On the performance of Kalman filtering with intermittent observations: a geometric approach with fractals

Andrea Censi

Abstract—This paper describes the stationary distribution of the a-posteriori covariance matrix of a Kalman filter when the availability of measurements is subject to random phenomena such as lossy network links. If a certain non-overlapping condition is satisfied, the distribution has a fractal nature, and there exists a closed-form expression for the cdf, which is a singular function. If the condition is not satisfied, deciding whether the cdf is singular or not, even in the scalar case, is at least as hard as some open problems in measure and number theory.
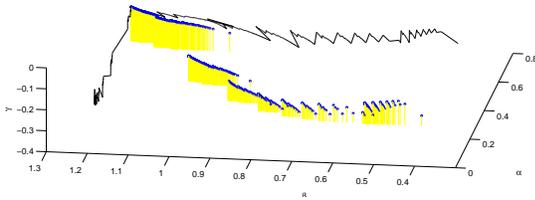

Figure 1. The sample distribution of $\mathbf{P}$ can be visualized by letting $\mathbf{Y} = \mathbf{P}^{-1} = [\alpha^2, \alpha\beta\gamma; \alpha\beta\gamma, \beta^2]$ and plotting it in the (α, β, γ) space. This picture for the system $\mathbf{A} = [2, 0.5; 0, 1]$, $\mathbf{Q} = [2, -0.5; -0.5, 1]$, $\mathcal{I} = [1, 0; 0, 1]$ suggests that the distribution has a fractal nature. The blue dots are the samples, the black lines are the projection on the (β, γ) and (α, γ) subspaces.

(a) Varying Q: ■ $Q = 0$; ■ $Q = 1$; ■ $Q = 2$; ■ $Q = 3$

(b) Varying the independent drop rate ($\alpha = 1 - d$, $\beta = d$): ■ $d = 0.5$; ■ $d = 0.3$; ■ $d = 0.1$; ■ $d = 0.9$

(c) Varying the parameters of the Markov Chain: ■ $\alpha = 0.5$, $\beta = 0.5$; ■ $\alpha = 0.3$, $\beta = 0.3$; ■ $\alpha = 0.1$, $\beta = 0.1$; ■ $\alpha = 0.9$, $\beta = 0.9$;
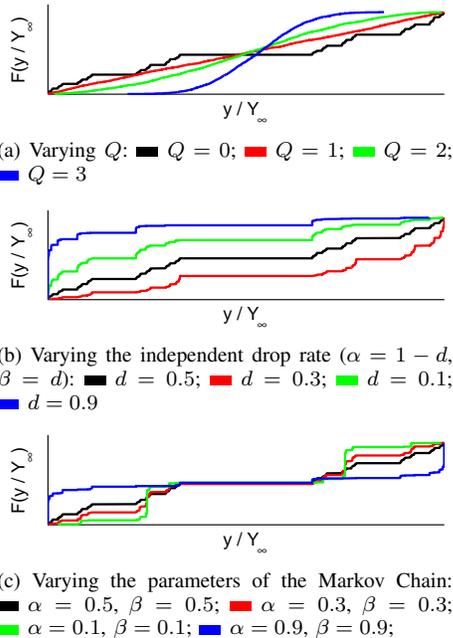
Figure 2. The scalar case contains most of the complexity of the matrix case. The figures show the cdf $F(y) = \mathbb{P}\{Y < y\}$ (the plots are normalized to Y_∞). The black plot is for the nominal system $A = 1/\sqrt{3}$, $Q = 0$, $\mathcal{I} = 1$ and packet dropping governed by a Markov chain with transition matrix $T = [\alpha, 1 - \alpha; 1 - \beta, \beta]$, $\alpha = \beta = 0.5$.

The author is with the Control & Dynamical Systems department, California Institute of Technology. Address: MC 107-81, 1200 E. California Blvd., 91125, Pasadena, CA. E-mail: andrea@cds.caltech.edu

