  is the closure of its interior.

**Remark 4** ([Matveev & Pogromsky, 2019](#)). If [Assumption 3](#) is dropped, the first relation from (4) remains true provided that  $\leq$  is put in place of  $=$  in it.

The topological entropy and Lyapunov exponents are classic and long-studied concepts. However, their use in assessing  $H_{\text{res}}$  is highly impeded by the fact that constructive practical evaluation of both  $H_{\text{top}}$  and the limit in (4) has earned the reputation of an intricate matter and is in fact a long-standing challenge that still remains unresolved in many respects.

We offer an alternative machinery for evaluation of  $H_{\text{res}}$ . By the foregoing, it can also be used to upper estimate  $H_{\text{top}}$  and to quantify finite-time Lyapunov exponents for large times.

### 3. Main ideas and results

Due to practical needs, the definition (3) operates with balls in the Euclidean metric. Meanwhile, balls and Lyapunov exponents are standard in studies of dynamics on Riemannian manifolds, where its own metric tensor is assigned to every point. A homogeneous, state-independent metric was used in Section 2. The idea to modify the material of Section 2 via assigning its own metric tensor to any point  $x \in \mathbb{R}^n$ , thus altering the neighboring balls and the finite-time Lyapunov exponents, may look as nothing but a complication with no reason. However, this hint is a keystone for a method that not only aids in practical estimation of  $H_{\text{res}}$  but also carries a potential to compute  $H_{\text{res}}$  with as high exactness as desired.

To specify this, we denote by  $\mathcal{S}$  the linear space of the real symmetric  $n \times n$  matrices, and by  $\mathcal{S}^+ \subset \mathcal{S}$  the subset of the positive definite ones. A continuous mapping  $P : K \rightarrow \mathcal{S}^+$  gives rise to a Riemannian metric on  $K$  ([Chavel, 2006](#)) by defining the state-dependent inner product<sup>1</sup>

$$\langle v, w \rangle_{P, x} := \langle P(x)v, w \rangle,$$

where  $\langle \cdot, \cdot \rangle$  is the standard Euclidean inner product.

This Riemannian metric produces its own finite-time Lyapunov exponents and, on a more general level, singular values of  $A(x) := \mathbf{D}\phi(x)$ ,  $x \in K$  for any map  $\phi \in C^1(\mathbb{R}^n \rightarrow \mathbb{R}^n)$  such that  $\phi(K) \subset K$ . Indeed, then  $A(x)$  should be viewed as an operator between two spaces, endowed with the inner products  $\langle \cdot, \cdot \rangle_{P, x}$  and  $\langle \cdot, \cdot \rangle_{P, \phi(x)}$ , respectively. This obliges to treat the singular values of  $A(x)$  as the square roots of the eigenvalues of the operator  $\mathbf{D}\phi(x)^* \mathbf{D}\phi(x)$ , where  $\mathbf{D}\phi(x)^*$  is the adjoint to  $\mathbf{D}\phi(x)$  with respect to that pair of inner products.

**Lemma 5.** The singular values  $\alpha_1^P(x|\phi) \geq \dots \geq \alpha_n^P(x|\phi) \geq 0$  of the matrix  $A(x) := \mathbf{D}\phi(x)$  in the metric  $\langle \cdot, \cdot \rangle_P$  are the square roots of the solutions  $\lambda$  for the following algebraic equation:

$$\det[A(x)^T P(\phi(x))A(x) - \lambda P(x)] = 0. \quad (5)$$

These solutions are also the eigenvalues of the positive semi-definite matrix  $B(x)^T B(x)$ , where

$$B(x) := [P(\phi(x))]^{1/2} A(x) P(x)^{-1/2}. \quad (6)$$

**Proof.** To compute the adjoint to  $A(x)$ , we observe that

$$\begin{aligned} \langle A(x)v, w \rangle_{P, \phi(x)} &= \langle P(\phi(x))A(x)v, w \rangle \\ &= \langle A(x)v, P(\phi(x))w \rangle = \langle v, A(x)^T P(\phi(x))w \rangle \\ &= \langle v, P(x)P(x)^{-1}A(x)^T P(\phi(x))w \rangle \end{aligned}$$

<sup>1</sup> Here  $P$  is not a Riemannian metric in the strict sense, insofar as  $K$  is not necessarily a manifold.

$$= \langle P(x)v, P(x)^{-1}A(x)^T P(\phi(x))w \rangle \\ = \langle v, P(x)^{-1}A(x)^T P(\phi(x))w \rangle_{P,x}$$

implying  $A(x)^* = P(x)^{-1}A(x)^T P(\phi(x))$ . It remains to note that the associated singular value equation

$$\det [P(x)^{-1}A(x)^T P(\phi(x))A(x) - \lambda I_n] = 0 \quad (7)$$

is equivalent to both (5) and  $\det[B(x)^T B(x) - \lambda I] = 0$ , which can be seen via multiplying the matrix inside [...] in (7) by  $P(x)[\dots]$  and  $P(x)^{1/2}[\dots]P(x)^{-1/2}$  in the first and second case, respectively. ■

### 3.1. Discrete-time systems

Now we are in a position to state our first main result.

**Theorem 6.** *Let the case (d-t) and Assumption 2 hold. Then the following statements are true (where  $\log 0 := -\infty$ ):*

- (i) Any map  $P(\cdot) \in C^0(K, S^+)$  gives rise to the following bound on the restoration entropy of the system (1):

$$H_{\text{res}}(\varphi, K) \leq \max_{x \in K} \Sigma^P(x|\varphi),$$

$$\text{where } \Sigma^P(x|\varphi) := \sum_{i=1}^n \max\{0, \log \alpha_i^P(x|\varphi)\}. \quad (8)$$

- (ii) Let the set  $K$  satisfy Assumption 3 and the Jacobian matrix  $\mathbf{D}\varphi(x)$  be invertible for every  $x \in K$ . Then for any  $\varepsilon > 0$ , there exists  $P \in C^0(K, S^+)$  such that

$$H_{\text{res}}(\varphi, K) \geq \max_{x \in K} \Sigma^P(x|\varphi) - \varepsilon.$$

The proof of this theorem is given in Appendix B.

In Theorem 6,  $\max_{x \in K}$  is attained since  $K$  is compact by Assumption 2 and  $\alpha_i^P(x|\varphi)$  depend continuously on  $x$ . The latter holds since  $\alpha_i^P(x|\varphi)$  are the singular values of the matrix (6), which is continuous in  $x$ , whereas the singular values continuously depend on the matrix (Horn & Johnson, 2013, Cor. 7.4.3.a).

The claim (i) converts any positive definite matrix function  $P$  into an upper estimate on  $H_{\text{res}}(\varphi, K)$ , while taking no limits as  $t \rightarrow \infty$ , unlike (3) or (4). Meanwhile, (ii) shows that this estimate can be made as tight as one wishes by a proper choice of  $P$ . So the presented method allows for computation of  $H_{\text{res}}(\varphi, K)$  with as high exactness as desired.

The claim (i) of Theorem 6 extends (Matveev & Pogromsky, 2016, Thm. 12) from the case of only constant  $P$ 's to the class of metrics of the most general type. It is shown in Matveev and Pogromsky (2016) that the discussed method is largely constructive: intelligent choices of  $P$  and simple lower bounds on  $H_{\text{res}}$  result in closed-form expressions of  $H_{\text{res}}$  in terms of the parameters of some classic prototypical chaotic systems. The claim (ii) proves that these are not accidental successes but are manifestations of a fundamental trait of the approach.

**Remark 7.** Under the assumptions of (ii) in Theorem 6, the claim (ii) yields the following new and exact formula:

$$H_{\text{res}}(\varphi, K) = \inf_{P \in C^0(K, S^+)} \max_{x \in K} \Sigma^P(x|\varphi). \quad (9)$$

In fact,  $\inf_P$  can be taken only over  $P$ 's that are extendible to an open neighborhood  $V_P$  of  $K$  as  $C^\infty$ -smooth mappings to  $S^+$ . (Since this is not used in the paper, the proof is omitted.)

**Corollary 8.** *Let  $K$  meet Assumption 3 and  $\mathfrak{G} := \{\varphi \in C^1(\mathbb{R}^n \rightarrow \mathbb{R}^n) : \varphi(K) \subset K, \exists \mathbf{D}\varphi(x)^{-1} \forall x \in K\}$  be endowed with the topology of uniform convergence of the function and its first derivatives on*

$K$ . The map  $\varphi \in \mathfrak{G} \mapsto H_{\text{res}}(\varphi, K)$  is upper semi-continuous:  $H_{\text{res}}(\varphi, K) \geq \lim_{\phi \rightarrow \varphi, \phi \in \mathfrak{G}} H_{\text{res}}(\phi, K) \forall \varphi \in \mathfrak{G}$ .

Indeed,  $\alpha_i^P(x|\varphi)$  continuously depend not only on  $x$  but also on the map  $\varphi \in \mathfrak{G}$ , as can be easily seen by the analysis of the arguments in the second paragraph following Theorem 6. Hence,  $\Sigma^P(x|\varphi)$  and  $\max_{x \in K} \Sigma^P(x|\varphi)$  also depend on  $\varphi \in \mathfrak{G}$  continuously. It remains to note that the infimum (over  $P$ 's, in our case) of continuous functions is upper semi-continuous; see, e.g., Schechter (1997, 15.23).

In a practical setting, Corollary 8 asserts that any upper estimate of  $H_{\text{res}}$  is robust: if an upper estimate is established for a nominal model (let the above  $\varphi$  be associated with this model), this estimate remains true, possibly modulo small correction, under small uncertainties in the model and perturbations of its parameters. Here the “smallness” of the correction goes to zero as so do the uncertainties and perturbations.

### 3.2. Continuous-time systems

For continuous-time systems, an analog to (5) looks as follows:

$$\det\{2[P(x)\mathbf{D}\varphi(x)]^{\text{sym}} + \dot{P}(x) - \lambda P(x)\} = 0, \quad (10)$$

where  $B^{\text{sym}} := (B + B^T)/2$  is the symmetric part of the matrix  $B \in \mathbb{R}^{n \times n}$  and  $\dot{P}$  is the orbital derivative:

$$\dot{P}(a) := \lim_{\tau \rightarrow 0+} \frac{P[x(\tau, a)] - P[a]}{\tau}. \quad (11)$$

This derivative is equal to  $\mathbf{D}P(x)\varphi(x)$  for continuously differentiable  $P$ 's. However, we shall consider maps  $P(\cdot)$  with a limited differentiability, as is described in the following.

**Assumption 9.** The map  $P(\cdot) \in C^0(K, S^+)$  is such that for any  $a \in \mathbb{R}^n$ , the limit in (11) exists and is orbitally continuous, i.e., the function  $t \mapsto \dot{P}[x(t, a)]$  is continuous.

