Optimal State Estimation Over Gaussian Channels with Noiseless Feedback

Dapeng Li and Naira Hovakimyan

Abstract

This paper addresses the optimal state estimation problem when the estimator and the dynamics are connected only by an additive white Gaussian channel with input power constraint. We present a new communication and estimation strategy based on Kalman-Bucy filtering and water filling optimization algorithms. The optimality is established with respect to the minimal mean-square estimation error with an exponentially decaying rate. As an example, it illustrates the use of analogue amplitude modulation for an unstable system.

I. Introduction

During the last decade much attention has been drawn to control problems in the presence of communication channel constraints. This class of problems has been investigated in different settings, including finite alphabets channels (quantization) [1], [2], erasure channels [3] [4] and additive Gaussian channels [5], [6]. Among all the channel modelings, *channel capacity* plays a key role in characterizing the fundamental limitations of control design imposed by limited communication. The relationship between the channel capacity and the plant dynamics are revealed in all above channels.

Gaussian channel and its variants have been one of the central topics in information and communication theory for their capability of capturing several important aspects of real-life communication systems. To consider the relationship between control and communication, Gaussian channel is also a popular choice. Ref. [5] has captured the relation between the state (output) feedback stabilization of a linear time-invariant (LTI) system and the signal-to-noise ratio (SNR) constraint of the channel for both continuous-time and discrete-time cases; [7] and [8] have considered the linear quadratic Gaussian framework to derive the data-rate bound and provide a fairly complete scheme for design of the encoder, the controller and the decoder. In [9], Gaussianity plays an important role in obtaining the Bode's integrals in terms of log integral of relevant power spectral densities in the closed loop.

The state estimation under communication limitations has been investigated for its close relationship with controls as well as its own importance. Refs. [10]and [11] tried to fit the problem into the framework developed in [9] and [12] with the hope to use the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control theory in this context. In a more general setting, feedback has long been used to improve the performance of the communication systems in terms of better convergence rate of

This material is based upon work supported by the United States Air Force under Contract No. FA9550-08-1-0135.

Dapeng Li and Naira Hovakimyan are with Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign (UIUC), Urbana, IL 61801, USA, {li63,nhovakim}@illinois.edu

the error probability. In the discrete-time setting in case of additive white gaussian noise (AWGN) channel, inspired by Robbins-Monro stochastic iterative root seeking algorithm from [13] K-S feedback coding is presented [14]. A large number of results followed this seminal work along with various of extensions. Recently, this classical result has caught much attention from control community, starting from [15], which linked the optimal estimation with optimal encoding/decoding, with a fundamental observation unifying control, estimation and communication (see also [16]). Another similar development from the information theory perspective is reported in [17], where colored gaussian channel with the capacity of coding is discussed in a fairly general setting. The continuous-time version of K-S scheme is presented in [18], where the derivation heavily relies on the stochastic calculus and optimal filtering theory.

The objective of this paper is to solve the continuous-time optimal estimation problem in the presence of an AWGN channel with an input power constraint. The contribution of the paper is three-fold:

- It establishes a framework to analyze some important quantities in a stable closed loop, such as minimal mean-square error (MMSE) and channel capacity (or signal to noise ratio), where stationarity is not assumed;
- Based on this framework, we not only recover the existing relation between channel capacity and the open-loop instability in stable closed loops, but also provide a tighter bound to guarantee an exponentially decaying mean square of estimation error.
- The detailed procedure and algorithms are provided for the transmitter and estimator design, together with the rigorous proof of optimality.

The paper is organized as follows. In Section II, we introduce the models for both the channel and the plant, and the design problem statement. Section III discusses a scalar version of the problem, which leads to the development of the solution in Section IV. A numerical example is analyzed in Section V. We conclude the paper with different problems for future research directions in Section VI.

II. Problem Formulation

In this section we state the problem formulation. The scheme is depicted in Fig. 1 where the transmitter has the access to the time-history of the channel output via a noiseless feedback.

- The plant of interest is given by the following n dimensional linear SDE

$$\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x}(t)\,, \mathbf{x}(0) = \mathbf{x}_0\,. \tag{II.1}$$

  where $A \in \mathbb{R}^{n\times n}$. To ensure the solution $\mathbf{x}(t)$ of (II.1) is Gaussian, the initial value $\mathbf{x}_0$ is also assumed to be Gaussian. Also, $\mathbf{E}\mathbf{x}_0\mathbf{x}_0^\top$ is not singular.
- The communication part of the closed loop is modeled as an additive white Gaussian channel

$$d\mathbf{v}(t) = \mathbf{z}(t)dt + \sigma d\mathbf{W}(t)\,, \tag{II.2}$$

  where $\mathbf{z}(t)$ is the channel input generated by the signal $\mathbf{x}_0^t$, $\mathbf{W}(t)$ is a standard Wiener process and $\mathbf{v}(t)$ is

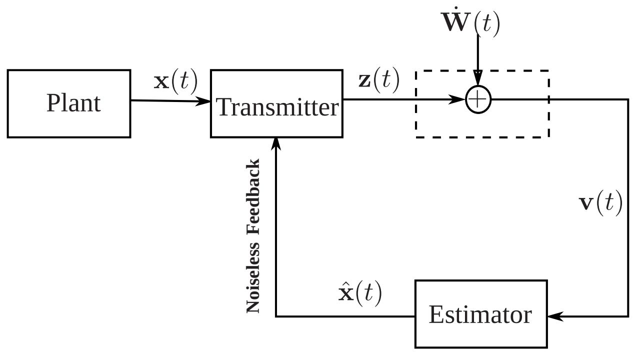

Fig. 1. State Estimation via Noiseless Feedback

the channel output. An average power constraint is imposed on the channel input:

$$\limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{E}\mathbf{v}^2(t)dt \le \mathscr{P},$$

for some $\mathscr{P} > 0$. Slightly different from most of the communication theory literature, the power constraint here is defined over an infinite time horizon to get aligned with some notions in control theory such as asymptotic stability. We also define the noise-to-signal ratio of the channel as

$$\mathsf{SNR} \triangleq \frac{\mathscr{P}}{\sigma^2}.$$

It is well-known that the channel capacity is $\mathcal{C} = \mathsf{SNR}/2$ [19].

- The transmitter (encoder) is a causal map defined as $\mathbf{z}(t) \triangleq f(t, \mathbf{x}_0, \mathbf{v}_0^t)$. The receiver(decoder)/estimator is also a causal map $\hat{\mathbf{x}}(t) \triangleq g(t, \mathbf{v}_0^t)$, where $\hat{\mathbf{x}}(t)$ is the estimation of the state $\mathbf{x}(t)$. The error signal is defined as $\tilde{\mathbf{x}}(t) \triangleq \mathbf{x}(t) - \hat{\mathbf{x}}(t)$.
- As a standard assumption, all the random variables (processes) in this system are defined in a filtered probability space $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathbb{P})$.

