2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Intermittent Kalman Filtering: Eigenvalue Cycles and Nonuniform Sampling

Se Yong Park and Anant Sahai {separk,sahai}@eecs.berkeley.edu

***Abstract*— We develop the concept of an *eigenvalue cycle* to completely characterize the critical erasure probability for intermittent Kalman filtering. It is also proved that eigenvalue cycles can be easily broken if the original physical system is considered to be continuous-time — randomly-dithered nonuniform sampling of observations makes the critical erasure probability depend only on the dominant eigenvalue, making it almost surely $\frac{1}{|\lambda_{max}|^2}$.**

I. INTRODUCTION

This paper can best be viewed as a response to [1], which itself enhanced our understanding of the *intermittent Kalman filtering* problem introduced by Sinopoli *et al.* in [2]. The situation that this problem is modeling is that of estimation over a so-called *packet drop channel*. A memoryless[1] sensor samples the output of an unstable[2] continuous-time system, quantizes it to a "sufficient" number of bits, binds these bits together into a single packet, and transmits the packet to the estimator through a communication system. Due to network congestion or wireless fading, the entire transmitted packet may be lost[3] with a certain probability and this erasure process is further simplified to be independent and identically distributed (i.i.d.). The problem is designed to focus attention on the effect of losing packets and so the number of bits per packet is unconstrained. Formally, the problem is usually formulated in discrete time:

$$\mathbf{x}[n+1] = \mathbf{A}\mathbf{x}[n] + \mathbf{B}\mathbf{w}[n] \quad (1)$$

$$\mathbf{y}[n] = \beta[n]\left(\mathbf{C}\mathbf{x}[n] + \mathbf{v}[n]\right). \quad (2)$$

Here n is the non-negative integer-valued time index and the system variables can take on complex values — *i.e.* $\mathbf{x}[n] \in \mathbb{C}^m, \mathbf{w}[n] \in \mathbb{C}^g, \mathbf{y}[n] \in \mathbb{C}^l, \mathbf{v}[n] \in \mathbb{C}^l$. $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ are complex matrices with appropriate dimensions. The underlying randomness comes from the initial state $\mathbf{x}[0]$, the persistent driving disturbances $\mathbf{w}[n]$, the observation noises $\mathbf{v}[n]$ and the Bernoulli packet-drops $\beta[n]$. $\beta[n] = 0$ with probability p_e.

The objective is to find the best causal estimator $\widehat{\mathbf{x}}[n]$ of $\mathbf{x}[n]$ that minimizes the mean square error (MMSE) $\mathbb{E}[(\mathbf{x}[n] - \widehat{\mathbf{x}}[n])^\dagger(\mathbf{x}[n] - \widehat{\mathbf{x}}[n])]$. Without loss of generality, $\mathbf{x}[0], \mathbf{w}[n]$ and $\mathbf{v}[n]$ are assumed to be zero mean. $\mathbf{x}[0], \mathbf{w}[n]$ and $\mathbf{v}[n]$ are independent and have uniformly bounded second moments so that there exists a positive σ such that

$$\mathbb{E}[\mathbf{x}[0]\mathbf{x}[0]^T] \leq \sigma^2\mathbf{I}$$
$$\mathbb{E}[\mathbf{w}[n]\mathbf{w}[n]^T] \leq \sigma^2\mathbf{I}$$
$$\mathbb{E}[\mathbf{v}[n]\mathbf{v}[n]^T] \leq \sigma^2\mathbf{I}.$$

To prevent degeneracy, we also assume that there exists a positive σ' such that

$$\mathbb{E}[\mathbf{w}[n]\mathbf{w}[n]^T] \geq \sigma'^2\mathbf{I}.$$

We refer to the system (1) and (2) as *intermittent observable* if the MMSE is uniformly bounded. Since more erasures can be simulated from fewer erasures, it is obvious that there must be a threshold on p_e for intermittent observability:

Theorem 1 (Theorem 2. of [2]): Let $(\mathbf{A}, \mathbf{B})$ be controllable.[4] Then, there exists a threshold $p_e^\star$, such that for $p_e < p_e^\star$ the system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is intermittent observable and for $p_e \geq p_e^\star$ the system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is not intermittent observable.

In fact, we can see that the condition that induces $p^\star = 1$ is the system's stability and the condition that induces $p^\star = 0$ is the lack of system observability. Thus, we can think the intermittent observability as a concept that connects stability and observability. In [2], Sinopoli *et al.* thought of intermittent observability as a generalization of stability, and gave a lower bound to the critical erasure probability based on Lyapunov stability. However, their lower bound given in a linear matrix inequality (LMI) form, while computable, is neither tight in general nor does it give any insight into the solution. A more intuitive bound was given by Elia in [4].

Theorem 2 (Corollary 8.4. of [4]):

$$\frac{1}{\prod_i |\lambda_i|^2} \leq p_e^\star \leq \frac{1}{|\lambda_{max}|^2},$$

where the λ_is are the unstable eigenvalues (including those with multiplicity multiple times) of $\mathbf{A}$ and λ_{max} is the one with the largest magnitude.

Therefore, the core question in this area boils down to understanding the gap between $\frac{1}{\prod_i |\lambda_i|^2}$ and $\frac{1}{|\lambda_{max}|^2}$.