**Remark 10.** The  $x$ -dependent roots of the algebraic equation (10) are the eigenvalues of the symmetric matrix

$$P(x)^{-1/2} \{2[P(x)\mathbf{D}\varphi(x)]^{\text{sym}} + \dot{P}(x)\} P(x)^{-1/2}.$$

So these roots are real. Being repeated in accordance with their multiplicity, they are denoted by  $\varsigma_i^P(x)$ ,  $i = 1, \dots, n$ .

**Theorem 11.** *Suppose that the case (c-t) and Assumption 2 hold. Then the following statements are true:*

- (i) For any map  $P(\cdot)$  satisfying Assumption 9,

$$H_{\text{res}}(\varphi, K) \leq \frac{1}{2 \ln 2} \max_{x \in K} \sum_{i=1}^n \max\{0, \varsigma_i^P(x)\}. \quad (12)$$

- (ii) Suppose that the set  $K$  satisfies Assumption 3. Then for any  $\varepsilon > 0$ , there exists a map  $P(\cdot)$  for which Assumption 9 and the following inequality are fulfilled:

$$H_{\text{res}}(\varphi, K) \geq \frac{1}{2 \ln 2} \max_{x \in K} \sum_{i=1}^n \max\{0, \varsigma_i^P(x)\} - \varepsilon.$$

The proof of this theorem is given in Appendix C. As in the case of discrete time, analogs of Remark 7 and Corollary 8 follow from Theorem 11.



#### 4. Example: The Lanford system

Consider the following system:

$$\begin{aligned}\dot{x} &= (a-1)x - y + xz \\ \dot{y} &= x + (a-1)y + yz, \quad x, y, z \in \mathbb{R}, \quad a > 0 \\ \dot{z} &= az - (x^2 + y^2 + z^2)\end{aligned}\quad (13)$$

This system is attributed to Lanford and was studied in many publications, see, e.g. [Belozyorov \(2015\)](#). It is well-known that the system (13) has only two equilibrium points:

$$O_1 = [0, 0, 0]^\top, \quad O_2 = [0, 0, a]^\top.$$

The value  $a = 2/3$  is of particular interest, since then there is a heteroclinic orbit connecting the equilibria ([Belozyorov, 2015](#)).

Let  $K$  be some compact forward-invariant set of (13).

**Remark 12.** In  $K$ , we necessarily have  $z \geq 0$ .

Indeed, if  $z(0) < 0$ , then the third equation of (13) implies that the solution escapes to  $-\infty$  in finite time ( $\dot{z} < -z^2$ ).

In (10), the Jacobian matrix is now given as follows:

$$\mathbf{D}\varphi(x, y, z) = \begin{bmatrix} a-1+z & -1 & x \\ 1 & a-1+z & y \\ -2x & -2y & a-2z \end{bmatrix}.$$

In (i) of [Theorem 11](#), we take the following matrix function<sup>2</sup>

$$P(x, y, z) = P_0 e^{w(x, y, z)} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}}_{=:P_0} \underbrace{\exp\left(\frac{2z}{a}\right)}_{=:w}. \quad (14)$$

Straightforward calculations yield

$$\begin{aligned}P\mathbf{D}\varphi(x) &= \begin{bmatrix} a-1+z & -1 & x \\ 1 & a-1+z & y \\ -x & -y & \frac{1}{2}(a-2z) \end{bmatrix} e^w, \\ \mathbf{D}\varphi(x)^\top P &= \begin{bmatrix} a-1+z & 1 & -x \\ -1 & a-1+z & -y \\ x & y & \frac{1}{2}(a-2z) \end{bmatrix} e^w,\end{aligned}$$

and therefore

$$\begin{aligned}2[P(x)\mathbf{D}\varphi(x)]^{\text{sym}} &= \mathbf{D}\varphi(x)^\top P + P\mathbf{D}\varphi(x) \\ &= e^w \begin{bmatrix} 2(a-1+z) & 0 & 0 \\ 0 & 2(a-1+z) & 0 \\ 0 & 0 & a-2z \end{bmatrix}.\end{aligned}$$

At the same time,

$$\begin{aligned}\dot{P} - \lambda P &= \left(\frac{2\dot{z}}{a} - \lambda\right) e^w P_0 \\ &\stackrel{(13)}{=} e^w \left(\frac{2}{a}(az - x^2 - y^2 - z^2) - \lambda\right) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}.\end{aligned}$$

Finally, the solutions of (10) can easily be found:

$$\begin{aligned}\lambda_1 &= 2(a-2z) + 2\frac{az - z^2 - x^2 - y^2}{a} \\ &\leq -\frac{2z^2}{a} - 2z + 2a \stackrel{(z \geq 0)}{\leq} 2a,\end{aligned}\quad (15)$$

$$\begin{aligned}\lambda_{2,3} &= 2(a-1+z) + 2\frac{az - z^2 - x^2 - y^2}{a} \\ &\leq -\frac{2z^2}{a} + 4z + 2(a-1) \leq 2(2a-1).\end{aligned}\quad (16)$$

<sup>2</sup> For a more detailed treatment of the metric in this form for related problems of stability of forced oscillations, see [Pogromsky and Matveev \(2016b\)](#).

By (i) of [Theorem 11](#), the following upper estimate holds:

$$\begin{aligned}H_{\text{res}}(K) &\leq \frac{1}{2\ln 2} \max_K [\max\{0, \lambda_1\} + 2\max\{0, \lambda_{2,3}\}] \\ &= \frac{1}{2\ln 2} \max\{\max_K \lambda_1, 2\max_K \lambda_{2,3}, \max_K (\lambda_1 + 2\lambda_{2,3})\}.\end{aligned}$$

Maximizing  $\lambda_1 + 2\lambda_{2,3}$  over  $z \in \mathbb{R}$  yields

$$\begin{aligned}\max_{x,y,z} (\lambda_1 + 2\lambda_{2,3}) &\leq \max_z \left[6a - 4 + 6z - \frac{6}{a}z^2\right] \\ &= \frac{15}{2}a - 4.\end{aligned}$$

By using (15) and (16), we thus arrive at the following.

**Theorem 13.** Let  $K$  be a compact forward-invariant set of the system (13) with  $a \geq 2/3$ . Then

$$H_{\text{res}}(\varphi, K) \leq \frac{2(2a-1)}{\ln 2}.$$

Our next step is to estimate  $H_{\text{res}}(\varphi, K)$  from below under an extra assumption about  $K$ . Such an estimate is given by the so-called *proximate topological entropy*  $H_L(O)$  around the system equilibria  $O$ , which is defined as  $H_L(O) := \frac{1}{\ln 2} \sum_{j=1}^n \max\{\text{Re}\beta_j, 0\}$  ([Matveev & Pogromsky, 2019](#)). Here  $\beta_1, \dots, \beta_n$  are the eigenvalues of  $\mathbf{D}\varphi(O)$  repeated in accordance with their algebraic multiplicities and ordered so that their real parts  $\text{Re}\beta_i$  do not increase as  $i$  grows. It is easy to see that now

$$H_L(O_1) = \frac{1}{\ln 2} \begin{cases} a & \text{if } 0 < a \leq 1 \\ 3a-2 & \text{if } a \geq 1, \end{cases}$$

$$H_L(O_2) = \frac{1}{\ln 2} \begin{cases} 0 & \text{if } 0 < a \leq 1/2 \\ 2(2a-1) & \text{if } a \geq 1/2, \end{cases}$$

$$\max\{H_L(O_1), H_L(O_2)\} \stackrel{a \geq 2/3}{=} H_L(O_2) = \frac{2(2a-1)}{\ln 2}.$$

The last relation together with [Matveev and Pogromsky \(2019, Cor. 12\)](#) and [Theorem 13](#) gives the following result.

**Theorem 14.** Assume that  $a \geq 2/3$ . Let  $K$  be any compact forward-invariant set for system (13), which satisfies [Assumption 3](#) and the inclusion  $O_2 \in \text{int}K$ . Then

$$H_{\text{res}}(\varphi, K) = \frac{2(2a-1)}{\ln 2}.$$

At this point, it is worth mentioning that the matrix  $P$  from (14) not only provides an upper estimate of the restoration entropy according to the statement (i) of [Theorem 11](#), but also gives a Riemannian metric for which the lower estimate (see the statement (ii) of [Theorem 11](#)) holds true with  $\varepsilon = 0$ .

#### 5. Conclusions, future work, and final remarks

In this paper, we have provided a new characterization of the restoration entropy, which measures the minimal channel capacity above which a system can be regularly or finely observed from a remote location. This characterization is based on the observation that the upper bound provided in [Matveev and Pogromsky \(2019\)](#) in terms of the singular values of the system's linearization is true for any choice of the inner product with respect to which singular values are computed, even if the inner product varies from point to point, i.e., constitutes a Riemannian metric. Moreover, the corresponding metric tensor plays a role somewhat similar to the Lyapunov matrix from the direct Lyapunov method. The main results of this paper show that a proper choice of this metric leads to the exact value of the

restoration entropy, more precisely, the exact value is attained as the infimum value of the above bound over all metrics. This holds for both discrete- and continuous-time systems.

Moreover, our proofs explicitly disclose a minimizing sequence of metrics  $P_N(x)$ . This issue is discussed in more detail in [Appendix D](#). However, constructive methods for building such a sequence or even computing an exact minimizer are beyond the scope of this paper, partly due to the paper length limitations; they form a topic in the schedule of our ongoing and future research.

We further remark that the afore-required compactness for the set  $K$  is not, in fact, essential in our results. Indeed, already in the definition of restoration entropy, it is only necessary that the image of a closed  $\delta$ -ball is compact. Wherever we use compactness in our proof, it can be replaced by some uniformity condition, which is another important fact that distinguishes restoration entropy from topological entropy.

Finally, another possible generalization of our results concerns the abstraction of the state space from a Euclidean space to a differentiable manifold. Using the appropriate technical language, we think that our proofs carry over to this more general setup without the need of introducing new ideas.

## Acknowledgments

A. Pogromsky acknowledges his partial support by the UCoCoS project which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675080. A. Matveev acknowledges his support by the Ministry of Science and Higher Education of the Russian Federation (appl. 2020-220-08-4683). C. Kawan is supported by the German Research Foundation (DFG) through the grant ZA 873/4-1. A preliminary version of this paper was presented at the 2020 IFAC World Congress ([Kawan, Matveev, & Pogromsky, 2020](#)).