Definition 2.1: The unique solution $X(t)$ of a stochastic differential equation is said to be mean-square exponentially stable with convergence rate $\nu < 0$ if

$$\limsup_{t\to\infty} \frac{1}{t}\log \mathbf{E}\|X(t)\|^2 \le \nu$$

The objective of joint estimation/communication design is to identify *a transmitter and receiver/estimator combination such that the error dynamics with state* $\tilde{\mathbf{x}}(t)$ *is mean-square exponentially stable with minimal decaying rate.*

III. Estimation, Communication and Control over Gaussian Channel: A Scalar case study

In this section we review a scalar estimation problem with communication constraint, which was originated by [19] and [18]. Some modifications and innovative observations are made to shed a light on the main result to be presented in the next section.

A. Transmitting a Gaussian Random Variable

We consider the simplest case, where an analog scalar Gaussian variable $\mathbf{m}$ is to be transmitted through a continuous-time AWNG channel. We further assume that the transmitter (encoder) takes the following affine structure for easy computation and Guassianity of f, given by

$$f(t, \mathbf{m}, \mathbf{v}_0^t) \triangleq \phi(t, \mathbf{v}_0^t) + \psi(t, \mathbf{v}_0^t)\mathbf{m}\,. \tag{III.1}$$

For this given structure of information transmission scheme, the minimal mean-square error for each time instance t is achieved by choosing the estimation $\hat{\mathbf{m}}(t) = \mathbf{E}[\mathbf{m}|\mathbf{v}_0^t]$, which is not readily calculable in general case. So one needs to show a way to construct the corresponding receiver/estimator, which yields $\hat{\mathbf{m}}(t)$. Upon that, constrained by the channel input power level $\mathscr{P}$, parameter optimization for f and g needs to be conducted to reach minimal mean square error. In other words, the problem of optimal estimation is solved in two steps:

1) For the given transmitter (III.1), obtain the estimation scheme g with output $\hat{\mathbf{m}}(t)$;
2) Solve the optimization problem $\min_{g,f} \mathbf{E}(\tilde{\mathbf{m}}^2(t))$ subject to power constraint $\mathscr{P}$.

The first step is straightforwardly obtained by the conditional Kalman-Bucy filter.

Lemma 3.1: Consider the linear transmission strategy in (III.1). Then

$$\begin{aligned} d\hat{\mathbf{m}}(t) &= \frac{1}{\sigma^2} P(t)\psi(t, \mathbf{v}_0^t)[d\mathbf{v}_t - \phi(t, \mathbf{v}_0^t)dt - \psi(t, \mathbf{v}_0^t)\hat{\mathbf{m}}(t)dt] \\ \frac{dP(t)}{dt} &= -\frac{1}{\sigma^2} P^2(t)\psi^2(t, \mathbf{v}_0^t)\,, \end{aligned} \tag{III.2}$$

where $P(t) \triangleq \mathbf{E}[(\tilde{\mathbf{m}}(t))^2|\mathbf{v}_0^t]$, $P(0) = \mathbf{E}(\tilde{m}(0))^2$ and $\hat{\mathbf{m}}(0) = \mathbf{E}\mathbf{m}$.

Proof: The proof is just an application of Kalman-Bucy filter for the dynamic system with $\mathbf{m}(t)$ as the system state and $\mathbf{v}(t)$ as the noise corrupted observation.

$$\begin{aligned} d\mathbf{m}(t) &= 0 \\ d\mathbf{v}(t) &= [\phi(t, \mathbf{v}_0^t) + \psi(t, \mathbf{v}_0^t)\mathbf{m}]dt + \sigma d\mathbf{W}(t)\,. \end{aligned}$$

■

The second step is solved by the following lemma.

Lemma 3.2: Within the class of linear transmission strategies, which satisfy the condition of (3.6) and the power constraint, optimal transmission strategy ϕ^* and ψ^* are given by

$$\phi^*(t, \mathbf{v}_0^t) = -\sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp\left(\frac{\mathsf{SNR}}{2}t\right)\hat{\mathbf{m}}(t) \qquad \psi^*(t, \mathbf{v}_0^t) = \sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp\left(\frac{\mathsf{SNR}}{2}t\right).$$

The optimal mean square error for this strategy is

$$\mathbf{E}\tilde{\mathbf{m}}^2(t) = P(0)\exp\left(-\mathsf{SNR}t\right)$$

The proof of the lemma can be found in [18].

Remark 3.3: Not surprisingly, this feedback coding strategy design can be regarded as feedback stabilization problem, where the state to be stabilized, in the mean-square sense, is defined as $\tilde{\mathbf{m}}(t)$. The stabilization problems can be solved conveniently by using Lyapunov's indirect method. More specifically, one can employ the Lyapunov argument developed in stochastic setting by choosing the candidate Lyapunov function as $V(\tilde{\mathbf{m}}(t)) = \frac{1}{2}\tilde{\mathbf{m}}^2(t)$, and ensure its negative derivative by designing proper transmission schemes. The details of this approach are not discussed here.

Remark 3.4: It is also shown in [18] that the solution $\phi^*(t, \mathbf{v}_0^t) + \psi^*(t, \mathbf{v}_0^t)\mathbf{m}$ is optimal among nonlinear functionals of $\mathbf{m}$ (i.e. $f(t, \mathbf{m}, \mathbf{v}_0^t)$).

Remark 3.5: This feedback communication scheme can be regarded as an continuous-time extension of the K-S method.

B. Transmission of a signal

Next we go one step further by replacing the constant source $\mathbf{m}$ by a dynamic one $\mathbf{x}(t)$, evolving according to the linear scalar differential equation with parameter $\lambda \in \mathbb{R}$ and a Gaussian initial value $\mathbf{x}_0$

$$\frac{d\mathbf{x}(t)}{dt} = \lambda \mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}_0. \tag{III.3}$$

Following the same idea in (III.2), we can consider the Kalman-Bucy filter for the dynamics

$$d\mathbf{x}(t) = \lambda \mathbf{x}(t) dt,$$

$$d\mathbf{v}(t) = [\phi(t, \mathbf{v}_0^t) + \psi(t, \mathbf{v}_0^t)\mathbf{x}(t)]dt + \sigma d\mathbf{W}(t).$$

Next, we proceed with the two-step strategy. The following lemma provides a structure of decoder/estimator, which yields the optimal estimation $\hat{\mathbf{x}}(t) = \mathbf{E}[\mathbf{x}(t)|\mathbf{v}_0^t]$.