This work was supported by the Samsung Scholarship and the National Science Foundation (CCF-729122).

The authors are with the Department of Electrical Engineering and Computer Sciences at the University of California at Berkeley.

[1]This sensor memorylessness is at the heart of what Sinopoli, *et al.* are trying to model. Otherwise, the earlier results of [3] immediately reveal that the critical erasure probability only depends on the magnitude of the largest eigenvalue. However, Sahai's work implicitly pushed all of the estimation complexity into the sensor rather than at the so-called estimator that sees the unreliable output of a channel. Thus, there was a need to carefully separate out the impact of intermittent observations themselves.

[2]Following Sahai in [3], unstable systems are used to abstract real problems of system performance into simpler binary questions of stabilizability.

[3]Such losses need not come from a network — they could occur because of sensor occlusion or other sensor-level issue. That is why the issue of intermittent observations needs to be studied on its own.

[4]Since this assumption is trivially necessary, this is assumed to be true thoughout the paper unless it's mentioned otherwise.

In [1], Mo and Sinopoli found two interesting cases that give further insight into this question. The first is when $\mathbf{A}$ is diagonalizable and all eigenvalues of $\mathbf{A}$ have distinct magnitudes — then the critical erasure probability is $\frac{1}{|\lambda_{max}|^2}$ just it would be in the formulation of [3]. The second case is when $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} 1 & 1 \end{bmatrix}$ — the critical erasure probability is $\frac{1}{\prod_i |\lambda_i|^2} = \frac{1}{2^4}$. This second case showed that the gap is real and requiring packets to be about a scalar observation can have serious consequences.

In this paper, we introduce the concept of an *eigenvalue cycle* to formalize the insight of Mo and Sinopoli and characterize the critical erasure probability based on it. As a corollary, we show that in the absence of eigenvalue cycles the critical value becomes $\frac{1}{|\lambda_{max}|^2}$. Furthermore, we show that simply by introducing nonuniform sampling at the physical sensor, the eigenvalue cycles can be broken and the critical erasure probability becomes effectively $\frac{1}{|\lambda_{max}|^2}$.

The basic idea is to consider intermittent observability as an extension of observability. The condition for observability is that for all $s \in \mathbb{C}$

$$\begin{bmatrix} s\mathbf{I} - \mathbf{A} \\ \mathbf{C} \end{bmatrix} \text{ is full rank.}$$

Moreover, without loss of generality we can assume that $\mathbf{A}$ is in Jordan form. With this additional assumption, the observability condition can be further simplified.

Theorem 3 ([5]): The system $(\mathbf{A}, \mathbf{C})$ with a Jordan matrix $\mathbf{A}$ is observable iff for the Jordan blocks with the same eigenvalue, the column vectors of $\mathbf{C}$ that correspond to the first columns of these Jordan blocks are linearly independent.

For example, let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix}.$$

Then, $(\mathbf{A}, \mathbf{C})$ is observable if both $\begin{bmatrix} c_1 & c_3 \end{bmatrix}$ and $\begin{bmatrix} c_4 \end{bmatrix}$ are full rank. This characterization reminds us of a *divide-and-conquer* approach — first divide the observability problem into smaller problems according to the identical eigenvalues, check whether the smaller sub-problems are observable, and finally the answer for the original observability question is yes iff all the sub-problems' answers are positive. This divide-and-conquer approach informs our characterization of intermittent observability. To use divide-and-conquer, we have to answer the following three questions:
(a) What are the minimal irreducible sub-problems?
(b) How can we solve each sub-problem?
(c) How can we combine the answers of the sub-problems?
The next section gives the intuitive characterization of intermittent observability by resolving these questions.

Roughly speaking, the intermittent observability problem can be divided along Jordan block lines into smaller sub-problems as long as the ratio of eigenvalues from different sub-problems is not a root of unity. Therefore, all ratios of the eigenvalues belonging the smallest sub-problems are roots of unity. As the system evolves, these roots of unity show periodic behaviors, and we can find the critical erasure probability for each sub-problem by inspecting the periodic behavior and seeing how this thickens the relevant tail probabilities. Finally, the intermittent observability of the original system can be shown to be equivalent to that of the hardest-to-estimate sub-problem.

II. Eigenvalue Cycles and the Critical Erasure Probability

We begin with the formal definition of a cycle.

Definition 1: A multiset (a set that allows repetitions of its elements) $\{a_1, a_2, \cdots, a_l\}$ is called a cycle with length l and period p if $\left(\frac{a_i}{a_j}\right)^p = 1$ for all $i, j \in \{1, 2, \cdots, l\}$ and some $p \in \mathbb{N}$. Following convention, p is denoted[5] as $\min\left\{n \in \mathbb{N} : \left(\frac{a_i}{a_j}\right)^n = 1, \forall i, j \in \{1, 2, \cdots, l\}\right\}$.
For example, $\{a\}$ is a cycle with length 1 and period 1 by itself. $\{e^{j\omega}, e^{j(\omega + \frac{2\pi}{6})}\}$ is a cycle with length 2 and period 6. $\{e^j, e^{j\sqrt{2}}\}$ and $\{1, 2\}$ are not cycles. One trivially necessary condition for a_1, a_2 to belong to the same cycle is $|a_1| = |a_2|$. It can be also shown that cycles are closed under overlapping unions, meaning that if $\{a_1, a_2\}$ and $\{a_2, a_3\}$ are cycles, $\{a_1, a_2, a_3\}$ is also a cycle.