## Appendix

The remainder of the paper is devoted to the proofs of [Theorems 6](#) and [11](#). [Appendix A](#) summarizes some preliminaries required in the proofs of these results. The main result for the discrete-time case is proven in [Appendix B](#), while the proof for the continuous-time case is presented in [Appendix C](#). [Appendix D](#) highlights a technical conclusion from the proofs: it discusses an explicit formula for a minimizing sequence of metrics.

### Appendix A. Technical preamble to proofs of [Theorems 6](#) and [11](#)

From now on, we use the following notations:  $\text{Gl}(n, \mathbb{R})$  – group of real invertible  $n \times n$  matrices,  $I$  –  $n \times n$  identity matrix,  $\|\cdot\|_2$  – standard Euclidean norm on  $\mathbb{R}^n$ ,  $x_i$  – components of  $x \in \mathbb{R}^k$ .

#### A.1. Riemannian geometry on the set $S^+$

Since the set  $S^+$  is open in the vector space  $S$  of all symmetric  $n \times n$ -matrices, the first candidate for a Riemannian metric tensor on  $S^+$  is the Euclidean metric inherited from  $S$ . However, many applications motivate the use of the so-called *trace metric* ([Bhatia, 2007](#)). It varies with the point  $p \in S^+$  and defines the following inner product on the space  $T_p S^+$  tangential to  $S^+$  at  $p$ , which is in fact a copy of  $S$ :

$$\langle v, w \rangle_p := \text{tr}(p^{-1}vp^{-1}w), \quad \forall v, w \in T_p S^+ \cong S.$$

This metric makes  $S^+$  a complete Riemannian manifold ([Chavel, 2006](#)), assigns a special length to any smooth curve, and gives

birth to the Riemannian distance  $d(p, q)$ , defined as the minimal length of curves bridging the points  $p$  and  $q$  of  $S^+$ . The minimizer exists, is unique, is called a *geodesic*, and is given by [Bhatia \(2007, Thm. 6.1.6\)](#):

$$p \#_t q := p^{1/2}(p^{-1/2}qp^{-1/2})^t p^{1/2}, \quad t \in [0, 1]. \quad (\text{A.1})$$

Let  $\alpha_1(g) \geq \dots \geq \alpha_n(g) > 0$  be the singular values of  $g \in \text{Gl}(n, \mathbb{R})$ . We put

$$\begin{aligned} \bar{\sigma}(g) &:= [\log \alpha_1(g), \dots, \log \alpha_n(g)] \in \mathfrak{a}^+ \\ &:= \{\xi \in \mathbb{R}^n : \xi_1 \geq \xi_2 \geq \dots \geq \xi_n\} \end{aligned} \quad (\text{A.2})$$

and endow  $\mathfrak{a}^+$  with the following partial order:

$$\xi \leq \eta \stackrel{\text{def}}{\iff} \begin{cases} \sum_{i=1}^k \xi_i \leq \sum_{i=1}^k \eta_i & \forall 1 \leq k \leq n-1, \\ \sum_{i=1}^n \xi_i = \sum_{i=1}^n \eta_i & \text{for } k = n. \end{cases} \quad (\text{A.3})$$

Any matrix  $g \in \text{Gl}(n, \mathbb{R})$  defines an action (mapping)  $g*$  on  $S^+$ , specifically,  $p \in S^+ \mapsto g*p := gpg^T \in S^+$ . It is easy to see that  $g_1*(g_2*p) = (g_1g_2)*p$ ,  $I*p = p$ ,  $p_2 = g*p_1 \Leftrightarrow p_1 = g^{-1}*p_2$ , and for any  $p, q \in S^+$  there exists  $g \in \text{Gl}(n, \mathbb{R})$  (e.g.,  $g := q^{1/2}p^{-1/2}$ ) such that  $q = g*p$ .

The following proposition is based on [Bochi \(2018\)](#) and lists some properties of the so-called *vectorial distance*

$$\bar{d}(p, q) := 2\bar{\sigma}(p^{-1/2}q^{1/2}) \in \mathfrak{a}^+, \quad p, q \in S^+. \quad (\text{A.4})$$

**Proposition 15.** *The following statements hold:*

- (a)  $\bar{d}(I, p) = \bar{\sigma}(p)$  and  $\|\bar{d}(p, q)\|_2 = d(p, q) \forall p, q \in S^+$ ;
- (b)  $\bar{d}(p_1, q_1) = \bar{d}(p_2, q_2)$  if and only if  $g*p_1 = p_2$  and  $g*q_1 = q_2$  for some  $g \in \text{Gl}(n, \mathbb{R})$ ;
- (c)  $\bar{d}(p, p) = 0$ ;  $\bar{d}(p, q) \leq \bar{d}(p, r) + \bar{d}(r, q)$ ;
- (d)  $\bar{d}(q, p) = i(\bar{d}(p, q))$ , where  $i(\xi) := -(\xi_n, \xi_{n-1}, \dots, \xi_1)$ ;
- (e) The curve [\(A.1\)](#) is a geodesic segment for the vectorial distance, i.e., there is  $\xi \in \mathfrak{a}^+$  such that  $\bar{d}(p \#_t q, p \#_s q) = (s-t)\xi$  for all  $s \geq t$ ;
- (f)  $\bar{d}(r \#_{1/2} p, r \#_{1/2} q) \leq \frac{1}{2}\bar{d}(p, q)$  for all  $p, q, r \in S^+$ .

Here (b) shows that the map  $g*$  is an isometry (preserves distances); (c) and (d) are akin to the axioms of the metric, and (f) replicates formula (4–7) in [Bochi \(2018\)](#).

**Proposition 16.** *The following holds for geodesics on  $S^+$ :*

$$g*(p \#_t q) = (g*p) \#_t (g*q), \quad (\text{A.5})$$

$$\bar{d}(p \#_t q, r \#_t o) \leq (1-t)\bar{d}(p, r) + t\bar{d}(q, o) \forall t \in [0, 1]. \quad (\text{A.6})$$

**Proof.** Here [\(A.5\)](#) holds since the map  $g*$  is an isometry of the Riemannian space  $S^+$ . To prove [\(A.6\)](#), we use (c) and (f) in [Proposition 15](#) and the symmetry  $p \#_{\frac{1}{2}} q = q \#_{\frac{1}{2}} p$  ([Bhatia, 2007, Sec. 4.1, 6.1.7](#)) and see that

$$\begin{aligned} &\bar{d}(p \#_{1/2} q, r \#_{1/2} o) \\ &\leq \bar{d}(p \#_{1/2} q, p \#_{1/2} o) + \bar{d}(p \#_{1/2} o, r \#_{1/2} o) \\ &= \bar{d}(p \#_{1/2} q, p \#_{1/2} o) + \bar{d}(o \#_{1/2} p, o \#_{1/2} r) \\ &\leq \frac{1}{2}\bar{d}(q, o) + \frac{1}{2}\bar{d}(p, r), \end{aligned}$$

[\(A.6\)](#) with  $t = 1/2$ . For all numbers  $t$  of the form  $t = k/2^n$  with integers  $k, n$  such that  $0 \leq k \leq 2^n$ , [\(A.6\)](#) follows by induction. Since these numbers are dense in  $[0, 1]$  and [\(A.1\)](#) is continuous in  $t$ , [\(A.6\)](#) extends to the entirety of  $[0, 1]$  by continuity. ■

The point  $p \#_{1/2} q$  is in fact the barycenter of the two-point set  $\{p, q\}$ , i.e., the minimizer  $r$  of  $d(r, p)^2 + d(r, q)^2$ . This notion of

barycenter has a far-reaching generalization. Specifically, let

$$\Delta_m := \{\omega \in \mathbb{R}^m : \omega_i \geq 0, \sum \omega_i = 1\}$$

be the standard simplex in  $\mathbb{R}^m$ , let  $\omega \in \Delta_m$ , and let  $p_1, \dots, p_m \in S^+$ . Then the following point is well-defined (see, e.g., [Lim & Palfia, 2014](#)):

$$\text{bar}(\omega; p_1, \dots, p_m) := \arg \min_{q \in S^+} \sum_{i=1}^m \omega_i d(q, p_i)^2$$

and is called a *weighted barycenter* of  $p_1, \dots, p_m$  (also known under other names). The *unweighted* barycenter is given by

$$\text{bar}(p_1, \dots, p_m) := \text{bar}(\omega^=; p_1, \dots, p_m), \text{ where } \omega_i^= := m^{-1} \forall i.$$

There is no closed-form expression for  $\text{bar}(\omega; p_1, \dots, p_m)$ , in general. The following theorem from [Lim and Palfia \(2014\)](#) (see [Holbrook, 2012](#) for the unweighted case) characterizes the barycenter as the limit of a certain iterative process, which acts through a binary operation at each step  $k = 1, 2, \dots$ . Specifically, let  $i(\text{mod } m)$  stand for the remainder after dividing  $i$  by  $m$ . We put  $l(k) := \sum_{i=1}^k \omega_{i(\text{mod } m)}$ ,  $s_k := \omega_{k(\text{mod } m)} / l(k)$  and consider the following process:

$$\bar{p}_1 := p_1, \quad \bar{p}_k := \bar{p}_{k-1} \#_{s_k} p_{k(\text{mod } m)}.$$

**Theorem 17** ([Holbrook, 2012](#); [Lim & Palfia, 2014](#)).  $\text{bar}(\omega; p_1, \dots, p_m) = \lim_{k \rightarrow \infty} \bar{p}_k$ .

Now we list some properties of the weighted barycenter.

**Proposition 18.** *The following relations hold:*

$$g * \text{bar}(\omega; p_1, \dots, p_m) = \text{bar}(\omega; g * p_1, \dots, g * p_m) \quad \forall g \in \text{Gl}(n, \mathbb{R}); \quad (\text{A.7})$$

$$\bar{d}(u, v) \leq \omega_m \bar{d}(p_m, p'_m) \quad \forall p_1, \dots, p_m, p'_m \in S^+, \quad (\text{A.8})$$

$$\text{where } \begin{cases} u = \text{bar}(\omega; p_1, \dots, p_{m-1}, p_m), \\ v = \text{bar}(\omega; p_1, \dots, p_{m-1}, p'_m); \end{cases}$$

$$\text{bar}(\omega; p_1, \dots, p_m) = \text{bar}(\omega_\sigma; p_{\sigma(1)}, \dots, p_{\sigma(m)}) \quad (\text{A.9})$$

for any permutation  $\sigma$  of  $\{1, \dots, m\}$ , where  $\omega_\sigma$  is the result  $(\omega_{\sigma(1)}, \dots, \omega_{\sigma(m)})$  of its action on  $\omega$ .