Lemma 3.6: Consider the linear transmission strategy in (III.1) (where $\mathbf{m}$ is replaced by $\mathbf{x}$) and the source (III.3). Then the optimal estimation of $\mathbf{x}(t)$ is given as

$$\begin{aligned} d\hat{\mathbf{x}}(t) &= \lambda \hat{\mathbf{x}}(t) + \frac{1}{\sigma^2} P(t)\psi(t, \mathbf{v}_0^t)[d\mathbf{v}_t - \phi(t, \mathbf{v}_0^t)dt - \psi(t, \mathbf{v}_0^t)\hat{\mathbf{x}}(t)dt] \\ \frac{dP(t)}{dt} &= 2\lambda P(t) - \frac{1}{\sigma^2} P^2(t)\psi^2(t, \mathbf{v}_0^t)\,, \end{aligned} \tag{III.4}$$

where $P(t) \triangleq \mathbf{E}[\tilde{\mathbf{x}}^2|\mathbf{v}_0^t]$, $P(0) = \mathbf{E}\mathbf{x}_0^2$ and $\hat{\mathbf{x}}(0) = \mathbf{E}\mathbf{x}_0$.

Next we proceed to the step two. Towards this end, the differential equation with equality of $P(t)$ in (III.4) is rewritten as

$$\dot{P}(t) = \left(\lambda - \frac{1}{\sigma^2} P(t)\psi^2(t, \mathbf{v}_0^t)\right) P(t)\,,$$

and solved by

$$P(t) = P(0) \exp\left(\int_0^t \left(2\lambda - \frac{1}{\sigma^2} P(\tau)\psi^2(\tau, \mathbf{v}_0^\tau)\right) d\tau\right)\,.$$

Taking the expectation and using Jensen's inequality, we have

$$E\tilde{\mathbf{x}}^2(t) = \mathbf{E}P(t) = P(0) \exp\left(\int_0^t \left(2\lambda - \frac{1}{\sigma^2} \mathbf{E}P(\tau)\psi^2(\tau, \mathbf{v}_0^\tau)\right) d\tau\right)\,,$$

where Fubini's theorem is also used to interchange integration and expectation. The Lyapunov exponent can be calculated as

$$\limsup_{T\to\infty}\frac{1}{T}\log \mathbf{E}P(T) \geq 2\lambda - \frac{1}{\sigma^2}\liminf_{T\to\infty}\frac{1}{T}\int_0^t \mathbf{E}P(t)\psi^2(t,\mathbf{v}_0^t,t)dt$$
$$\geq 2\lambda - \frac{1}{\sigma^2}\limsup_{T\to\infty}\frac{1}{T}\int_0^t \mathbf{E}P(t)\psi^2(t,\mathbf{v}_0^t,t)dt\,. \tag{III.5}$$

It is clear that the minimization of $P(t)$ is reduced to the choice of ψ that minimizes $\frac{1}{\sigma^2}\limsup_{T\to\infty}\frac{1}{T}\int_0^t \mathbf{E}P(t)\psi^2(t,\mathbf{v}_0^t,t)dt$. Towards this end, we have

$$\begin{aligned}\mathscr{P} &\geq \limsup_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{E}[\phi(t,\mathbf{v}_0^t)+\psi(t,\mathbf{v}_0^t)\mathbf{x}(t)]^2\\ &= \limsup_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{E}[\phi(t,\mathbf{v}_0^t)+\psi(t,\mathbf{v}_0^t)\hat{\mathbf{x}}(t)]^2 + \limsup_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{E}\psi^2(t,\mathbf{v}_0^t)P(t)dt\\ &\geq \limsup_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{E}\psi^2(t,\mathbf{v}_0^t)P(t)dt\,.\end{aligned}$$

A lower bound of the Lyapunov exponent of $\mathbf{E}P(t)$ is given as

$$\limsup_{T\to\infty}\frac{1}{T}\log \mathbf{E}P(T) \geq 2\lambda - \frac{\mathscr{P}}{\sigma^2} = 2\lambda - \mathsf{SNR}\,. \tag{III.6}$$

The above lower bound can be achieved on

$$\psi^2(t,\mathbf{v}_0^t)P(t) = \mathscr{P} \text{ and } \phi(t,\mathbf{v}_0^t)+b(t,\mathbf{v}_0^t)\hat{\mathbf{x}}(t) = 0\,,\forall t\geq 0\,,$$

which in turn gives the optimal solution of

$$\psi^*(t,\mathbf{v}_0^t) = \sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp\left(\frac{\mathsf{SNR}-2\lambda}{2}t\right) \text{ and } \phi^*(t,\mathbf{v}_0^t) = -\sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp\left(\frac{\mathsf{SNR}-2\lambda}{2}t\right)\hat{\mathbf{x}}(t)\,.$$

Remark 3.7: Eqn. (III.6) shows that for the variance of $\tilde{\mathbf{x}}(t)$ to be exponentially decaying, one needs $\lambda < \frac{\mathsf{SNR}}{2} = \mathcal{C}$. In other words, converging estimation is achievable provided that the degree of instability of the source is less than the channel capacity. This observation can be roughly explained by Shannon's source-channel separation principle [20]. The unstable process produces extra information at the steady rate $\lambda(\geq 0)$, which needs to me transmitted in a timely manner for the vanishing of the mean square error (or rate distortion function). Therefore adequate channel capacity needs to be allocated. For an alternative in-depth treatment of unstable sources, by resorting to the concept of *any time capacity*, one is referred to [3].

C. *Estimation Without Feedback*

As a special case, the non-feedback communication scheme can be considered by proceeding to a similar argument as in the case when feedback is available. In fact, without the knowledge of $\mathbf{v}_0^t$, the optimal estimation of $\mathbf{x}(t)$, utilized on the transmitter's sid,e reduces to its expectation: $\mathbf{E}\mathbf{x}(t) = \exp(\lambda t)\mathbf{E}\mathbf{x}_0$ and $\phi(t,\mathbf{u}_0^t)$ becomes $\phi(t)$, which is a non-random function. Consequently the output of the estimator verifies the following dynamics:

$$\frac{dP(t)}{dt} = 2\lambda P(t) - \frac{1}{\sigma^2}P^2(t)b(t)\,,$$

which is solved by

$$P(t) = \frac{\exp(2\lambda t)}{P^{-1}(0) + \frac{1}{\sigma^2}\int_0^t \psi^2(\tau, \mathbf{v}_0^\tau)\exp(2\lambda\tau)d\tau}\,.$$