Now, we can define an eigenvalue cycle. By changing coordinates, the system equation (1) can always be equivalently written with a Jordan matrix $\mathbf{A}$. Even though the MMSE's value can also change by such a coordinate change, the condition for it to be bounded remains the same. Therefore, we can assume that $\mathbf{A}$ is a Jordan matrix without loss of generality. Then, the matrix $\mathbf{A}$ and $\mathbf{C}$ can be written as the following form:

$$\mathbf{A} = diag\{\mathbf{A}_{(\mathbf{1},\mathbf{1})}, \mathbf{A}_{(\mathbf{1},\mathbf{2})}, \cdots, \mathbf{A}_{(\mathbf{k},\mathbf{l_k})}\}$$
$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{(\mathbf{1},\mathbf{1})} & \mathbf{C}_{(\mathbf{1},\mathbf{2})} & \cdots & \mathbf{C}_{(\mathbf{k},\mathbf{l_k})} \end{bmatrix}$$

where

$\mathbf{A}_{(\mathbf{i},\mathbf{j})}$ is a Jordan block matrix with an eigenvalue $\lambda_{i,j}$
$\{\lambda_{i,1}, \cdots, \lambda_{i,l_i}\}$ is a cycle with length l_i and period p_i
For $i \neq i'$, $\{\lambda_{i,j}, \lambda_{i',j'}\}$ is not a cycle
$\mathbf{C}_{(\mathbf{i},\mathbf{j})}$ is a $l \times \dim \mathbf{A}_{(\mathbf{i},\mathbf{j})}$ matrix. (3)

Since cycles are closed under overlapping unions, the eigenvalues of $\mathbf{A}$ can be uniquely partitioned into maximal cycles, $\{\lambda_{i,1}, \cdots, \lambda_{i,l_i}\}$. We call these cycles *eigenvalue cycles* and we say $\mathbf{A}$ has no eigenvalue cycle if all of its eigenvalue cycles are period 1.
Denote

$$\mathbf{A_i} = diag\{\lambda_{i,1}, \cdots, \lambda_{i,l_i}\}$$
$$\mathbf{C_i} = \begin{bmatrix} {\mathbf{C}_{(\mathbf{i},\mathbf{1})}}_1 & \cdots & {\mathbf{C}_{(\mathbf{i},\mathbf{l_i})}}_1 \end{bmatrix}$$

where ${\mathbf{C}_{(\mathbf{i},\mathbf{j})}}_1$ implies the first column of $\mathbf{C}_{(\mathbf{i},\mathbf{j})}$.

Let l_i' be the minimum cardinality among the sets $S' \subseteq \{0, 1, \cdots, p_i - 1\}$ whose resulting $S := \{0, 1, \cdots, p_i - 1\} \setminus$

[5] We use $\frac{0}{0} = 1$, $\frac{1}{0} = \infty$, $1^\infty = \infty$ and $\frac{1}{\infty} = 0$.

$S' = \{s_1, s_2, \cdots, s_{|S|}\}$ makes

$$\begin{bmatrix} \mathbf{C_i}\mathbf{A_i}^{s_1} \\ \mathbf{C_i}\mathbf{A_i}^{s_2} \\ \vdots \\ \mathbf{C_i}\mathbf{A_i}^{s_{|S|}} \end{bmatrix}$$

be rank deficient, i.e. the rank is strictly less than l_i.

Then, the critical erasure probability is given as the following main theorem of the paper.

Theorem 4: The critical erasure probability of $(\mathbf{A}, \mathbf{C})$ is $\frac{1}{\max_i |\lambda_{i,1}|^{2\frac{p_i}{l'_i}}}$.

Proof: See [6]. ■

Here, we can notice that there is no assumption about stability or observability of the system. Let's first run a "reality check" on the theorem by looking at stable modes and unobservable modes. If $|\lambda_{i,1}| < 1$, $\frac{1}{|\lambda_{i,1}|^{2\frac{p_i}{l'_i}}} > 1$. Therefore, the stable modes do not contribute to the characterization of the critical erasure probability. If $(\mathbf{A_i}, \mathbf{C_i})$ are unobservable, $l'_i = 0$. So, $\frac{1}{|\lambda_{i,1}|^{2\frac{p_i}{0}}} = 0$ if $|\lambda_{i,1}| \geq 1$ and $\frac{1}{|\lambda_{i,1}|^{2\frac{p_i}{0}}} = \infty$ if $|\lambda_{i,1}| < 1$. Therefore, if the unobservable modes are stable they do not affect the intermittent observability of the system and if they are not the system is not intermittent observable even if $p_e = 0$.

Even though in general l'_i does not admit a closed form, it is computable for special cases.

Corollary 1: Let $(\mathbf{A}, \mathbf{C})$ be observable and $\mathbf{A}$ have no eigenvalue cycles (*i.e.* $\left(\frac{\lambda_i}{\lambda_j}\right)^n \neq 1$ for all $\lambda_i \neq \lambda_j$ and $n \in \mathbb{N}$). Then, the critical erasure probability of intermittent Kalman filtering is $\frac{1}{|\lambda_{max}|^2}$.