**Proof.** Here (A.7) and (A.9) are straightforward from the definition of the barycenter and the fact that  $g*$  is an isometry. To prove (A.8), we consider the sequences  $(\bar{p}_k)_{k \in \mathbb{N}}$  and  $(\bar{q}_k)_{k \in \mathbb{N}}$  defined as in [Theorem 17](#) from the data  $(\omega; p_1, \dots, p_{m-1}, p_m)$  and  $(\omega; p_1, \dots, p_{m-1}, p'_m)$ , respectively. By using (A.6), we obtain

$$\begin{aligned} \bar{d}(\bar{p}_{km}, \bar{q}_{km}) &= \bar{d}(\bar{p}_{km-1} \#_{s_{km}} p_m, \bar{q}_{km-1} \#_{s_{km}} p'_m) \\ &\leq (1 - s_{km}) \bar{d}(\bar{p}_{km-1}, \bar{q}_{km-1}) + s_{km} \bar{d}(p_m, p'_m) \\ &= (1 - s_{km}) \bar{d}(\bar{p}_{km-2} \#_{s_{km-1}} p_{m-1}, \bar{q}_{km-2} \#_{s_{km-1}} p'_{m-1}) \\ &\quad + s_{km} \bar{d}(p_m, p'_m) \leq (1 - s_{km}) \left[ (1 - s_{km-1}) \bar{d}(\bar{p}_{km-2}, \bar{q}_{km-2}) \right. \\ &\quad \left. + s_{km-1} \underbrace{\bar{d}(p_{m-1}, p'_{m-1})}_{=0} \right] + s_{km} \bar{d}(p_m, p'_m) \\ &= (1 - s_{km})(1 - s_{km-1}) \bar{d}(\bar{p}_{km-2}, \bar{q}_{km-2}) + s_{km} \bar{d}(p_m, p'_m) \\ &\leq \dots \\ &\leq \left[ \prod_{i=0}^{m-1} (1 - s_{km-i}) \right] \bar{d}(\bar{p}_{(k-1)m}, \bar{q}_{(k-1)m}) + s_{km} \bar{d}(p_m, p'_m). \end{aligned}$$

Now we observe that  $s_{km} = \omega_m / k$  and

$$\prod_{i=0}^{m-1} (1 - s_{km-i}) = \prod_{i=0}^{m-1} \left( 1 - \frac{\omega_{m-i}}{k-1 + \sum_{j=1}^{m-i} \omega_j} \right)$$

$$= \prod_{i=0}^{m-1} \frac{k-1 + \sum_{j=1}^{m-i} \omega_j - \omega_{m-i}}{k-1 + \sum_{j=1}^{m-i} \omega_j} = \frac{k-1}{k}.$$

Hence, we end up with

$$\bar{d}(\bar{p}_{km}, \bar{q}_{km}) \leq \frac{k-1}{k} \bar{d}(\bar{p}_{(k-1)m}, \bar{q}_{(k-1)m}) + \frac{\omega_m}{k} \bar{d}(p_m, p'_m).$$

Iterating this estimate all the way down to  $k = 1$ , we find that

$$\begin{aligned} \bar{d}(\bar{p}_{km}, \bar{q}_{km}) &\leq \left( \frac{\omega_m}{k} + \frac{k-1}{k} \frac{\omega_m}{k-1} + \frac{k-1}{k} \frac{k-2}{k-1} \frac{\omega_m}{k-2} + \dots \right. \\ &\quad \left. + \prod_{i=1}^k \frac{k-i}{k-i+1} \omega_m \right) \bar{d}(p_m, p'_m) = \omega_m \bar{d}(p_m, p'_m). \end{aligned}$$

Letting  $k \rightarrow \infty$  with regard to [Theorem 17](#) completes the proof. ■

The set of  $m$  points  $p_i \in S^+$  and the tuple of weights  $\omega \in \Delta_m$  give birth to a probability measure on  $S^+$ : this is the convex combination  $\mu$  of  $m$  Dirac measures  $\delta_{p_i}$  with the coefficients  $\omega_i$ . The weighted barycenter is the minimizer  $p$  of the function

$$p \mapsto \int_{S^+} d(p, q)^2 d\mu(q). \quad (\text{A.10})$$

This observation permits to extend the notion of the barycenter to any probability measure  $\mu$  on the Borel  $\sigma$ -algebra of  $S^+$ .

More rigorously, let  $\mathcal{P}^1(S^+)$  be the set of all Borel probability measures on  $S^+$  with finite first moment, i.e.,

$$\mathcal{P}^1(S^+) := \left\{ \mu : \int_{S^+} d(I, p) d\mu(p) < \infty \right\}.$$

This set can be equipped with the 1-Wasserstein metric ([Villani, 2009](#)):

$$W_1(\mu, \nu) := \inf_{P \in (\mu|\nu)} \int_{S^+ \times S^+} d(p, q) dP(p, q),$$

where  $(\mu|\nu)$  is the set of all probability measures  $P$  on  $S^+ \times S^+$  whose projection along the first and second coordinate coincides with  $\mu$  and  $\nu$ , respectively. Simultaneously ([Villani, 2009](#), Rem. 6.5),

$$W_1(\mu, \nu) = \sup_{\|\psi\|_{\text{Lip}} \leq 1} \left[ \int_{S^+} \psi d\mu - \int_{S^+} \psi d\nu \right], \quad (\text{A.11})$$

where  $\|\psi\|_{\text{Lip}} := \sup_{p_1 \neq p_2, p_i \in S^+} |\psi(p_1) - \psi(p_2)| / d(p_1, p_2)$ .

Rem. 4.2 and Lem. 5.1 in [Sturm \(2003\)](#) give rise to the following.

**Lemma 19.** (i) *The sums of Dirac measures with identical weights  $\frac{1}{m}(\delta_{p_1} + \dots + \delta_{p_m})$  form, in total, a set that is dense in  $\mathcal{P}^1(S^+)$ .* (ii) *For any isometry  $\mathcal{J} : S^+ \rightarrow S^+$ , the associated push-forward transformation of probability measures  $\mu \mapsto \nu, \nu(E) := \mu(\mathcal{J}^{-1}E)$  isometrically maps  $\mathcal{P}^1(S^+)$  into itself.*

The next proposition asserts existence of the measure barycenter and reports some its properties.

**Proposition 20** (Lem. 5.1, Thm. 6.3, Prop. 4.3 in [Sturm, 2003](#)). *There exists a map  $\text{bar} : \mathcal{P}^1(S^+) \rightarrow S^+$  with the following properties:*

- (1) *Whenever an isometry  $\mathcal{J} : S^+ \rightarrow S^+$  pushes a probability measure  $\mu$  forward into the measure  $\nu$ , we have  $\text{bar}[\nu] = \mathcal{J}(\text{bar}[\mu])$ ;*
- (2) *The map  $\text{bar}[\cdot]$  is 1-Lipschitz:*

$$d(\text{bar}[\mu], \text{bar}[\nu]) \leq W_1(\mu, \nu) \quad \forall \mu, \nu \in \mathcal{P}^1(S^+); \quad (\text{A.12})$$

- (3) *The minimum of (A.10) is attained at  $\text{bar}[\mu]$  provided that the r.h.s. is well-defined, i.e.,  $\int_{S^+} d(I, q)^2 d\mu(q) < \infty$ .*

The last inequality implies that in (A.10),  $\int_{S^+} d(p, q)^2 d\mu(q) < \infty$  for any  $p \in S^+$  since  $d(p, q) \leq d(p, I) + d(I, q)$  and the constant  $d(p, I)$  is of class  $L_2$  with respect to the probability measure.

**Corollary 21.** *The unweighted barycenter  $\text{bar}(p_1, \dots, p_m)$  continuously depends on  $p_i \in S^+$ .*

This is immediate from (2) in Proposition 20, since (A.11) implies that  $W_1(\mu, \nu) \leq \max_i d(p_i, p'_i)$  if  $\mu$  and  $\nu$  are the sums of equi-weighted Dirac measures at points  $p_1, \dots, p_m$  and  $p'_1, \dots, p'_m$ , respectively.

## A.2. Derivatives of singular values and other matrix functionals

The  $m$ th largest eigenvalue of a symmetric matrix  $A \in \mathcal{S}$  is the  $m$ th term  $\lambda_m(A)$  of the nonincreasing sequence  $\lambda_1(A) \geq \dots \geq \lambda_n(A)$  of its eigenvalues, each repeated with regard to its algebraic multiplicity  $r_m$ . We put  $i_m := m - m_{\text{first}} + 1$ , where  $m_{\text{first}}$  is the first position in the sequence that accommodates the number  $\lambda_m(A)$ , whereas  $m_{\text{last}}$  is the last position:

$$\lambda_1(A) \geq \dots \geq \lambda_{m-i_m}(A) > \lambda_{m_{\text{first}}}(A) = \dots = \lambda_{m_{\text{last}}}(A) > \lambda_{m_{\text{last}}+1}(A) \geq \dots \geq \lambda_n(A).$$

Also let  $U$  be an orthogonal matrix that reduces  $A$  to a diagonal matrix  $U^T A U$  with nonincreasing diagonal entries. We denote by  $\tilde{U}_m$  the matrix that is obtained by depriving  $U$  of all columns, except for those in the range from  $m_{\text{first}}$  to  $m_{\text{last}}$ .

**Theorem 22** (Hiriart-Urruty & Ye, 1995, Thm. 4.5). *Let  $\mathcal{O} \subset \mathbb{R}^p$  be an open set and  $A : \mathcal{O} \rightarrow \mathcal{S}$  be a map of class  $C^1$ . For any  $m = 1, \dots, n$ ,  $\xi_0 \in \mathcal{O}$  and unit-length vector  $d \in \mathbb{R}^p$ , the derivative*

$$\lim_{\varepsilon \rightarrow 0+} \frac{f(\xi_0 + \varepsilon d) - f(\xi_0)}{\varepsilon}$$

*of the function  $\xi \in \mathcal{O} \mapsto f(\xi) := \lambda_m[A(\xi)]$  in direction  $d$  exists. It equals the  $i_m$ -largest eigenvalue of the following  $r_m \times r_m$  matrix built from  $A(\cdot)$  (along with  $\tilde{U}_m, r_m, i_m, \xi_0$ )*

$$\tilde{U}_m^T \left[ \sum_{j=1}^p d_j \frac{\partial A}{\partial \xi_j}(\xi_0) \right] \tilde{U}_m.$$

**Corollary 23.** *For any  $C^1$ -smooth function  $g : \mathbb{R} \rightarrow \text{Gl}(n, \mathbb{R})$  with  $g(0) = I$  and  $\dot{g}(0) = H \in \mathbb{R}^{n \times n}$ , the l.h.s. of (A.2) with  $g = g(t)$  has the right derivative at  $t = 0$  and*

$$\frac{d}{dt} \bar{\sigma}[g(t)] \Big|_{t=0+} = \frac{1}{2 \ln(2)} [\lambda_1(H + H^T), \dots, \lambda_n(H + H^T)].$$