Similar to the previous case, we have the optimal solution

$$\phi^*(t) = -\sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp(-\lambda t)\mathbf{E}\mathbf{x}_0 \text{ and } \psi^*(t) = \sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp(-\lambda t)$$

Remark 3.8: The following discussion further reveals the dependency of the optimal performance on the nature of the source dynamics:

- *Stable source ($\lambda < 0$)*: $P^*(t)$ is exponentially decaying at the rate $|\lambda|$, which is given by the inequality

$$P^*(t) \leq P(0)\exp(-|\lambda|t)$$

- *Neutrally stable source ($\lambda = 0$)*: $P(t)$ presents a much slower decay rate given by

$$P^*(t) = \frac{P(0)}{1 + \mathsf{SNR}t}\,.$$

  The behavior of $P(t)$ in above equation is similar to the one that has been achieved by traditional sphere-packing coding strategy in discrete-time setting, with code word length n replaced by the time t.
- *Unstable source ($\lambda > 0$)*: $P(t)$ diverges with arbitrary instability rate, since

$$P^*(t) = \frac{P(0)\exp(|\lambda|t)}{1 + \mathsf{SNR}t}\,.$$

  However, if only the finite horizon problem is considered, one can always find a global minimum.

IV. Main Result: Optimal Estimation Over A Gaussian Channel

With the clear identification of the relation between communication and estimation in the previous section,we are now ready to tackle the main problem. The solution is given by using a water-filling type of argument.

A. Estimation Structure & And a Dual Control Problem

Consider the dynamics shown in Fig. **?????**. Like in the scalar case, we first consider the optimal estimation problem for the dynamics

$$d\mathbf{x}(t) = A\mathbf{x}(t)dt,$$
$$d\mathbf{v}(t) = \phi(t, \mathbf{v}_0^t)dt + \psi^\top(t, \mathbf{v}_0^t)\mathbf{x}(t) + \sigma d\mathbf{W}(t)\,.$$

The transmitter is expressed as $\phi(t, \mathbf{v}_0^t)dt + \psi^\top(t, \mathbf{v}_0^t)\mathbf{x}(t)$. The functions $\phi(t, \mathbf{v}_0^t) \in \mathbb{R}$ $\psi(t, \mathbf{v}_0^t) \in \mathbb{R}^n$ are nonlinear functions to be determined to minimize the Lyapunov index of the error variance, while ensuring the average power of channel input below the constrained level $\mathscr{P}$.

For the given transmitting scheme, the following Kalman-Bucy filter is adopted for the optimal estimation of $\mathbf{x}(t)$,

$$\begin{aligned} d\hat{\mathbf{x}}(t) &= A\hat{\mathbf{x}}(t)dt + \frac{1}{\sigma^2}P(t)\psi(t, \mathbf{v}_0^t)[d\mathbf{v} - \phi(t, \mathbf{v}_0^t)dt - \psi^\top(t, \mathbf{v}_0^t)\hat{\mathbf{x}}(t)dt], \\ \dot{P}(t) &= AP(t) + P(t)A^\top - \frac{1}{\sigma^2}P(t)\psi(t, \mathbf{v}_0^t)\psi^\top(t, \mathbf{v}_0^t)P(t)\,, \end{aligned} \quad \text{(IV.1)}$$

where $P(t) := \mathbf{E}\left[\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^\top(t)|\mathbf{v}_0^t\right]$.

Remark 4.1: One can consider the dual control problem with plant dynamics given by

$$\begin{aligned} \frac{d\mathbf{x}(t)}{dt} &= A\mathbf{x}(t) + B\mathbf{u}(t), \\ d\mathbf{v}(t) &= \psi^\top(t, \mathbf{v}_0^t)\mathbf{x}(t)dt + \sigma d\mathbf{W}(t)\,, \end{aligned} \tag{IV.2}$$

where the second equation models the AWGN channel identical to (II.2). If the control signal $\mathbf{u}(t)$ is designed via the typical LQG method [21], then the separation principle further shows that the variance of the error between the state and its estimated value is identical to $\mathbf{E}P(t)$ in (IV.1). Therefore, to control the plant (IV.2) over the AWGN channel, one can design a proper estimator to cope with the communication constraint, and the control part, which falls into the classical linear quadratic framework, is relatively independent, given the convergence of the estimation. Admittedly, the overall closed loop performance is fundamentally restricted by the communication-constrained estimation, no matter how well the controller is designed. On can further refer to [22] for the same property in general nonlinear systems. This estimation-control separation also explains why our focus is on the estimation part, whose relationship with communication constraint is unveiled in detail subsequently.

B. Solving The Estimation Problem: A water-filling approach

We first introduce a space $\mathcal{B}$, which is a real Hilbert space with internal product defined as

$$\langle \alpha, \gamma \rangle \triangleq \lim_{T\to\infty} \frac{1}{T}\int_0^T \alpha^\top(t)\gamma(t)dt \qquad \alpha(\cdot), \gamma(\cdot) \in \mathcal{B}\,. \tag{IV.3}$$

We say $\beta(\cdot) \in \mathcal{B}$, if $\langle \beta, \beta \rangle$ exits and is less than ∞. If $\beta(\cdot) \in \mathcal{B}$, then the $\lim_{T\to} \frac{1}{T}\int_0^T \beta(t)\beta^\top(t)$ exists.

Next, we define a new quantity $\beta(t) \triangleq \frac{1}{\sigma}P^{1/2}(t)\psi(t, \mathbf{v}_0^t)$, and assume that $\beta(\cdot) \in \mathcal{B}$.

Remark 4.2: Rigorously speaking, rather than a deterministic function of t as its notation suggests, $\beta(t)$ is a stochastic process on the σ-algebra generated by $\mathbf{v}_0^t$. However, we implicitly drop the randomness for three reasons: (1) We can always choose $\psi(t, \mathbf{v}_0^t) = \sigma P^{-1/2}(t)\beta(t)$ to make it non-stochastic; (2) The scalar cases in the previous section suggest that deterministic choices of $\beta(t)$ suffice for the optimality, which is also verified in the later discussion for this vector case; (3) This simplification reduces an otherwise accusive math discussion, while keeps the main point clear. For example, we see obviously that $\mathbf{E}P(t) = P(t)$, which will be useful in the later discussion.

The next lemma links Lyapunov exponent of the the variance of $\tilde{\mathbf{x}}$ with a matrix eigenvalue.