I. INTRODUCTION

Consider the linear time-invariant dynamical system:

$$\begin{aligned} \boldsymbol{x}(k+1) &= \mathbf{A}\,\boldsymbol{x}(k) + \mathbf{B}\,\boldsymbol{\omega}(k) \\ \boldsymbol{y}(k) &= \mathbf{C}\,\boldsymbol{x}(k) + \boldsymbol{v}(k) \end{aligned}$$

where the sequences $\boldsymbol{\omega}(k)$, $\boldsymbol{v}(k)$ are white Gaussian with zero mean and covariance matrix equal to the identity. If the initial estimate of the state is Gaussian, the conditional distribution of the state given all previous measurements is still Gaussian. The Kalman filter computes this distribution exactly by propagating a mean and a covariance matrix $\mathbf{P}(k)$. Under the assumption that $(\mathbf{A}, \mathbf{B})$ is stabilizable and $(\mathbf{A}, \mathbf{C})$ is detectable, the covariance $\mathbf{P}(k)$ converges to a unique value $\mathbf{P}_\infty$ from any initial value $\mathbf{P}(0)$. If some measurements are missing (for example, because measures are sent over an unreliable network), then the Kalman filter is still the correct estimator, provided that only the prediction update is done when a packet is missing [1]. If there is some statistical description about the loss of measurements in time, then it makes sense to consider the stationary distribution of $\mathbf{P}(k)$. This distribution exists if the sequence of packet drops has certain ergodic properties; see [2] for an extensive discussion of existence of the stationary random variable $\mathbf{P}$.

Once it is ascertained that the stationary distribution exists, it is of interest to describe its properties. Sinopoli *et al.* [1] proved that, for independent packet drops, there is a critical value d_c of the packet drop rate d such that, for $d > d_c$, $\mathbb{E}\{\mathbf{P}\}$ is unbounded. This is of theoretical interest, e.g., in LQG control. The actual knowledge of d_c is less useful, because a filter operating in conditions of d close to d_c would have arbitrarily high expected covariance not matching the underlying physical process. Of more interest is the computation of an upper bound for $\mathbb{E}\{\mathbf{P}\}$, which [1] shows can be numerically computed as the fixed point of a certain modified algebraic Riccati equation. Other works that improved on such results are [3], [4], [5].

Shi *et al.* [6] considered this problem from another point of view. Recognizing that the unboundedness of $\mathbb{E}\{\mathbf{P}\}$ depends on zero-probability events, they proposed that a better characterization of the performance must take into account the whole distribution of $\mathbf{P}$. In particular, they derived upper and lower bounds for $\mathbb{P}(\{\mathbf{P} \preceq \mathbf{M}\})$ for any matrix $\mathbf{M}$, in the case of an invertible matrix $\mathbf{C}$; in [7], the limitation on $\mathbf{C}$ is removed. The present paper improves on this line of research by describing the whole distribution of $\mathbf{P}$ and giving closed-form expressions for $\mathbb{P}(\{\mathbf{P} \preceq \mathbf{M}\})$ in some special cases.

The parametrization used in this paper is the inverse of covariances, also known as information matrices. Let $\mathcal{P}$ be the set of positive semidefinite matrices and let $\mathcal{P}_* \subset \mathcal{P}$ be the set of positive definite matrices. Let $\mathbf{Y} \triangleq \mathbf{P}^{-1}$, $\mathcal{I} \triangleq \mathbf{C}^T\mathbf{C} \succeq 0$

and $\mathbf{Q} \triangleq \mathbf{B}\mathbf{B}^T \succeq 0$. Then the update functions for $\mathbf{Y}$ in the case a packet is received (g) and in the case it is not (h) can be written as follows:

$$g: \quad \mathbf{Y} \mapsto \quad (\mathbf{A}\mathbf{Y}^{-1}\mathbf{A}^T + \mathbf{Q})^{-1} + \mathcal{I} \tag{1}$$

$$h: \quad \mathbf{Y} \mapsto \quad (\mathbf{A}\mathbf{Y}^{-1}\mathbf{A}^T + \mathbf{Q})^{-1} \tag{2}$$

The advantages in using this formulation are that equations have a particularly simple expression; in particular, the difference between h and g is a constant term *not* depending on the current $\mathbf{Y}$, and that the distribution has support in the bounded set $\{\mathbf{Y} \mid \mathbf{0} \preceq \mathbf{Y} \preceq \mathbf{P}_\infty^{-1}\}$. Of course, this is absolutely not the preferred form one might use in an actual implementation, as g and h contain two matrix inversions the size of the state. Note also that, because of the matrix inversions, they are formally defined on $\mathcal{P}_*$; however, they can be extended by continuity on $\mathcal{P}$ (in particular, $h(\mathbf{0}) = \mathbf{0}$ and $g(\mathbf{0}) = \mathcal{I}$). In the following, "$h$" and "$g$" refer to such extensions.