Proof: Since $\mathbf{A}$ has no eigenvalue cycles, p_i equal to 1 for all i and $\mathbf{A_i}$ are diagonal matrices. Moreover, by the observability condition and Theorem 3, $\mathbf{C_i}$ is full-rank. Thus, $l'_i = 1$ for all i and by Theorem 4 the critical erasure probability is $\frac{1}{\max_i |\lambda_{i,1}|^2} = \frac{1}{|\lambda_{max}|^2}$. ■

For a more precise understanding of the critical erasure probability, we will focus on the case of a row vector $\mathbf{C}$ — *i.e.* single-output systems. Heuristically, a row vector $\mathbf{C}$ is the worst among $\mathbf{C}$ matrices since a vector observation is clearly better than a scalar observation.

Furthermore, we will also restrict the periods of the all eigenvalue cycles of $\mathbf{A}$ to be primes[6]. The technical reason for this restriction is that prime periods give us a useful invariance property of sub-eigenvalue cycles. Let $\{\lambda_1, \lambda_2, \cdots, \lambda_l\}$ be an eigenvalue cycle with a prime period p'. Then, all subsets of $\{\lambda_1, \lambda_2, \cdots, \lambda_l\}$ with distinct elements are eigenvalue cycles with the same period p'. This invariance property need not hold for cycles with composite periods.

Corollary 2: Let $(\mathbf{A}, \mathbf{C})$ be observable, $\mathbf{C}$ be a row vector, and $\mathbf{A}$ have only prime-period eigenvalue cycles. Then, the critical erasure probability is $\frac{1}{\max_i |\lambda_{i,1}|^{\frac{2p_i}{p_i - l_i + 1}}}$.

[6]For convenience, we include 1 as a prime number here.

Proof: The proof follows from Theorem 4 and the following fact shown in [7]. Let p be a prime, $a_1, \cdots, a_n$ be pairwise incongruent mod p and $b_1, \cdots, b_n$ be pairwise incongruent mod p. Then,

$$\begin{bmatrix} e^{j\frac{a_1 b_1}{p}} & e^{j\frac{a_1 b_2}{p}} & \cdots & e^{j\frac{a_1 b_n}{p}} \\ e^{j\frac{a_2 b_1}{p}} & e^{j\frac{a_2 b_2}{p}} & \cdots & e^{j\frac{a_2 b_n}{p}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\frac{a_n b_1}{p}} & e^{j\frac{a_n b_2}{p}} & \cdots & e^{j\frac{a_n b_n}{p}} \end{bmatrix}$$

is full rank. ■

Since the full proof of Theorem 4 is too long to be conveyed here, we here briefly explain the key ideas based on three properties of the critical erasure probability that answer the three questions raised in the previous section. The basic idea of the proof is to look at the reverse process — just as it was done in [1]. For example, consider a scalar system

$$\begin{cases} x[n+1] = 2x[n] + w[n] \\ y[n] = \beta[n]x[n] \end{cases}.$$

Let $n - S$ be the most recent non-erased observation at time n, i.e. $S := \inf\{k \geq 0 : \beta[n-k] = 1\}$. The stopping time S is a geometric random variable, so if we use the estimator $\widehat{x}[n] = 2^S y[n-S] = 2^S x[n-S]$ the estimation error is upper bounded by

$$\mathbb{E}\left[(x[n] - \widehat{x}[n])^2\right] \leq \frac{\sigma^2}{2^2 - 1}\left(\left(\sum_{i=0}^{\infty}(1-p_e)(p_e 2^2)^i\right) - 1\right)$$

The right-hand side is finite if $p_e < \frac{1}{2^2}$, which gives a sufficiency proof. For necessity, a similar analysis can be done using the fact that the driving noise $w[n-S+1]$ is independent of the non-erased observations present up to the time n. This scalar example reveals that the most important quantity is the p.m.f. tail of S, $\exp \limsup_{s \to \infty} \ln \mathbb{P}\{S = s\}$, which is p_e in this case.

(1) Power property: The power property answers the question (b) of the previous section. Consider the example of [1].

$$\begin{cases} \mathbf{x}[n+1] = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \mathbf{x}[n] + \mathbf{w}[n] \\ y[n] = \beta[n] \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{x}[n] \end{cases}.$$

If we write the observability matrix, we immediately notice cyclic behavior:

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \mathbf{CA^{-1}} &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \\ \mathbf{CA^{-2}} &= \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \end{bmatrix} \\ \mathbf{CA^{-3}} &= \begin{bmatrix} \frac{1}{8} & -\frac{1}{8} \end{bmatrix} \cdots \end{aligned}$$