Lemma 4.3: If $P(0)$ is non-singular, the following equality holds:

$$\limsup_{t\to\infty} \frac{1}{t}\log \mathbf{E}\|\tilde{\mathbf{x}}(t)\|^2 = \lambda_{\max}\left(A^\top + A - \lim_{t\to\infty}\frac{1}{t}\int_0^t \beta(\tau)\beta^\top(\tau)d\tau\right)\,. \tag{IV.4}$$

Remark 4.4: If $P(0)$ is singular, then

$$\limsup_{t\to\infty} \frac{1}{t}\log \mathbf{E}\|\tilde{\mathbf{x}}(t)\|^2 \le \lambda_{\max}\left(A^\top + A - \lim_{t\to\infty}\frac{1}{t}\int_0^t \beta(\tau)\beta^\top(\tau)d\tau\right)\,. \tag{IV.5}$$

Note that $\lambda_{\max}$ cannot made arbitrarily small due the power constraint, clearly shown by the following inequality

$$\begin{aligned}\mathscr{P} &\geq \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{E}[\phi(t,\mathbf{v}_0^t)+\psi^\top(t,\mathbf{v}_0^t)\mathbf{x}(t)]^2 dt \\ &\geq \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{E}[\phi(t,\mathbf{v}_0^t)+\psi^\top(t,\mathbf{v}_0^t)\hat{\mathbf{x}}(t)]^2 dt + \mathbf{E}\psi^\top(t,\mathbf{v}_0^t)P(t)\psi(t,\mathbf{v}_0^t)dt \\ &\geq \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{E}\psi^\top(t,\mathbf{v}_0^t)P(t)\psi(t,\mathbf{v}_0^t)dt \\ &= \sigma^2\langle\beta,\beta\rangle\,,\end{aligned} \tag{IV.6}$$

where the second inequality follows from the orthogonality between $\tilde{\mathbf{x}}(t)$ and $\hat{\mathbf{x}}(t)$.

Hence, an optimization problem could be formulated to achieve the lowest Lyapunov exponent by the choice of $\beta(\cdot)$.

$$\begin{aligned}&\inf_{\beta(\cdot)\in\mathcal{B}} \lambda_{\max}\left(A^\top + A - \lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt\right) \\ s.t.\quad &\langle\beta,\beta\rangle \leq \mathsf{SNR} \text{ and } A^\top + A - \lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt \prec 0\,.\end{aligned} \tag{IV.7}$$

Another related optimization problem can be formulated in the same fashion, where the optimal $\beta(\cdot)$ must achieve a minimal channel SNR, subject to closed loop stability:

$$\begin{aligned}&\inf_{\beta(\cdot)\in\mathcal{B}} \langle\beta,\beta\rangle \\ s.t.\quad &A^\top + A - \lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt \prec 0\,.\end{aligned}$$

For both problems, once the optimal decision function $\beta^*(\cdot)$ is obtained, the optimal transmitter and estimator are straightforwardly obtained. Unfortunately, it is very hard, if not impossible to obtain $\beta^*(t)$ by using numerical routines, because these optimization problems are all inherently infinite-dimensional. Here we propose a solution inspired by the water-filling strategy.

Before jumping into the detailed development, an immediate observation can be made regarding the minimal SNR for mean square stability.

Proposition 4.5: If the error dynamics are mean-square exponentially stable, then channel SNR statistics for any causal transmission and decoding/control is given by

$$\frac{\mathsf{SNR}}{2} > \frac{1}{2}\sum_i \lambda_i^+(A+A^\top) \geq \sum_j \Re^+\left(\lambda_j(A)\right) \tag{IV.8}$$

Remark 4.6: Compared with the existing result ($\frac{\mathsf{SNR}}{2} > \sum_j \Re^+\lambda_j(A)$), Proposition 4.5 provides a tighter lower bound for SNR (or capacity) of a channel in a stable closed loop. We use a double integrator to demonstrate the failure of the existing bound to capture the relationship between channel capacity and open loop dynamics. Consider the plant given as $\frac{1}{s^2}$. The two bounds are calculated as

$$\sum_j \Re^+\left(\lambda_j(A)\right) = 0 \text{ and } \frac{1}{2}\sum_i \lambda_i^+(A+A^\top) = 0.5$$

respectively, which implies that a communication channel with capacity at least $.5nats/sec.$ is needed to estimate (and stabilize) a double integrator.

Proof of Proposition 4.5: Note that matrices $A + A^\top$, $\lim_{T\to\infty} \frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt$ and the difference of the two are Hermitian, so all their eigenvalues are real and can be ordered as $\lambda_1 \geq \lambda_2, ..., \geq \lambda_n$ for convenience. Then using Theorem III.4.1 of [23] and noting the fact that $\lim_{T\to\infty} \frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt - (A + A^\top) \succ 0$, we have

$$\begin{aligned} 0 &< \sum_{i=1}^{k} \lambda_i \left(\lim_{T\to\infty} \frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt - (A + A^\top) \right) \\ &\leq \sum_{i=1}^{k} \lambda_i \left(\lim_{T\to\infty} \frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt \right) - \sum_{i=1}^{k} \lambda_i \left(A + A^\top \right) , \end{aligned} \tag{IV.9}$$

for all $k \geq 1$. Particularly, the inequality (IV.9) is also valid for $k = \kappa \triangleq \max_i \left\{ i | \lambda_i(A + A^\top) \geq 0 \right\}$, in which we have

$$\begin{aligned} \sum_{i=1}^{\kappa} \lambda_i \left(A + A^\top \right) &< \sum_{i=1}^{\kappa} \lambda_i \left(\lim_{T\to\infty} \frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt \right) \\ &\leq \sum_{i=1}^{n} \lambda_i \left(\lim_{T\to\infty} \frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt \right) \\ &= \lim_{T\to\infty} \frac{1}{T}\int_0^T \beta^\top(t)\beta(t)dt \leq \mathsf{SNR} . \end{aligned} \tag{IV.10}$$

The first inequality in (IV.8) is straightforward to obtain. The second inequality is a direct application of **what** from the Proposition III.5.3 of (3.22) in [23]. The detailed proof is omitted. □

Now we are ready to construct an optimal information transmission scheme. More specifically, given the channel SNR level, the smallest mean-square convergence rate ν of the state is obtained via the choice of $\beta(\cdot)$. The complete algorithm follows these steps.