**Proof.** The map  $t \mapsto A(t) := g(t)g(t)^T \in \mathcal{S}$  is of class  $C^1$  and

$$\bar{\sigma}[g(t)] \stackrel{(A.2)}{=} \frac{1}{2} [\log \lambda_1(A(t)), \dots, \log \lambda_n(A(t))], \quad (A.13)$$

$$A(0) = I \Rightarrow r_m = n, i_m = m, \tilde{U}_m = I, \quad \dot{A}(0) = H + H^T.$$

By applying Theorem 22 to  $p = 1$ ,  $\mathcal{O} := \mathbb{R}$ ,  $\xi_0 := 0$ , we see that the right derivative of  $\lambda_m[A(t)]$  exists and is equal to  $\lambda_m[\dot{A}(0)] = \lambda_m(H + H^T)$ . This and (A.13) complete the proof. ■

**Lemma 24.** *The function  $(p, q) \in S^+ \times S^+ \mapsto \Upsilon(p, q) := p^{-1/2} q^{1/2} \in \text{Gl}(n, \mathbb{R})$  is of class  $C^1$  and its derivative at any point with  $p = q$  is given by*

$$\mathbf{D}\Upsilon(p, p) \begin{bmatrix} v_p \\ v_q \end{bmatrix} = p^{-1/2} h(v_q - v_p) \quad \forall v_p, v_q \in \mathcal{S},$$

where  $h = h(v) \in \mathcal{S}$  is the unique solution of the equation

$$hp^{1/2} + p^{1/2}h = v \in \mathcal{S}. \quad (A.14)$$

**Proof.** Existence and uniqueness of the solution for the Lyapunov equation (A.14) is a well-known fact; see e.g., Khalil (2002, Ch. 3). We put  $\mathcal{E}(p) := p^{-1}$ ,  $\Omega_{\pm}(p) := p^{\pm 1/2} \forall p \in S^+$ . It is immediate from Del Moral and Niclas (2018, Thm. 1.1) and van den Bos (2007, Thm. D2) that  $\mathcal{E}, \Omega, \Upsilon$  are  $C^1$ -smooth and

$$\mathbf{D}\mathcal{E}(p)v = -p^{-1}vp^{-1}, \quad \mathbf{D}\Omega_{\pm}(p)v = h(v) \quad \forall v \in \mathcal{S}.$$

Since  $\Omega_- = \mathcal{E} \circ \Omega_+$ , we have  $\mathbf{D}\Omega_-(p) = \mathbf{D}\mathcal{E}[\Omega_+(p)] \circ \mathbf{D}\Omega_+(p) \Leftrightarrow \mathbf{D}\Omega_-(p)v = -p^{-1/2}h(v)p^{-1/2}$ . Finally,  $\Upsilon(p, q) = \Omega_-(p) \times \Omega_+(q)$ , hence

$$\begin{aligned} \mathbf{D}\Upsilon(p, p) \begin{bmatrix} v_p \\ v_q \end{bmatrix} &= \mathbf{D}\Omega_-(p)v_p \times \Omega_+(p) + \Omega_-(p) \times \mathbf{D}\Omega_+(p)v_q \\ &= -p^{-1/2}h(v_p)p^{-1/2} \times p^{1/2} + p^{-1/2} \times h(v_q) \\ &= p^{-1/2}[h(v_q) - h(v_p)] = p^{-1/2}h(v_q - v_p). \quad \blacksquare \end{aligned}$$

## A.3. Linear cocycles

A dynamical system on a metric space  $X$  is given by its evolution function (dynamic semiflow)  $\Phi : \mathbb{T}_+ \times X \rightarrow X$ , where  $\mathbb{T}_+$  is as in (1) and  $\Phi(0, x) = x \forall x$ ,  $\Phi[s, \Phi(t, x)] = \Phi[s+t, x] \forall s, t \in \mathbb{T}_+, x \in X$ . For this system, a linear cocycle is defined as a mapping  $(t, x) \mapsto A^{(t)}(x) \in \text{Gl}(k, \mathbb{R})$  such that

$$A^{(0)}(x) = I \text{ and } A^{(s+t)}(x) = A^{(s)}[\Phi(t, x)]A^{(t)}(x) \quad (A.15)$$

for all  $x \in X$ ,  $s, t \in \mathbb{T}_+$ . In the discrete-time case, the semiflow is determined by  $\Psi(\cdot) := \Phi(1, \cdot)$  (specifically,  $\Phi(t, \cdot)$  is the  $t$ th iterate of  $\Psi(\cdot)$  for  $t > 0$  and the identity map for  $t = 0$ ), and the cocycle by its generator  $\mathcal{A}(\cdot) := A^{(1)}(\cdot)$  since (A.15) shapes into

$$A^{(0)}(x) = I, \quad A^{(t)}(x) = \mathcal{A}[\Psi^{t-1}(x)] \dots \mathcal{A}[\Psi^1(x)]\mathcal{A}[x].$$

Two linear cocycles  $A, B : \mathbb{T}_+ \times X \rightarrow \text{Gl}(k, \mathbb{R})$  for a common dynamical system are said to be conjugate if there exists a continuous map  $V : X \rightarrow \text{Gl}(k, \mathbb{R})$  (the conjugacy) such that  $B^{(t)}(x) = V[\Phi(t, x)]^{-1}A^{(t)}(x)V(x) \forall x \in X, t \in \mathbb{T}_+$ . In the discrete-time case, it suffices to test only  $t = 1$ , i.e., the relation  $\mathcal{B}(x) = V[\Psi(x)]^{-1}\mathcal{A}(x)V(x) \forall x \in X$ , where  $\mathcal{B}(\cdot) := B^{(1)}(\cdot)$ .

## Appendix B. Proof of Theorem 6

In this section, the assumptions of this theorem are adopted; in particular, we consider the discrete-time system (1) (the case (d-t)) and denote by  $\mathbb{Z}_+$  the set of all nonnegative integers. The first step towards proving (i) in Theorem 6 is the following.

**Lemma 25.** *There are  $C_{\pm} \in \mathbb{R}$  such that for all  $t \in \mathbb{Z}_+, x \in K$ ,*

$$-\infty < C_- \leq \Sigma^P(t, x) - \Sigma^I(t, x) \leq C_+ < \infty. \quad (B.1)$$

Here  $\Sigma^P(t, x) := \Sigma^P[x|\varphi^t]$  is defined in (8) and so  $\Sigma^I(t, x)$  is built from  $\alpha_i^I(t, x)$ , which are the ordinary singular values (by the definition in Lemma 5).

**Proof.** Let  $\omega_k(C)$  be the product of  $k$  largest singular values of the square matrix  $C$  if  $k \geq 1$ ; and  $\omega_0(C) := 1$ . We apply Lemma 5 to the  $t$ th iterate  $\phi := \varphi^t$  of  $\varphi$ , denote by  $A^{(t)}(x)$  and  $B^{(t)}(x)$  the respective  $A(x)$  and  $B(x)$  from that lemma, and see that thanks to the last claim of that lemma,

$$\Sigma^P(t, x) - \Sigma^I(t, x) = \log \frac{\max_{0 \leq k \leq n} \omega_k(B^{(t)}(x))}{\max_{0 \leq k \leq n} \omega_k(A^{(t)}(x))}.$$

By Horn's inequality (Boichenko, Leonov, & Reitman, 2005, Ch. I, Prop. 2.3.1), see also (Pogromsky & Matveev, 2011, Lemma 8.3)

$$\omega_k(B^{(t)}(x)) \leq \omega_k(P(\varphi^t(x)))^{1/2} \omega_k(A^{(t)}(x)) \omega_k(P(x)^{-1/2}).$$



Here  $x \in K \Rightarrow y := \varphi^t(x) \in K \ \forall t$  by [Assumption 2](#). Since the singular values continuously depend on the matrix ([Horn & Johnson, 2013](#), Thm. 2.6.4), so does  $\omega_k$ . Hence the following functions are continuous as well:

$$y \in K \mapsto \omega_k(P(y)^{1/2}) \quad \text{and} \quad y \in K \mapsto \omega_k(P(y)^{-1/2})$$

So the maximum over the compact set  $K$  is attained and finite for both quantities. This observation yields the upper estimate in [\(B.1\)](#). The lower estimate is obtained likewise by applying Horn's inequality to  $\omega_k(A^{(t)}(x))$ . Thus we see that [\(B.1\)](#) does hold. ■

**Proof of (i) in Theorem 6.** By combining [\(B.1\)](#) with [Remark 4](#), we get

$$H_{\text{res}}(\varphi, K) \leq \max_{x \in K} \limsup_{t \rightarrow \infty} t^{-1} \Sigma^P(t, x). \quad (\text{B.2})$$

The generalized Horn's inequality ([Boichenko et al., 2005](#), Ch. I, Prop. 7.4.3) implies that  $\Sigma^P[t + s, x] \leq \Sigma^P[s, \varphi^t(x)] + \Sigma^P[t, x]$ . Then, by [Morris \(2013, Thm. A.3\)](#) and the definition of  $\Sigma^P(t, x)$ , the r.h.s. of [\(B.2\)](#) equals

$$\begin{aligned} & \inf_{t > 0} \max_{x \in K} \frac{1}{t} \sum_{i=1}^n \max\{0, \log \alpha_i^P(t, x)\} \\ & \leq \max_{x \in K} \sum_{i=1}^n \max\{0, \log \alpha_i^P(x|\varphi)\}. \quad \blacksquare \end{aligned}$$

A key to the proof of [\(ii\)](#) in [Theorem 6](#) is the following lemma, which uses the concepts and notations from [Appendix A.3](#).

**Lemma 26.** *Let  $A$  be a continuous linear cocycle over a continuous discrete-time dynamical system  $\Phi$  on a compact metric space  $X$ . For any natural  $N$ , there exists a continuous cocycle  $B$ , conjugate to  $A$ , such that for its generator  $\mathcal{B}$ , the following holds:*

$$\bar{\sigma}[\mathcal{B}(x)] \leq N^{-1} \bar{\sigma}[A^{(N)}(x)] \quad \text{for all } x \in X. \quad (\text{B.3})$$

The involved conjugacy  $V$  can be chosen so that  $V(x) \in S^+ \ \forall x$ .