Notice that $\mathbf{C}, \mathbf{CA^{-2}}, \mathbf{CA^{-4}}, \cdots$ are linearly dependent and $\mathbf{CA^{-1}}, \mathbf{CA^{-3}}, \mathbf{CA^{-5}}, \cdots$ are linearly dependent. As observed in [1], we therefore need both even and odd time observations to estimate the states. Therefore, the stopping time S until we get enough observations to estimate the current states is $\inf\{k \geq 0 : k \geq k_1, k \geq k_2, \beta[n-k_1] = 1, \beta[n-k_2] = 1, k_1 \neq k_2 (\bmod 2)\}$. Now, the p.m.f. tail of S becomes $\exp \limsup_{s \to \infty} \ln \mathbb{P}\{S = s\} = p_e^{\frac{1}{2}}$. The condition for bounded mean-squared error is thus $p_e^{\frac{1}{2}} < \frac{1}{2^2}$, which is $p_e < \frac{1}{2^{2 \cdot 2}}$. This gives sufficiency. For necessity, denote the stopping time S' to be $\inf\{k \geq 0 : \beta[n-k] = 1, k \text{ is even}\}$. Then, the p.m.f. tail of S' is also $p_e^{\frac{1}{2}}$. $\mathbf{w}[n-S'+2]$ can be decomposed to $\mathbf{w}[n-S'+2] = \begin{bmatrix} 1 \\ 1 \end{bmatrix} w_1[n-S'+2] + \begin{bmatrix} 1 \\ -1 \end{bmatrix} w_2[n-S'+2]$. Using the fact that $w_2[n-S'+2]$ is independent of all non-erased observations up to time n, we can similarly prove necessity.

To generalize this idea, we have to use the idea of large deviations [8] which is equivalent to a union bound for simple cases. The idea goes as follows.

Consider test channels that are erasure-type channels which would make the observability Gramian rank-deficient. For the previous example, these would be the channel that erases every odd-time observations, the channel that erases every even-time observations and the channel that erase all observations. In the characterization, the set S' is a proxy for these test channels.

Measure the distance from the true channel to the test channels. In our case, the true channel is the channel without any restriction and the distance measure is the Hamming distance. For the test channels considered above, the distance to the odd-time-erasure channel is 1, the even-time-erasure channel is 1 and the all-erasure channel is 2.

Then, the large deviation principle says that the performance is decided by the minimum-distance test channel. For the example, the odd-time or even-time erasure channel whose distances are 1 will govern the performance. This minimum distance is denoted as l_i' in the characterization.

So the effect of the eigenvalue cycle is to thicken the tail of the stopping time until you get enough observations to estimate the states. Analytically, the effect is equivalent to taking a proper power to the p_e and hence the name "power property."

(2) Max combining : This answers the question (c) i.e. how we go from a single eigenvalue cycle to multiple eigenvalue cycles. Consider the following example with two eigenvalue cycles:

$$\begin{cases} \begin{bmatrix} x_1[n+1] \\ x_2[n+1] \\ x_3[n+1] \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1[n] \\ x_2[n] \\ x_3[n] \end{bmatrix} + \mathbf{w}[n] \\ y[n] = \beta[n] \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \mathbf{x}[n] \end{cases} . \quad (4)$$

Let $S_1 := \{k \geq 0 : \beta[n-k] = 1\}$, $S_2 := \{k > S_1 : \beta[n-k] = 1\}$ and $S_3 := \{k > S_2 : \beta[n-k] = 1\}$. Then, one can easily check that the p.m.f. tail of s_3 is p_e, *i.e.* $\exp \limsup_{s \to \infty} \ln \mathbb{P}\{S_3 = s\} = p_e$. Since we have three observations at time $n - S_1$, $n - S_2$ and $n - S_3$, by the pigeon-hole principle at least two among them have to be congruent mod 2. Without loss of generality, let S_1 and S_2 be even numbers. Then, $\mathbf{A}^{-S_1} \begin{bmatrix} x_1[n] \\ x_2[n] \\ x_3[n] \end{bmatrix} = \begin{bmatrix} 3^{-S_1} & 0 & 0 \\ 0 & 2^{-S_1} & 0 \\ 0 & 0 & 2^{-S_1} \end{bmatrix} \begin{bmatrix} x_1[n] \\ x_2[n] \\ x_3[n] \end{bmatrix} = \begin{bmatrix} 3^{-S_1} & 0 \\ 0 & 2^{-S_1} \end{bmatrix} \begin{bmatrix} x_1[n] \\ x_2[n] + x_3[n] \end{bmatrix}$ and likewise $\mathbf{A}^{-S_2} \begin{bmatrix} x_1[n] \\ x_2[n] \\ x_3[n] \end{bmatrix} = \begin{bmatrix} 3^{-S_2} & 0 \\ 0 & 2^{-S_2} \end{bmatrix} \begin{bmatrix} x_1[n] \\ x_2[n] + x_3[n] \end{bmatrix}$. Therefore, instantaneously at time $n - S_1$ and $n - S_2$ the system equation (4) behaves like

$$\begin{cases} \begin{bmatrix} x_1[n+1] \\ x_2[n+1] + x_3[n+1] \end{bmatrix} \\ \qquad = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1[n] \\ x_2[n] + x_3[n] \end{bmatrix} + \mathbf{w}'[n] \\ y[n] = \beta[n] \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{x}[n] \end{cases}$$

, which has no eigenvalue cycles. It is known in [1] that this new system is intermittent observable if $p_e < \frac{1}{3^2}$ which is equivalent condition to the case when we only have the state $x_1[n]$. When $p_e < \frac{1}{3^2}$ we can estimate $x_1[n]$ within a bounded error. By considering the residual estimation error $x_1[n] - \widehat{x}_1[n]$ as an additional observation noise, we can estimate $x_2[n]$ and $x_3[n]$ without worrying about the existence of $x_1[n]$. This idea is known as successive decoding in information theory [9].