1) Basis Construction: Choose a set of orthonormal basis functions $\beta_i(\cdot) \in B, i = 1, 2, ..., n$ such that

$$\langle \beta_i, \beta_j \rangle = \delta_{ij}, \quad i, j = 1, 2, ..., n$$

where δ_{ij} is the Kronecker's delta. Whereas the internal product $\langle \cdot, \cdot \rangle$ has been defined for n dimensional vectors in (IV.3), we slightly abuse the notation here to accommodate the scalar cases as well. There are a number of ways to construct the basis functions, e.g. if $n = 2$, we can simply choose

$$\beta_1(t) = \sqrt{2}\sin(\omega t), \text{ and } \beta_2(t) = \sqrt{2}\cos(\omega t) \qquad \omega > 0.$$

2) Weight Choice by Water-filling: Choose an orthonormal matrix $Q \in \mathbb{R}^{n\times n}$ such that

$$Q^\top (A + A^\top) Q = \text{diag}\{\lambda_1, \lambda_2, ..., \lambda_n\} ,$$

where λ_i is short for $\lambda_i(A + A^\top)$. Then $\beta(\cdot)$ can be parameterized by the basis constructed in 1) with a set of weighting factors $\eta_1, \eta_2, ..., \eta_n \geq 0$ as

$$Q^\top \beta(t) = [\eta_1\beta_1(t), \eta_2\beta_2(t), ..., \eta_n\beta_n(t)]^\top .$$

Based on this fact, the following identity is evident and will be useful later for

$$\langle \beta, \beta \rangle = \langle Q^\top \beta, Q^\top \beta \rangle = \sum_{i=1}^{n} \eta_i^2 .$$

Then the convergence rate minimization problem (IV.7) can be reduced to the following finite dimensional case

$$\min_{\eta_i, \nu} \nu$$
$$s.t. \sum_{i=1}^{n} \eta_i^2 \leq \mathsf{SNR} \text{ and } (\lambda_i - \nu)^+ \leq \eta_i^2 ,$$

where the positivity of η_i^2 brings up $(\lambda_i - \nu)^+ \leq \eta_i^2$. This standard optimization problem can be solved by using the Lagrange multipliers $\xi_i \in \mathbb{R}, i = 1, 2, ..., n$ and $L \in \mathbb{R}$. The objective function is re-written as

$$J \triangleq \nu + \sum_{i=1}^{n} \xi_i((\lambda_i - \nu)^+ - \eta_i^2) + L \left(\sum_{i=1}^{n} \eta_i^2 - \mathsf{SNR} \right) .$$

Differentiating with respect to $\eta_1^2, ..., \eta_n^2$ and ν respectively, we have

$$0 = \frac{\partial J}{\partial \eta_i^2} = -\xi_i + L$$
$$0 = \frac{\partial J}{\partial \nu} = 1 - \sum_{i \in \mathbb{S}} \xi_i \, , \mathbb{S} \triangleq \{ i | (\lambda_i - \nu) \geq 0 \}$$

Solving the set of equations and using Kuhn-Tucker conditions, we have the optimal assignment of the energy

$$\eta^{*2}_i = (\lambda_i - \nu^*)^+ , \quad \sum_{i=1}^{n} \eta^{*2}_i = \mathsf{SNR}$$

The optimal convergence rate ν^* solves

$$\sum_{i=1}^{n} (\lambda_i - \nu^*)^+ = \mathsf{SNR}$$

The solution is depicted graphically in Fig. 2. The vertical levels indicate the eigenvalues of the matrix $A + A^\top$, and the vertical axis is downward pointing. As the input power is increased from zero, we allocate the power to the eigenspace associated with the largest eigenvalue. When more power becomes available, it will be spilled over other eigenspaces to achieve an even ”water level”.

3) Optimal Transmitter and Estimator: Notice that (from last step)

$$\langle \beta^*, \beta^* \rangle = \sum_{i=1}^{n} \eta^{*2}_i = \mathsf{SNR} ,$$

and the equality in (IV.6) holds. Then we have the optimality achieved on

$$\phi^*(t, \mathbf{v}_0^t) + {\psi^*}^\top(t, \mathbf{v}_0^t) \hat{\mathbf{x}}(t) = 0 .$$

Expressed in terms of $\beta^*(t)$, we have the optimal transmitter design:

$$\phi^*(t, \mathbf{v}_0^t) = -{\beta^*}^\top(t) {P^*}^{-\frac{1}{2}}(t) \hat{\mathbf{x}}(t) \quad \psi^*(t, \mathbf{v}_0^t) = {P^*}^{-\frac{1}{2}}(t) \beta^*(t) ,$$

where $P^*(t)$ solves a variation of differential Lyapunov equation given by

$$\dot{P}^*(t) = P^*(t) A + A^\top P^*(t) - {P^*}^{\frac{1}{2}}(t) \beta^*(t) {\beta^*}^\top(t) {P^*}^{\frac{1}{2}}(t) \quad P^*(0) = P(0) . \tag{IV.11}$$

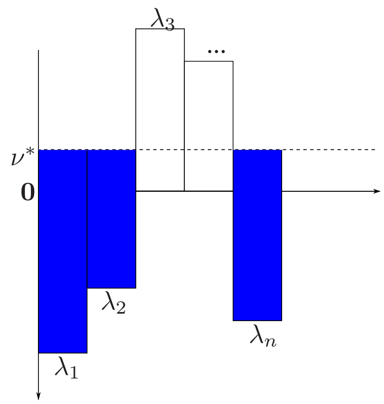

Fig. 2. Water Filling For Optimal Energy Distribution

and the estimator/receiver is given as

$$d\hat{\mathbf{x}}(t) = A\hat{\mathbf{x}}(t)dt + \frac{1}{\sigma^2}P^{*-\frac{1}{2}}(t)\beta^*(t)d\mathbf{v}(t)\,, \hat{\mathbf{x}}(0) = \mathbf{x}_0$$

Remark 4.7: Note that the time profile of $P^*(t)$ (and hence $\psi^*(t, \mathbf{v}_0^t)$) can be determined off-line by integrating (IV.11).

Remark 4.8: One needs $\nu < 0$ to have mean square stability. This requirement also implies that $\sum_{i=1}^n \lambda_i^+(A + A^\top) < \mathsf{SNR}$ is sufficient to ensure mean-square exponentially converging error.

V. Simulation: Estimation via Amplitude Modulation

In this section we demonstrate our approach by using an analog amplitudes modulation (AM) method to transmit the estimation error. The schematic block diagram is shown in Fig. 3, where we do not assume any digitalization (A/D, D/A, quantization etc.) for simplicity. Here the plant is given as

$$\frac{d\mathbf{x}(t)}{dt} = \begin{bmatrix} 0 & 1 \\ -6 & 3.5 \end{bmatrix} \mathbf{x}(t)\,, \mathbf{x}(0) = [1 \quad 1]^\top .$$

The communication channel is corrupted by a standard white Gaussian noise ($\dot{\mathbf{W}}(t)$, $\sigma^2 = 1$) and is assumed to have the power constraint $\mathscr{P} = 13$ ($\mathsf{SNR} = \mathscr{P}/\sigma^2 = 13$).