**Proof.** For  $x \in X$ , let  $Q(x) \in S^+$  be the unweighted barycenter:

$$Q(x) := \text{bar}(I, A^{(1)}(x)^{-1} * I, \dots, A^{(N-1)}(x)^{-1} * I). \quad (\text{B.4})$$

Recall that we denote the time-1 map of  $\Phi$  by  $\psi$ . By [\(A.7\)](#) and [\(A.9\)](#), then the matrix  $A^{(1)}(x) * Q(x)$  is equal to

$$\begin{aligned} & \text{bar}(A^{(1)}(x) * I, I, A^{(1)}(\psi(x))^{-1} * I, \dots, A^{(N-2)}(\psi(x))^{-1} * I) \\ & = \text{bar}(I, A^{(1)}(\psi(x))^{-1} * I, \dots, A^{(N-2)}(\psi(x))^{-1} * I, A^{(1)}(x) * I). \end{aligned}$$

Meanwhile,  $Q[\psi(x)]$  is the following matrix

$$\text{bar}(I, A^{(1)}(\psi(x))^{-1} * I, \dots, A^{(N-1)}(\psi(x))^{-1} * I).$$

For the generator  $\mathcal{A}(x) = A^{(1)}(x)$ , we thus have

$$\begin{aligned} \bar{d}[Q(\psi(x)), \mathcal{A}(x) * Q(x)] & \stackrel{(\text{A.8})}{\leq} \frac{\bar{d}[A^{(N-1)}(\psi(x))^{-1} * I, \mathcal{A}(x) * I]}{N} \\ & \stackrel{(\text{b) in Proposition 15}}{\leq} N^{-1} \bar{d}[I, A^{(N)}(x) * I]. \end{aligned} \quad (\text{B.5})$$

Thanks to [Corollary 21](#), the maps  $Q(\cdot)$  and  $x \in X \mapsto V(x) := Q(x)^{-1/2} \in S^+$  are continuous; also,  $V(x) * Q(x) = Q(x)^{-1/2} Q(x) Q(x)^{-1/2} = I$ . For the conjugate linear cocycle generated by  $\mathcal{B}(x) := V(\psi(x)) \mathcal{A}(x) V(x)^{-1}$ , we have

$$\begin{aligned} \bar{\sigma}[\mathcal{B}(x) * I] & = \bar{d}[I, \mathcal{B}(x) * I] \\ & \stackrel{(\text{b) in Proposition 15}}{=} \bar{d}\{Q[\psi(x)]^{1/2} * I, Q[\psi(x)]^{1/2} \mathcal{B}(x) * I\} \\ & = \bar{d}\{Q[\psi(x)], V[\psi(x)]^{-1} \mathcal{B}(x) * I\} \\ & = \bar{d}\{Q[\psi(x)], \mathcal{A}(x) V(x)^{-1} * I\} = \bar{d}\{Q[\psi(x)], \mathcal{A}(x) * Q(x)\} \end{aligned}$$

$$\stackrel{(\text{B.5})}{\leq} N^{-1} \bar{d}[I, A^{(N)}(x) * I] \stackrel{(\text{A.4})}{=} N^{-1} \bar{\sigma}[A^{(N)}(x) * I].$$

To get [\(B.3\)](#), it remains to note that  $\bar{\sigma}(g * I) = 2\bar{\sigma}(g)$  by [\(A.2\)](#). ■

**Proof of (ii) in Theorem 6.** By [\(4\)](#), there exists  $N$  such that

$$\begin{aligned} H_{\text{res}}(\varphi, K) & \geq \frac{1}{N} \max_{x \in K} \sum_{i=1}^n \max\{0, \log \alpha_i(N, x)\} - \varepsilon \\ & \stackrel{(\text{a})}{=} \max_{x \in K} \frac{1}{N} \max_{k=0, \dots, n} \sum_{i=1}^k \log \alpha_i(N, x) - \varepsilon, \end{aligned} \quad (\text{B.6})$$

where  $\sum_{i=1}^0 \dots := 0$ , and (a) is due to the fact that, first, we put the singular values in the nonincreasing order, and, second,  $\max\{0, \log \alpha_i(N, x)\} = 0$  whenever  $\alpha_i(N, x) \leq 1$ . We apply [Lemma 26](#) to the system [\(1\)](#) to acquire a cocycle  $B$  and conjugacy  $V(x) \in S^+$ . By [\(A.2\)](#), [\(A.3\)](#), [\(B.3\)](#), the singular values  $\alpha_i[\mathcal{B}(x)]$  of  $\mathcal{B}(x)$  are such that

$$\sum_{i=0}^k \log \alpha_i[\mathcal{B}(x)] \leq \frac{1}{N} \sum_{i=0}^k \log \alpha_i(N, x) \quad \forall k = 0, \dots, n$$

$$\Rightarrow H_{\text{res}}(\varphi, K) \geq \max_{x \in K} \max_{k=0, \dots, n} \sum_{i=1}^k \log \alpha_i[\mathcal{B}(x)] - \varepsilon$$

$$= \max_{x \in K} \sum_{i=1}^n \max\{0, \log \alpha_i[\mathcal{B}(x)]\} - \varepsilon.$$

It remains to note that  $\mathcal{B}(x) = V[\varphi(x)] \mathcal{A}(x) V(x)^{-1}$  for all  $x \in K$  by the proof of [Lemma 26](#), and so  $\alpha_i[\mathcal{B}(x)] = \alpha_i^P(x|\varphi)$  for  $P(x) := V(x)^2$  by the last claim of [Lemma 5](#). ■

## Appendix C. Proof of Theorem 11

In this section, the assumptions of this theorem are adopted; in particular, the system [\(1\)](#) is described by ODE  $\dot{x} = \varphi(x)$ . We use the dynamical flow  $\varphi^t$  introduced just after [Assumption 2](#) and the associated linear cocycle  $(t, x) \mapsto A^{(t)}(x) := \mathbf{D}\varphi^t(x) \in \text{Gl}(n, \mathbb{R})$ .

**Proof of (i) in Theorem 11.** is by following the arguments from the proof of Thm. 14 in [Matveev and Pogromsky \(2019\)](#), where  $v_d(x) \equiv 0$ . The only difference is that the claim  $P(\cdot) \in C^1$  from [Matveev and Pogromsky \(2019\)](#) is relaxed to [Assumption 9](#). But this does not destroy these arguments, which can be seen by inspection that encompasses the proof of the underlying Prop. 8.6 in [Pogromsky and Matveev \(2011\)](#).

In proving [\(ii\)](#), the role of [Lemma 26](#) is played by the following.

**Proposition 27.** *For every  $T > 0$ , there exists a mapping  $P : K \rightarrow S^+$  that obeys [Assumption 9](#) and the following inequality*

$$\frac{1}{2 \ln(2)} [\varsigma_1^P(x), \dots, \varsigma_n^P(x)] \leq \frac{1}{T} \bar{\sigma}[A^{(T)}(x)] \quad \forall x \in K, \quad (\text{C.1})$$

where  $\varsigma_1^P(x) \geq \varsigma_2^P(x) \geq \dots \geq \varsigma_n^P(x)$  are the solutions of [\(10\)](#).

Its proof is broken into a string of several lemmata, which deal with a given  $T > 0$ . The notation  $\int_a^b \delta_{\gamma(t)} dt$  stands for the measure on  $S^+$  obtained by pushing forward the Lebesgue measure on  $[a, b]$  by a continuous curve  $\gamma : [a, b] \rightarrow S^+$ . We put

$$\mu_{t \rightarrow \tau|x} := \frac{1}{\tau - t} \int_t^\tau \delta_{A^{(s)}(x)^{-1} * I} ds, \quad \forall t < \tau; \quad (\text{C.2})$$

$$Q(x) := \text{bar}[\mu_{0 \rightarrow T|x}] \in S^+, \quad x \in K, \quad (\text{C.3})$$

$$\text{where } g * p = gpg^\top \in S^+ \ \forall g \in \text{Gl}(n, \mathbb{R}), p \in S^+. \quad (\text{C.4})$$

**Lemma 28.** The mapping (C.3) is continuous.

**Proof.** For two continuous  $\gamma_i : [a, b] \rightarrow S^+$  and  $\psi \in C^0(S^+, \mathbb{R})$  with  $\|\psi\|_{\text{Lip}} \leq 1$ , the following holds for the probability measures  $\mu_i = \frac{1}{b-a} \int_a^b \delta_{\gamma_i(t)} dt$  by the definition of “pushing forward”

$$\begin{aligned} \left| \int \psi d\mu_1 - \int \psi d\mu_2 \right| &= \frac{1}{b-a} \left| \int_a^b \{\psi[\gamma_1(t)] - \psi[\gamma_2(t)]\} dt \right| \\ &\leq \frac{1}{b-a} \int_a^b |\psi[\gamma_1(t)] - \psi[\gamma_2(t)]| dt \\ &\leq \frac{1}{b-a} \int_a^b d[\gamma_1(t), \gamma_2(t)] dt. \end{aligned}$$

So (A.11) yields that  $W_1(\mu_1, \mu_2) \leq \int_a^b d[\gamma_1(t), \gamma_2(t)] dt \rightarrow 0$  as  $\gamma_2(t)$  goes to  $\gamma_1(t)$  uniformly in  $t \in [a, b]$ . It remains to note that this uniform convergence does hold for  $\gamma_i(t) := A^{(t)}(x_i)^{-1} * I$  as  $x_2 \rightarrow x_1$  and to employ (2) in Proposition 20. ■

**Lemma 29.** For any continuous curve  $\gamma : [a, b] \rightarrow S^+$  and map  $\mathcal{J} : S^+ \rightarrow S^+$ , pushing the measure  $\mu := \int_a^b \delta_{\gamma(t)} dt$  forward by  $\mathcal{J}$  results in the measure  $\nu := \int_a^b \delta_{\mathcal{J}[\gamma(t)]} dt$ . In particular, pushing  $\mu_{0 \rightarrow T|\varphi^t(x)}$  forward by the map  $p \in S^+ \mapsto A^{(t)}(x)^{-1} * p$  results in  $\mu_{t \rightarrow t+T|x}$ .

**Proof.** For any  $\psi \in C^0(S^+, \mathbb{R})$ , the change-of-variables formula for the push-forward measure yields

$$\int_{S^+} \psi d\nu = \int_{S^+} \psi \circ \mathcal{J} d\mu = \int_a^b \psi\{\mathcal{J}[\gamma(t)]\} dt.$$

It remains to note that, by the same argument, the last expression is the integral of  $\psi$  with respect to  $\int_a^b \delta_{\mathcal{J}[\gamma(t)]} dt$ . Now, for  $\mu_{0 \rightarrow T|\varphi^t(x)}$  and  $p \mapsto A^{(t)}(x)^{-1} * p$ , we obtain the measure  $\frac{1}{T} \int_0^T \delta_{A^{(t)}(x)^{-1} * [A^{(s)}(\varphi^t(x))^{-1} * I]} ds$  and it remains to invoke (A.15) (with  $\Phi(t, x) := \varphi^t(x)$ ) and (C.4). ■

Now we recall that a certain  $T > 0$  was fixed just after Proposition 27.