Therefore, once we have Corollary 1, we can solve problems with multiple eigenvalue cycles one by one without worrying about the existence of the other eigenvalue cycles and the performance is bottlenecked by the hardest-to-estimate eigenvalue cycle. In this example, the intermittent observability condition is $p_e < \frac{1}{\max\{3^2, 2^{2 \cdot 2}\}} = \frac{1}{16}$. Necessity can easily be proved by a genie argument, *i.e.* we can give all states except the ones that correspond to the bottleneck eigenvalue cycle as side information to the estimator.

Therefore, the critical erasure probability of $\mathbf{A}$ with multiple eigenvalue cycles is decided by the bottleneck eigenvalue cycle.

(3) Separability of Eigenvalue Cycles: The remaining question is what are the minimal irreducible sub-problems, whose answer can be expected to be eigenvalue cycles from the discussion up to now. While discussing the max-combining property, we saw that if systems without eigenvalue cycles can be divided into scalar sub-systems, the result can be leveraged to general systems. In other words, we can separate a linear system into sub-systems where each sub-system has a single eigenvalue cycle. Thus, the question pins down to systems with no eigenvalue cycles, which is Corollary 1. In fact, a special case of Corollary 1 is shown in [1] when the magnitudes of the eigenvalues of a diagonal matrix $\mathbf{A}$ are distinct. The main technical insight for this case is that when $|a| > |b|$, for every $k \in \mathbb{R}$ we can find $n \in \mathbb{N}$ such that $\left(\frac{|a|}{|b|}\right)^{n'} > k$ for all $n' \geq n$.

The complete proof of Corollary 1 can be done in two steps. First, just relax the distinct-eigenvalue-magnitude condition to no eigenvalue cycle condition, but keep the system matrix diagonal. This step can be done using Weyl's criterion [10], which gives a necessary and sufficient condition for a sequence to behave like a uniform distribution on the interval $[0,1]$. For example, let $\mathbf{A} = \begin{bmatrix} e^{j} & 0 \\ 0 & e^{j\sqrt{2}} \end{bmatrix}$. Then $\mathbf{A}^n$ is equal to $\begin{bmatrix} e^{jn} & 0 \\ 0 & e^{j\sqrt{2}n} \end{bmatrix}$. By Weyl's criterion $(e^{jn}, e^{j\sqrt{2}n})$ behaves like a sample from $(e^{j\theta_1}, e^{j\theta_2})$ where θ_1 and θ_2 are independent random variables uniformly distributed on $[0, 2\pi]$. Thus, we can conclude that $(e^{jn}, e^{j\sqrt{2}n})$ do not behave in a periodic way. In fact, the effect of the hypothetical random variables $(e^{j\theta_1}, e^{j\theta_2})$ is quite similar to the actually randomly-dithered nonuniform sampling discussed in the next section.

The next step is to generalize from a diagonal matrix to an arbitrary Jordan matrix. A non-trivial Jordan block introduces polynomial terms in n to the observability Gramian. Based on this observation, in [6] we reduce the Jordan block problem to facts like "x^2 cannot be written as a linear combination of 1 and x," which are trivially true.

III. Intermittent Kalman Filtering with Nonuniform Sampling

In the previous section, we argued that the cyclic behavior caused by eigenvalue cycles is the only factor that prevents us from having the critical erasure probability be $\frac{1}{|\lambda_{max}|^2}$. Based on this understanding, we can look for a way to avoid this troublesome phenomenon. Here, nonuniform sampling is proposed as a simple way of breaking eigenvalue cycles and achieving the critical value $\frac{1}{|\lambda_{max}|^2}$.

As an intuitive example, consider $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Then, $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \mathbf{A}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{A}^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \cdots$. What the eigenvalue cycle is capturing is that half of $\mathbf{A}, \mathbf{A}^2, \mathbf{A}^3, \cdots$ are identical. Therefore, the question is how we can make every matrix in $\mathbf{A}, \mathbf{A}^2, \mathbf{A}^3, \cdots$ distinct. To simplify the question, consider the sequence of $-1, 1, -1, 1, \cdots$ which corresponds to $(2,2)$ elements of $\mathbf{A}, \mathbf{A}^2, \mathbf{A}^3, \cdots$.

Rewrite this sequence $-1, 1, -1, 1, \cdots$ as $(e^{j\pi})^1, (e^{j\pi})^2, (e^{j\pi})^3, (e^{j\pi})^4, \cdots$ and introduce a jitter t_i to each sampling time. The resulting sequence becomes $(e^{j\pi})^{1+t_1}, (e^{j\pi})^{2+t_2}, (e^{j\pi})^{3+t_3}, (e^{j\pi})^{4+t_4}, \cdots$ and if t_is are uniformly distributed i.i.d. random variables on $[0,T]$ each element in the sequence is distinct almost surely as long as $T > 0$.