The design procedure follows the three steps proposed in the previous section, following an initialization stage:

1) The estimator is initialized with $\hat{\mathbf{x}}_0 = [0, 0]^\top$, and $P(0)$ is set to a 2×2 unit matrix;
2) We choose the basis functions as

$$\beta_1(t) = \sqrt{2}\sin(200\pi t) \text{ and } \beta_2(t) = \sqrt{2}\cos(200\pi t)$$

respectively .

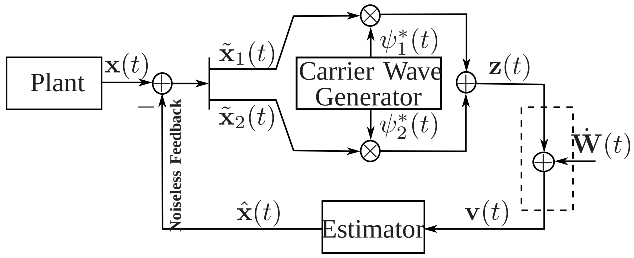

Fig. 3. Feedback AM Estimation

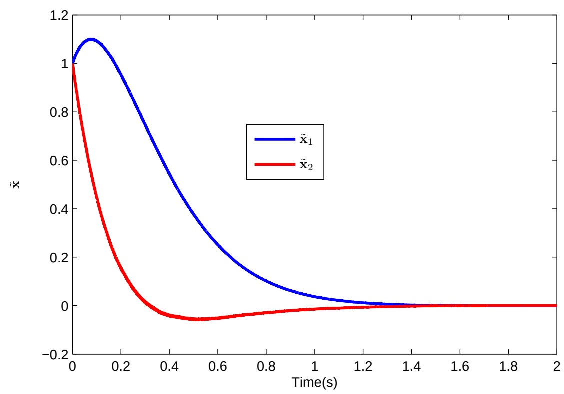

Fig. 4. State Error

3) We conduct the water filling algorithm to determine the optimal convergence rate $\nu^* = -3$ and weights $\eta_1 = 0.6299\,, \eta_2 = 3.5501$. In turn we have

$$\beta^*(t) = \begin{bmatrix} -0.7901\sin(200\pi t) - 2.3186\cos(200\pi t) \\ 0.4114\sin(200\pi t) + 4.4532\cos(200\pi t) \end{bmatrix}$$

4) The carrier waves $\psi_1^*(t)$ and $\psi_2^*(t)$, as well as the estimator, can be generated by solving the matrix differential equation (Ricatti).

Figure 4 shows the time-history of the state error $\tilde{\mathbf{x}}(t)$; Fig. 5 shows the modulated channel input and Fig. 6 shows the noise-corrupted channel output.

The simulation result is consistent with the theory developed in this paper and exhibits fast estimation error convergence in the presence of channel noise and power constraint. Compared with traditional amplitude modulation communications, where carrier waves are usually chosen as sinusoidal signals with constant amplitudes, this method explicitly uses the knowledge of the signal dynamics (A) to generate a set of carrier waves to meet the needs of optimal estimation. This example also suggests that the method can be extended to more practical scenarios for the simplicity of amplitude modulation in communication systems.

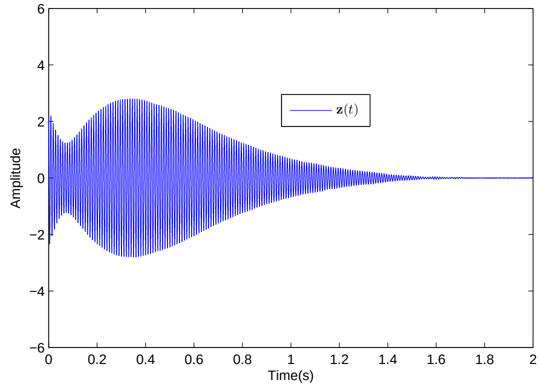

Fig. 5. Channel Input

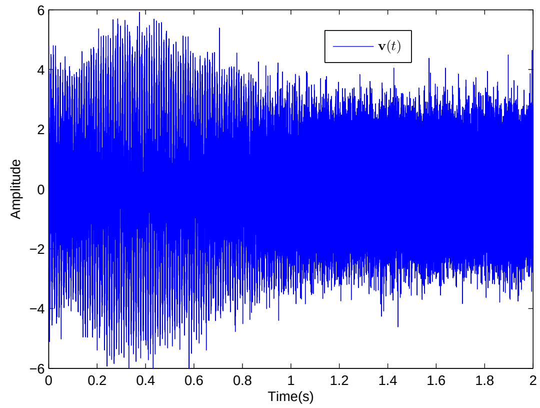

Fig. 6. Channel Output

VI. Conclusion

In this paper, we develop a design method to solve the optimal estimation problem with limited information. The objective is achieved by first fixing the structure of the transmitter and estimator by using conditional Kalman-Bucy filtering theory. Then the optimal parameters of the given structure are determined by a water-filling like technique by distributing the available channel input power to properly address the state-space of the dynamics to be estimated. The resulting communication/estimation scheme turns out to be surprisingly simple and fits into the conventional amplitude modulation framework with modified carrier waveforms, as shown in the example. The future research includes extension to digital communications and noisy feedbacks.

References

[1] G. Nair, F. Fagnani, S. Zampieri, and R. Evans, "Feedback control under data rate constraints: An overview," *PROCEEDINGS-IEEE*, vol. 95, no. 1, p. 108, 2007.

[2] R. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems," *IEEE Transactions on Automatic Control*, vol. 45, no. 7, pp. 1279–1289, 2000.

[3] A. SAHAI and S. MITTER, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link. part i: Scalar systems," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3369–3395, 2006.

[4] G. Como, F. Fagnani, and S. Zampieri, "Anytime reliable transmission of real-valued information through digital noisy channels," in *46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1473–1480, IEEE, 2008.

[5] J. Braslavsky, R. Middleton, and J. Freudenberg, "Feedback stabilization over signal-to-noise ratio constrained channels," *IEEE Transactions on Automatic Control*, vol. 52, no. 8, pp. 1391–1403, 2007.

[6] J. Freudenberg, R. Middleton, and J. Braslavsky, "Minimum variance control over a gaussian communication channel," *IEEE Transactions on Automatic Control*. Accepted for publication.

[7] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1196–1201, 2004.

[8] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1549–1561, 2004.

[9] N. Martins and M. Dahleh, "Feedback control in the presence of noisy channels:Bode-like fundamental limitations of performance," *IEEE Transactions on Automatic Control*, vol. 52, no. 7, pp. 1604–1615, 2008.