**Lemma 30.** For any  $\varepsilon > 0$  and the vector  $\mathbb{1} \in \mathbb{R}^n$  composed of 1's, the following relation holds for all small enough  $t > 0$ :

$$\frac{\vec{d}\{Q[\varphi^t(x)], A^{(t)}(x) * Q(x)\}}{t} \leq 2 \frac{\vec{\sigma}[A^{(T)}(x)]}{T} + \varepsilon \mathbb{1}. \quad (\text{C.5})$$

**Proof.** Since  $A^{(s)}(x)^{-1} * I \in S^+$  is continuous in  $s$ , there exists  $\theta \in (0, T)$  such that whenever  $|s - \tau| \leq \theta$  and  $\tau \in [0, T]$ , we have

$$d[A^{(\tau)}(x)^{-1} * I, A^{(s)}(x)^{-1} * I] \leq T\varepsilon/2. \quad (\text{C.6})$$

For  $t \in (0, \theta)$  and the map (C.3), we have

$$A^{(t)}(x)^{-1} * Q[\varphi^t(x)] \stackrel{(\text{C.3})}{=} A^{(t)}(x)^{-1} * \text{bar}[\mu_{0 \rightarrow T|\varphi^t(x)}]$$

(a,b) in Proposition 15 and (1) in Proposition 20

$$\text{bar}[\text{the measure addressed in Lemma 29}]$$

$$\stackrel{\text{Lemma 29}}{=} \text{bar}[\mu_{t \rightarrow t+T|x}];$$

and

$$\vec{d}\{Q[\varphi^t(x)], A^{(t)}(x) * Q(x)\}$$

$$\stackrel{(\text{b}) \text{ in Proposition 15}}{=} \vec{d}\{A^{(t)}(x)^{-1} * Q[\varphi^t(x)], Q(x)\}$$

$$= \vec{d}(\text{bar}[\mu_{t \rightarrow t+T|x}], \text{bar}[\mu_{0 \rightarrow T|x}]).$$

$$\mu_{t \rightarrow t+T|x} \stackrel{(\text{C.2})}{=} \underbrace{(1-t/T)}_{=:a_t} \underbrace{\mu_{t \rightarrow T|x}}_{=: \mu} + \underbrace{t/T}_{=:b_t} \mu_{T \rightarrow T+t|x},$$

$$\mu_{0 \rightarrow T|x} \stackrel{(\text{C.2})}{=} t/T \mu_{0 \rightarrow t|x} + (1-t/T) \mu_{t \rightarrow T|x}.$$

$$\vec{d}(\text{bar}[\mu_{t \rightarrow t+T|x}], \text{bar}[\mu_{0 \rightarrow T|x}])$$

$$\stackrel{(\text{c}) \text{ in Proposition 15}}{\leq} \vec{d}(\text{bar}[\mu_{t \rightarrow t+T|x}], \text{bar}[a_t \mu + b_t \delta_{A^{(T)}(x)^{-1} * I}])$$

$$+ \vec{d}(\text{bar}[a_t \mu + b_t \delta_{A^{(T)}(x)^{-1} * I}], \text{bar}[a_t \mu + b_t \delta_I])$$

$$+ \vec{d}(\text{bar}[a_t \mu(t) + b_t \delta_I], \text{bar}[\mu_{0 \rightarrow T|x}]).$$

The addends to the right of  $\leq$  are sequentially denoted by  $A_1, A_2, A_3$ . Thanks to (a) in Proposition 15 and (A.12), we have

$$\begin{aligned} \|A_3\|_2 &\leq W_1(a_t \mu + b_t \delta_I, a_t \mu + b_t \mu_{0 \rightarrow t|x}) \\ &\stackrel{(\text{A.11}), (\text{C.2})}{=} b_t \sup_{\|\psi\|_{\text{Lip}} \leq 1} \left( \int_{S^+} \psi d\delta_I - \frac{1}{t} \int_0^t \psi(A^{(s)}(x)^{-1} * I) ds \right) \\ &= \frac{1}{T} \sup_{\|\psi\|_{\text{Lip}} \leq 1} \int_0^t [\psi(I) - \psi(A^{(s)}(x)^{-1} * I)] ds \\ &\leq \frac{1}{T} \int_0^t d[I, A^{(s)}(x)^{-1} * I] ds \stackrel{(\text{C.6})}{\leq} t\varepsilon/2. \end{aligned}$$

By arguing likewise, we establish that  $\|A_1\|_2 \leq t\varepsilon/2$ . Bringing the pieces together and invoking (15) results in

$$t^{-1} \vec{d}\{Q[\varphi^t(x)], A^{(t)}(x) * Q(x)\} \leq \varepsilon \mathbb{1} + t^{-1} A_2. \quad (\text{C.7})$$

By (i) in Lemma 19,  $\mu$  is the  $W_1$ -limit of a sequence whose terms are measures of the form  $\mu' = \frac{1}{s}(\delta_{p_1} + \dots + \delta_{p_s})$ , where  $s \in \mathbb{N}$  and  $p_1, \dots, p_s \in S^+$  are individual for any term. Relation (A.8) with  $m := s + 1$ ,  $\omega := (a_t/s, \dots, a_t/s, b_t) \in \Delta_m$  yields

$$\begin{aligned} \vec{d}(\text{bar}[a_t \mu' + b_t \delta_{A^{(T)}(x)^{-1} * I}], \text{bar}[a_t \mu' + b_t \delta_I]) \\ = \vec{d}(\text{bar}[\omega; p_1, \dots, p_m, A^{(T)}(x)^{-1} * I], \text{bar}[\omega; p_1, \dots, p_m, I]) \\ \leq t/T \cdot \vec{d}(A^{(T)}(x)^{-1} * I, I). \end{aligned}$$

Letting  $\mu' \rightarrow \mu$  results in

$$\begin{aligned} t^{-1} A_2 &\leq \vec{d}(A^{(T)}(x)^{-1} * I, I)/T \\ &\stackrel{(\text{C.4})}{=} \vec{d}[A^{(T)}(x)^{-1} [A^{(T)}(x)^{-1}]^\top, I]/T \stackrel{(\text{A.4})}{=} \\ &2\vec{\sigma}\{[A^{(T)}(x)^\top A^{(T)}(x)]^{1/2}\}/T \stackrel{(\text{a})}{=} 2\vec{\sigma}[A^{(T)}(x)]/T, \end{aligned} \quad (\text{C.8})$$

where (a) holds since the involved matrices have the same singular values. The proof is completed by injecting (C.8) into (C.7). ■

**Proof of Proposition 27.** The remaining step of this proof largely comes to computing the limit of the l.h.s. in (C.5) as  $t \rightarrow 0+$ . In this l.h.s., the orbital derivative  $\dot{Q}(x) := \frac{d}{dt} Q(\varphi^t(x))|_{t=0+}$  exists due to Bochi and Navas (2015, Lem. 4.5). As for the other argument of  $\vec{d}[\cdot, \cdot]$ , we have

$$\begin{aligned} \frac{d}{dt} A^{(t)}(x) * Q(x)|_{t=0+} &= \frac{d}{dt} \{ \mathbf{D}\varphi^t(x) Q(x) \mathbf{D}\varphi^t(x)^\top \}|_{t=0+} \\ &\stackrel{\frac{d}{dt} \mathbf{D}\varphi^t(x)|_{t=0+} = \mathbf{D}\varphi(x) \text{ by Hartman (1982, Thm. 3.1, Ch. V)}}{=} \mathbf{D}\varphi(x) Q(x) \\ &+ Q(x) \mathbf{D}\varphi(x)^\top. \end{aligned}$$

Formula (A.4) is our incentive to use Lemma 24 and study

$$\begin{aligned} \frac{d}{dt} \gamma\{Q[\varphi^t(x)], A^{(t)}(x) * Q(x)\}|_{t=0+} \\ = \mathbf{D}\gamma[Q(x), Q(x)] \left( \frac{\dot{Q}(x)}{\mathbf{D}\varphi(x)Q(x) + Q(x)\mathbf{D}\varphi(x)^\top} \right) \stackrel{\text{Lemma 24}}{=} Q(x)^{-1/2} h, \end{aligned}$$

where  $h \in S$  is the unique solution of the Lyapunov equation

$$hQ(x)^{1/2} + Q(x)^{1/2}h = \mathbf{D}\varphi(x)Q(x) + Q(x)\mathbf{D}\varphi(x)^\top - \dot{Q}(x).$$

In the l.h.s. of (C.5), we replace  $\vec{d}$  by  $2\vec{\sigma}$  based on (A.4). Then, by Corollary 23, the l.h.s. converges to the following limit as  $t \rightarrow 0+$ :

$$\frac{1}{\ln(2)} \left\{ \lambda_1 [Q(x)^{-1/2} h + h Q(x)^{-1/2}], \dots, \lambda_n [Q(x)^{-1/2} h + h Q(x)^{-1/2}] \right\}.$$

The function  $P(x) := Q(x)^{-1} \in \mathcal{S}^+$  is continuous by Lemma 28; the foregoing and van den Bos (2007, Thm. D2) imply that its orbital derivative  $\dot{P}(x)$  exists and  $\dot{Q}(x) = -P(x)^{-1} \dot{P}(x) P(x)^{-1}$ . Multiplying the Lyapunov equation by  $Q(x)^{-1/2}$  on the left and right, we get

$$\begin{aligned} Q(x)^{-1/2} h + h Q(x)^{-1/2} &= \\ P(x)^{1/2} \left\{ \mathbf{D}\varphi(x) P(x)^{-1} + P(x)^{-1} \mathbf{D}\varphi(x)^\top \right. \\ &\quad \left. + P(x)^{-1} \dot{P}(x) P(x)^{-1} \right\} P(x)^{1/2} \\ &= P(x)^{-1/2} \left\{ P(x) \mathbf{D}\varphi(x) + \mathbf{D}\varphi(x)^\top P(x) + \dot{P}(x) \right\} P(x)^{-1/2}. \end{aligned}$$

Due to Remark 10, this means that the l.h.s. of (C.5) converges to twice the l.h.s. of (C.1) as  $t \rightarrow 0+$ . By letting  $t \rightarrow 0+$  and then  $\varepsilon \rightarrow 0+$  in (C.5), we arrive at (C.1). ■

**Proof of (ii) in Theorem 11.** By (4), (B.6) is still valid with some real  $N$ , now denoted by  $T$ . Hence

$$\begin{aligned} H_{\text{res}}(\varphi, K) &\geq \max_{x \in K} \frac{1}{T} \max_{k=0, \dots, n} \sum_{i=1}^k \log \alpha_i(T, x) - \varepsilon \\ &\stackrel{(A.2), (A.3), (C.1)}{\geq} \frac{1}{2 \ln 2} \max_{x \in K} \max_{k=0, \dots, n} \sum_{i=1}^k \log \varsigma_i^P(x) - \varepsilon \\ &= \frac{1}{2 \ln 2} \max_{x \in K} \sum_{i=1}^n \max\{0, \varsigma_i^P(x)\} - \varepsilon. \quad \blacksquare \end{aligned}$$

## Appendix D. Technical conclusions from the proofs

The above proofs in fact offer explicit formulas for a sequence of Riemannian metrics  $P_N(\cdot)$  that asymptotically provides the exact value of the restoration entropy, i.e., the infimum of the r.h.s. in (9). In the discrete-time case, such a formula is implied by (B.4) and the considerations just after (B.5), and is as follows:

$$P_N(x) = \text{bar}(I, \mathbf{D}\varphi(x)^{-1} * I, \dots, \mathbf{D}\varphi^{N-1}(x)^{-1} * I)^{-1}.$$

In the continuous-time case, an appropriate family of metrics  $P_T(x)$ ,  $T > 0$ , is described by an analogous but technically more sophisticated formula, which involves the barycenter of the continuum of matrices  $\mathbf{D}\varphi^t(x)^{-1} * I$ ,  $t \in [0, T]$ .