Operationally, this idea can be implemented as follows: at design-time, the sensor and the estimator agree on the nonuniform sampling pattern which is a realization of i.i.d. random variables whose distribution is uniform on $[0,T] (T > 0)$. Whenever the sensor samples the system, it jitters its sampling time according to this nonuniform pattern. Knowing the sampling time jitter, the sampled continuous-time system looks like a discrete *time-varying* system to the estimator. The joint Gaussianity between the observation and the state is preserved, and furthermore, Kalman filters are optimal even for time-varying systems! This intermittent Kalman filtering problem with nonuniform samples has the critical erasure probability $\frac{1}{|\lambda_{max}|^2}$ almost surely. Therefore, an eigenvalue cycle is breakable by nonuniform sampling.

One may be bothered by the probabilistic argument on the nonuniform sampling pattern. However, this probabilistic proof is an indirect argument for the existence of an appropriate deterministic nonuniform sampling pattern, which is similar to how the existence of capacity achieving codes is proved in information theory [11].

To write this more formally, consider a continuous-time dynamic system. A typical model for a linear continuous-time system is an Ornstein-Uhlenbeck Process [12],

$$d\mathbf{x_c}(t) = \mathbf{A_c}\mathbf{x_c}(t)dt + \mathbf{B_c}d\mathbf{W_c}(t) \tag{5}$$

where $t \geq 0$, $\mathbf{x_c} \in \mathbb{C}^m$ and $\mathbf{W_c}(t)$ is a g-dimensional Wiener process. The matrices $\mathbf{A_c}$ and $\mathbf{B_c}$ are complex matrices with proper sizes.
The observation process $\mathbf{y_c}(t)$ is modeled by

$$\mathbf{y_c}(t) = \mathbf{C_c}\mathbf{x_c}(t) + \mathbf{D_c}d\mathbf{V_c}(t) \tag{6}$$

where $t \geq 0$, $\mathbf{y_c} \in \mathbb{C}^l$ and $\mathbf{V_c}(t)$ is an h-dimensional Wiener process. The matrices $\mathbf{C_c}$ and $\mathbf{D_c}$ are complex matrices with proper sizes.

Let's say we want to sample the system with interval I. A common way to model sampling of the continuous-time process is by introducing an integration filter at the sensor, *i.e.* for the uniform sampling case, the nth sample $\mathbf{y}[n]$ is

$$\mathbf{y}[n] = \int_{(n-1)I}^{nI} \mathbf{y_c}(t)dt \tag{7}$$

where the integral is an Ito's integral [12].

Nonuniform sampling can be thought of in two ways with respect to the sensor's integrator: (1) The integration intervals are of non-uniform length but the starting time of the integration stays uniform. (2) The starting times of the integration are non-uniform but we keep the integration intervals of uniform length. Since the analysis and performance is similar in both cases, we will focus on the latter case. To take the nth sample of the system, the sensor integrates $\mathbf{y_c}(t)$ from $(n-1)I - t_n$ to $nI - t_n$:

$$\begin{aligned}
\mathbf{y_o}[n] &= \int_{(n-1)I-t_n}^{nI-t_n} \mathbf{y_c}(t)dt. \\
&= \underbrace{\left(\int_0^I \mathbf{C_c}e^{-\mathbf{A_c}(I-t)}dt\right)}_{:=\mathbf{C}} \mathbf{x_c}(nI - t_n) \\
&\quad - \int_{(n-1)I-t_n}^{nI-t_n} \int_t^{nI-t_n} \mathbf{C_c}e^{\mathbf{A_c}(t-t')}\mathbf{B_c}d\mathbf{W_c}(t')dt \\
&\quad + \int_{(n-1)I-t_n}^{nI-t_n} \mathbf{D_c}d\mathbf{V_c}(t).
\end{aligned} \tag{8}$$

since the solution of (5) is

$$\mathbf{x_c}(t) = e^{\mathbf{A_c}t}\mathbf{x_c}(0) + \int_0^t e^{\mathbf{A_c}(t-t')}\mathbf{B_c}d\mathbf{W_c}(t'). \tag{9}$$

Denote $\mathbf{v}[n]$ as the sum of the second and third term of (8). Here, we can notice that $\mathbf{v}[n]$s can be correlated due to possible overlapping in the integration intervals. Finally, the received observation at the estimator can be written as

$$\mathbf{y}[n] = \beta[n]\mathbf{y_o}[n] = \beta[n]\left(\mathbf{C}\mathbf{x_c}(nI - t_n) + \mathbf{v}[n]\right). \quad (10)$$

Since the observability of a continuous time system does not necessarily imply the observability of the sampled system, the observability condition of $(\mathbf{A_c}, \mathbf{C})$ is required, which is a necessary condition for $(\mathbf{A_c}, \mathbf{C_c})$ to be observable. Now, we can present the main theorem of this section.

Theorem 5: Let t_n be i.i.d. random variables uniformly distributed on $[0, T](T > 0)$ and the system $(\mathbf{A_c}, \mathbf{B_c}, \mathbf{C})$ is controllable and observable. The system is almost surely intermittent observable if and only if $P_e < \frac{1}{|e^{2\lambda_{max} I}|}$.