[10] E. Johannesson, A. Ghulchak, A. Rantzer, and B. Bernhardsson, "MIMO Encoder and Decoder Design for Signal Estimation," in *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems–MTNS*, vol. 5, 2010.

[11] E. Johannesson, A. Rantzer, B. Bernhardsson, and A. Ghulchak, "Encoder and decoder design for signal estimation," in *American Control Conference (ACC), 2010*, pp. 2132–2137, IEEE, 2010.

[12] N. Martins, M. Dahleh, and J. Doyle, "Fundamental limitations of disturbance attenuation in the presence of side information," *IEEE Transactions on Automatic Control*, vol. 52, no. 1, pp. 56–66, 2007.

[13] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[14] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 172–182, 1966.

[15] N. Elia, "When Bode meets Shannon: control-oriented feedback communication schemes," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, p. 1477, 2004.

[16] J. Liu and N. Elia, "Convergence of fundamental limitations in feedback communication, estimation, and feedback control over Gaussian channels," *Arxiv preprint arXiv:0910.0320*, 2009.

[17] Y.-H. Kim, "Feedback capacity of stationary gaussian channels," *IEEE Transactions on Information Theory*, vol. 56, pp. 57–85, Jan. 2010.

[18] R. Lipster and A. Shiryaev, *Statistics of Random Processes, Vol. 2: Applications*. Springer Verlag, Berlin, 2001.

[19] S. Ihara, *Information theory for continuous systems*. World Scientific Pub Co Inc, 1993.

[20] T. Cover, J. Thomas, J. Wiley, and W. InterScience, *Elements of information theory*. Wiley Wiley-Interscience, 2006.

[21] M. Davis, *Linear estimation and stochastic control*. Chapman and Hall Mathematics Series, 1977.

[22] S. Yu and P. Mehta, "Bode-like fundamental performance limitations in control of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 6, pp. 1390–1405, 2010.

[23] R. Bhatia, *Matrix analysis*. Springer Verlag, 1997.

APPENDIX

Proof Lemma 4.3: The trace differential Lyapunov equation derived from the Riccati equation in (IV.1) can be written as

$$\begin{aligned}\frac{d\text{trace}(P(t))}{dt} &= \text{trace}(A^\top + A - \frac{1}{\sigma^2} P^{1/2}(t)\psi(t,\mathbf{v}_0^t)\psi^\top(t,\mathbf{v}_0^t)P^{1/2}(t))(P(t)) \\ &= \text{trace}((A^\top + A - \beta(t)\beta^\top(t))P(t))\,,\end{aligned}$$

Let the symmetric matrix $M(t)$ be defined as $M(t) \triangleq \frac{1}{2}(A^\top + A - \beta(t)\beta^\top(t))$. We can define an auxiliary Lyapunov equation as

$$\frac{d\bar{P}(t)}{dt} = M(t)\bar{P}(t) + \bar{P}(t)M(t), \bar{P}(0) = P(0)$$

Obviously we see that trace$(\bar{P}(t)) = $ trace$(P(t))$. Solving the auxiliary Lyapunov equation, we have

$$\bar{P}(t) = \exp\left(\int_0^t M(\tau)d\tau\right)\bar{P}(0)\exp\left(\int_0^t M(\tau)d\tau\right).$$

Then we have

$$\begin{aligned}\text{trace}(P(t)) &= \text{trace}(\bar{P}(t))\\ &= \text{trace}\left(\exp\left(\int_0^t 2M(\tau)d\tau\right)P(0)\right)\end{aligned} \tag{VI.1}$$

The following bounds hold [**?**]

$$\lambda_{\min}(P(0))\text{trace}\left(\exp\left(\int_0^t 2M(\tau)d\tau\right)\right) \le \text{trace}\left(\exp\left(\int_0^t 2M(\tau)d\tau\right)P(0)\right) \le \lambda_{\max}(P(0))\text{trace}\left(\exp\left(\int_0^t 2M(\tau)d\tau\right)\right)$$

$\frac{1}{t}\int_0^t 2M(\tau)d\tau$ are $P(0)$ are symmetric and can be diagonalized by $n \times n$ orthornormal matrix $Q(t)$

$$Q^\top(t)\int_0^t 2M(\tau)d\tau Q(t) = \text{diag}\{\lambda_1^M(t), ..., \lambda_n^M(t)\}.$$

Note that

$$\max_i(\exp(\lambda_i^M(t)/t)) \le \left[\sum_{i=1}^n \exp(\lambda_i^M(t))\right]^{1/t} \le n^{1/t}\max_i(\exp(\lambda_i^M(t)/t))$$

Taking the limit and considering the continuity of $\max(\cdot)$ $\exp(\cdot)$ and $\lambda_{\max}$, and increasing $\exp(\cdot)$ we have

$$\begin{aligned}&\lim_{t\to\infty}\left[\sum_{i=1}^n \exp(\lambda_i^M(t))\right]^{1/t}\\ &= \lim_{t\to\infty}\max_i(\exp(\lambda_i^M(t)/t))\\ &= \lim_{t\to\infty}(\exp(\max_i(\lambda_i^M(t)/t)))\\ &= \lim_{t\to\infty}\exp\left(\lambda_{\max}\left(\frac{1}{t}\int_0^t 2M(\tau)d\tau\right)\right)\\ &= \exp\left(\lambda_{\max}\left(\lim_{t\to\infty}\frac{1}{t}\int_0^t 2M(\tau)d\tau\right)\right)\end{aligned} \tag{VI.2}$$

This leads to the bounds

$$(\lambda_{\min}(P(0)))^{1/t}\left[\sum_{i=1}^n \exp(\lambda_i^M(t))\right]^{1/t} \le \left[\text{trace}\left(\exp\left(\int_0^t 2M(\tau)d\tau\right)P(0)\right)\right]^{1/t} \le (\lambda_{\min}(P(0)))^{1/t}\left[\sum_{i=1}^n \exp(\lambda_i^M(t))\right]^{1/t}$$

Notice that $\mathbf{E}P(t) = P(t)$. The Lyapunov exponent can now be calculated as

$$\begin{aligned}&\limsup_{t\to\infty}\frac{1}{t}\log\mathbf{E}\|\tilde{\mathbf{x}}(t)\|^2\\ &= \limsup_{t\to\infty}\frac{1}{t}\log\text{trace}(P(t))\\ &= \limsup_{t\to\infty}\frac{1}{t}\log\text{trace}\left(\exp\left(\int_0^t 2M(\tau)d\tau\right)P(0)\right)\\ &= \lambda_{\max}\left(\lim_{t\to\infty}\frac{1}{t}\int_0^t 2M(\tau)d\tau\right)\end{aligned}$$

$\square$