Developments of effective numerical procedures for building a minimizing sequence of Riemannian metrics, based on both just sketched and other ideas, is basically a topic of our future research. An initial step towards this end was undertaken in Hafstein and Kawan (2019).

## References

- Andrievsky, B. R., Matveev, A. S., & Fradkov, A. L. (2010). Control and estimation under information constraints: Toward a unified theory of control, computation and communications. *Automation and Remote Control*, 71(4), 572–633.
- Belozyorov, V. E. (2015). Exponential-algebraic maps and chaos in 3d autonomous quadratic systems. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 25(4), Article 1550048.
- Bhatia, R. (2007). *Positive definite matrices*. Princeton and Oxford: Princeton University Press.

- Bochi, J. (2018). Ergodic optimization of Birkhoff averages and Lyapunov exponents. In *Proceedings of the international congress of mathematicians*. Rio de Janeiro.
- Bochi, J., & Navas, A. (2015). A geometric path from zero Lyapunov exponents to rotation cycles. *Ergodic Theory & Dynamical Systems*, 35, 374–402.
- Boichenko, V. A., Leonov, G. A., & Reitman, V. (2005). *Dimension theory for ordinary differential equations*. Wiesbaden, Germany: Teubner Verlag.
- Chavel, I. (2006). *Riemannian geometry: A modern introduction*. NY: Cambridge University Press.
- Colonius, F., & Kawan, C. (2009). Invariance entropy for control systems. *SIAM Journal on Control and Optimization*, 48(3), 1701–1721.
- Colonius, F., Kawan, C., & Nair, G. (2013). A note on topological feedback entropy and invariance entropy. *Systems & Control Letters*, 62(5), 377–381.
- Del Moral, P., & Niclas, A. (2018). A Taylor expansion of the square root matrix functional. *Journal of Mathematical Analysis and Applications*, 465(1), 259–266.
- Downarowicz, T. (2011). *Entropy in dynamical systems*. NY: Cambridge University Press.
- Fradkov, A. L., Andrievsky, B., & Ananyevskiy, M. S. (2015). Passification based synchronization of nonlinear systems under communication constraints and bounded disturbances. *Automatica*, 55, 287–293.
- Hafstein, S., & Kawan, C. (2019). Numerical approximation of the data-rate limit for state estimation under communication constraints. *Journal of Mathematical Analysis and Applications*, 473(2), 1280–1304.
- Hagihara, R., & Nair, G. N. (2013). Two extensions of topological feedback entropy. *Mathematics of Control, Signals and Systems*, 25(4), 473–490.
- Hartman, P. (1982). *Ordinary differential equations* (2nd ed.). Boston: Birkhäuser.
- Heemels, W. P. M. H., Teel, A. R., Van De Wouw, N., & Nešić, D. (2010). Networked control systems with communication constraints: Tradeoffs between transmission intervals, delays and performance. *IEEE Transactions on Automatic Control*, 55(8), 1781–1796.
- Hespanha, J. P., Naghshtabrizi, P., & Yonggang, X. (2007). A survey of recent results in networked control systems. *Proceedings of IEEE*, 95(1), 138–162.
- Hiriart-Urruty, J.-B., & Ye, D. (1995). Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numerische Mathematik*, 70, 45–72.
- Holbrook, J. (2012). No dice: a deterministic approach to the Cartan centroid. *Journal of the Ramanujan Mathematical Society*, 27, 509–521.
- Horn, R., & Johnson, C. (2013). *Matrix analysis*. Cambridge, UK: Cambridge University Press.
- Kawan, C. (2019). On the relation between topological entropy and restoration entropy. *Entropy*, 21(1), 7.
- Kawan, C., Matveev, A., & Pogromsky, A. (2020). Data rate limits for the remote state estimation problem. In *Proceedings of the 2020 IFAC world congress*. Berlin.
- Kawan, C., & Yüksel, S. (2017). Metric and topological entropy bounds on state estimation for stochastic non-linear systems. In *2017 IEEE international symposium on information theory (ISIT)* (pp. 1455–1459). IEEE.
- Kawan, C., & Yüksel, S. (2018). On optimal coding of non-linear dynamical systems. *IEEE Transaction on Information Theory*, 64(10), 6816–6829.
- Khalil, H. K. (2002). *Nonlinear systems* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Koiran, P. (2001). The topological entropy of iterated piecewise affine maps is uncomputable. *Discrete Mathematics & Theoretical Computer Science*, 4(2), 351–356.
- Liberzon, D., & Mitra, S. (2018). Entropy and minimal bit rates for state estimation and model detection. *IEEE Transactions on Automatic Control*, 63(10), 3330–3344.
- Lim, Y., & Palfia, M. (2014). Weighted deterministic walks for the least squares mean on Hadamard spaces. *Bulletin of the London Mathematical Society*, 46, 561–570.
- Matveev, A. S., & Pogromsky, A. Y. (2016). Observation of nonlinear systems via finite capacity channels: constructive data rate limits. *Automatica*, 70, 217–229.
- Matveev, A. S., & Pogromsky, A. Y. (2019). Observation of nonlinear systems via finite capacity channels, part II: Restoration entropy and its estimates. *Automatica*, 103, 189–199.
- Matveev, A. S., & Savkin, A. V. (2009). *Estimation and control over communication networks*. Boston: Birkhäuser.
- Morris, I. (2013). Mather sets for sequences of matrices and applications to the study of joint spectral radii. *Proceedings of the London Mathematical Society*, 107(1), 121–150.
- Nair, G. N., Fagnani, F., Zampieri, S., & Evans, R. J. (2007). Feedback control under data rate constraints: an overview. *Proceedings of the IEEE*, 95(1), 108–137.
- Pogromsky, A. Y., & Matveev, A. S. (2011). Estimation of topological entropy via the direct Lyapunov method. *Nonlinearity*, 24, 1937–1959.
- Pogromsky, A., & Matveev, A. (2016a). Data rate limitations for observability of nonlinear systems. *IFAC-PapersOnLine*, 49(14), 119–124.
- Pogromsky, A. Y., & Matveev, A. S. (2016b). Stability analysis via averaging functions. *IEEE Transactions on Automatic Control*, 61(4), 1081–1086.
- Rungger, M., & Zamani, M. (2017). Invariance feedback entropy of nondeterministic control systems. In *Proceedings of the 20th international conference on hybrid systems: computation and control* (pp. 91–100).

- Savkin, A. V. (2006). Analysis and synthesis of networked control systems: topological entropy, observability, robustness, and optimal control. *Automatica*, 42(1), 51–62.
- Schecter, E. (1997). *Handbook of analysis and its foundations*. Academic Press.
- Sturm, K.-Th. (2003). Probability measures on metric spaces of nonpositive curvature. In P. Auscher, T. Coulhon, & A. Grigor'yan (Eds.), *Heat kernels and analysis on manifolds, graphs, and metric spaces (contemporary mathematics, 338)* (pp. 357–390). Providence, RI: American Mathematical Society.
- van den Bos, A. (2007). *Parameter estimation for scientists and engineers* (3rd ed.). Upper Saddle River, NJ: John Wiley & Sons.
- Villani, C. (2009). *Optimal transport: old and new*. Berlin: Springer.
- Wong, W. S., & Brockett, R. G. (1997). Systems with finite communication bandwidth constraints – part i: State estimation problems. *IEEE Transactions on Automatic Control*, 42(9), 1294–1299.
- Yüksel, S., & Başar, T. (2013). *Stochastic networked control systems: Stabilization and optimization under information constraints*. Boston: Birkhäuser.



**Christoph Kawan** received the diploma and doctoral degree in mathematics (both under supervision of Prof. Fritz Colonius) from the University of Augsburg, Germany, in 2006 and 2009, respectively. He spent four months as a research scholar at the State University of Campinas, Brazil, in 2011, and nine months at the Courant Institute of Mathematical Sciences at New York University, in 2014. He is the author of the book 'Invariance Entropy for Deterministic Control Systems - An Introduction' (Lecture Notes in Mathematics 2089. Berlin: Springer, 2013). He currently works as a researcher at the Ludwig-Maximilians-Universität in Munich. His research interests include networked and information-based control and large-scale networks.



**Alexey S. Matveev** was born in Leningrad, USSR, in 1954. He received the M.S. and Ph.D. degrees in 1976 and 1980, respectively, both from the Leningrad University. Currently, he is a professor of the Department of Mathematics and Mechanics, Saint Petersburg University. His research interests include estimation and control over communication networks, hybrid dynamical systems, and navigation and control of mobile robots.



**Alexander Yu. Pogromsky** was born in Saint Petersburg, Russia, on March 1, 1970. He received the M.Sc degree (cum laude) from the Baltic State Technical University, Russia in 1991 and Ph.D (Candidate of Science) degree from the St.Petersburg Electrotechnical University in 1994. From 1995 till 1997 he was with the Laboratory "Control of Complex Systems" (Institute for Problems of Mechanical Engineering, St. Petersburg, Russia). From 1997 to 1998 he was a research fellow with the Department of Electrical Engineering, division on Automatic Control, Linköping University, Sweden.

In the period of 1999–2001 he was a research fellow with the department of Electrical Engineering of Eindhoven University and after that he joined the department of Mechanical Engineering of Eindhoven University of Technology, the Netherlands where he works till now. He occupies visiting professor positions in St. Petersburg State University and ITMO University. His research interests include theory of nonlinear, adaptive and robust control, nonlinear oscillations.