Proof: See [6]. ■

Remark 1: Since $\exp\left((\text{eigenvalue of } \mathbf{A_c})I\right)$ corresponds to the eigenvalue of the sampled discrete time system, the critical value of Theorem 5 is equivalent to that of Corollary 1. The nonuniform sampling allows us to no longer care if eigenvalue cycles could exist for the original continuous-time system.

Remark 2: The assumption that t_ns are identically and uniformly distributed is not minimal and can be relaxed as following: t_ns are independent random variables and there exists a, c such that $\mathbb{P}\{|t_n| \geq a\} = 0$ and $\mathbb{P}\{t_n \in B\} \leq c\mathcal{L}(B)$ for all $B \in \mathbb{B}$ and $n \in \mathbb{N}$ where $\mathbb{B}$ implies Borel σ-algebra and $\mathcal{L}$ implies Lebesgue measure.

Nonuniform sampling is the right way of breaking eigenvalue cycles from a practical point of view. So the critical erasure probability of $\frac{1}{|\lambda_{max}|^2}$ can be achieved not only by using the computationally challenging estimation-before-packetization strategy of [3], but also by the simple memoryless approach of dithered sampling before packetization. And so, even if the sensors were themselves distributed, the critical erasure probability with nonuniform sampling is still *critical value optimal* in a sense that they can achieve the same critical erasure probability to the sensors with causal or noncausal information about the erasure pattern or with any complexity.

IV. Discussion

The intermittent Kalman filtering problem was first motivated from control over communication channels. Therefore, the common belief about the problem is that it falls into the intersection of control and communication. However, if the plant is unstable the transmission power of the sensor diverges to infinity if it is really going to pack an ever increasing number of bits in there. Therefore, it is hard to say that intermittent Kalman filtering really has a strong connection to communication theory. The results of this paper argue instead that the intersection of control and signal processing — especially sampling theory — is the right conceptual category for intermittent Kalman filtering. It should thus be interesting to explore the connection between the result of this paper with the classical results of sampling theory.

Arguably, the closest problem to intermittent Kalman filtering is that of observability after sampling. As we mentioned earlier, the observability of $(\mathbf{A_c}, \mathbf{C_c})$ in (5) and (6) does not implies the observability of $(\mathbf{A_c}, \mathbf{C})$ in (5) and (8). The well-known sufficient condition is:

Theorem 6 (Theorem 6.9. of [5]): Suppose $(\mathbf{A_c}, \mathbf{C_c})$ is observable. A sufficient condition for its discretized system with sampling interval I to be observable is that $\frac{|\Im(\lambda_i - \lambda_j)I|}{2\pi} \notin \mathbb{N}$ whenever $\Re(\lambda_i - \lambda_j) = 0$.

Since the eigenvalue of the sampled system is given as $\exp(\lambda_i I)$, Corollary 1 can be written as the following corollary for a sampled system.

Corollary 3: Suppose $(\mathbf{A_c}, \mathbf{C_c})$ is observable. A sufficient condition for its discretized system with sampling interval I to have $\frac{1}{|e^{2\lambda_{max} I}|}$ as a critical erasure probability is that $\frac{|\Im(\lambda_i - \lambda_j)I|}{2\pi} \notin \mathbb{Q}$ whenever $\Re(\lambda_i - \lambda_j) = 0$.

The idea of breaking cyclic behavior using non-uniform sampling is also shown in the context of sampling multiband signals [13]. The lower bound on the sampling rate is known as the Lebesgue measure of the spectral support of the signal sampled. To achieve this lower bound for a general multiband signal, a nonuniform sampling pattern has to be used. Moreover, nonuniform sampling is also intimately connected to the "incoherence" conditions necessary for the currently hot field of compressed sensing [14].

References

[1] Y. Mo and B. Sinopoli, "A characterization of the critical value for Kalman filtering with intermittent observations," *Proceedings of the 47th IEEE Conference on Decision and Control CDC*, 2008.

[2] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Automat. Contr.*, vol. 49, no. 9, pp. 1453–1464, Sept. 2004.

[3] A. Sahai, "Anytime information theory," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.

[4] N. Elia, "Remote stabilization over fading channels," *Systems and Control Letters*, vol. 54, no. 3, pp. 239–249, Mar. 2005.

[5] C.-T. Chen, *Linear System Theory and Design*, 3rd ed. New York, NY: Oxford University Press, 1999.

[6] S. Y. Park and A. Sahai, "Intermittent Kalman filtering: Eigenvalue cycles and nonuniform sampling," In preparation.

[7] R.J.Evans and I.M.Isaacs, "Generalized Vandermonde determinants and roots of unity of prime order," *Proceedings of the American Mathematical Society*, vol. 58, July 1976.

[8] A.Dembo and O.Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer, 1998.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.

[10] K. Kuipers and H. Niederreitter, *Uniform Distribution of Sequences*. Wiley, 1974.

[11] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423,623–656, Oct. 1948.

[12] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. Springer, Nov. 2003.

[13] P.P.Vaidyanathan and V.C.Liu, "Efficient reconstruction of band-limited sequences from nonuniformly decimated versions by use of polyphase filter banks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 11, Nov. 1990.

[14] D. Donoho, "Compressed sensing." *Manuscript*. [Online]. Available: http://www-stat.stanford.edu/donoho/Reports/2004