The assumption throughout the paper is that $(\mathbf{A}, \mathbf{B})$ is controllable and $(\mathbf{A}, \mathbf{C})$ is detectable . The additional assumption with respect to the literature is that $\mathbf{A}$ is invertible; this is reasonable for a discrete-time linear system obtained by the discretization of a continuous-time system. The packet dropping process is parametrized in the form of a probability measure $\mathcal{M}$ on the packet dropping sequences, and there are no requirements on it, except that it must induce a stationary distribution for $\mathbf{P}$ (see the discussion in [2]); Bernoulli, Markov chains as well as Hidden Markov Models are allowed.

Some of the results in the paper rely on the following non-overlapping condition (NOC):

Condition 1: (Non-overlapping condition)

$$h\left(\mathbf{P}_\infty^{-1}\right) \prec g\left(\mathbf{0}\right) \tag{3}$$

This paper shows that, if the NOC is satisfied, then the resulting distribution of $\mathbf{Y}$, and hence the distribution of $\mathbf{P}$, has a fractal nature, in the sense that $\mathbb{P}(\{\mathbf{Y} \preceq \mathbf{M}\})$ is a singular function in $\mathbf{M}$. It is also possible to give a closed-form expression for $\mathbb{P}(\{\mathbf{Y} \preceq \mathbf{M}\})$ for any matrix $\mathbf{M}$. If the NOC is not satisfied, then, even in the scalar case, deciding whether the distribution is singular or not is equivalent to some open problems in number theory. In that case, iterative algorithms will be discussed.

It is worth noting that the NOC is a relatively strong assumption as it implies that $\mathcal{I} \succ \mathbf{0}$. This follows from (3) by noting that it is equivalent to $(\mathbf{A}\mathbf{P}_\infty\mathbf{A}^T + \mathbf{Q})^{-1} \prec \mathcal{I}$. In turn, $\mathcal{I} = \mathbf{C}^T\mathbf{C} \succ 0$ implies that $\mathbf{C}$ has full rank.

This paper tries to give a good intuition about the properties of the distribution of $\mathbf{P}$. Previous work used arguments which were essentially analytic and algebraic (Riccati equations, eigenvalues, limits, etc.); here the focus is on topological and geometric arguments. Section II introduces a number of mathematical preliminaries necessary for the study of fractals. Section II analyzes the scalar case, that presents most of the complexity of the full case, and is meant to be in an expository style. Section IV discusses the same arguments in a more precise way for the generic matrix case.

II. MATHEMATICAL PRELIMINARIES

In the words of Benoît Mandelbrot, fractals are the first mathematical objects that "capture the idea of scale-invariant roughness" [8]. There is no precise definition of "fractal": the word describes different phenomena with common properties, such as self-similarity and a fractional dimension (typically, they have Hausdorff dimension greater than its topological dimension). Fractals are studied through many tools, including topology, measure theory, and symbolic spaces; the rest of the section describes the relevant tools needed in the paper. A good reference text for the theory is [9], while [10] is a very nice introduction.

A. Symbolic spaces

A symbolic space $\{0, 1, \ldots, n\}^\mathbb{N}$ is the set of all infinite sequences over an ordered alphabet $\{0, 1, \ldots, n\}$. On this space, define the following metric: $d(x, y) = 2^{-k}$ where $k = \max\{m : x_i = y_i \text{ for } i < m\}$ is the index of the digit where the two sequences first differ. Let the topology be generated by this metric. The open sets are sets of sequences which have a common prefix p; these sets are commonly called "cylinders" and indicated as $[p]$. Let $\lhd$ be the lexicographical order on $\{0, \ldots, n\}^\mathbb{N}$: $x \lhd y$ if the two sequences x and y disagree for the first time at index i and $x_i < y_i$. Note that this is a strict total order: if $x \neq y$, either $x \lhd y$ or $y \lhd x$. Let "$\cdot$" indicate the composition operator; frequently the subsequent proofs use notations like $x = p \cdot 1 \cdot t$ to mean that x is composed by a (finite) prefix p, the digit 1, and a tail t. The symbol "$\overline{m}$" will refer to the infinite repetition of the digit m.

This formalism allows to use generic hidden Markov model for modeling packet drops. Rather than writing formulas for each case, the assumption is that there is some probability measure $\mathcal{M}$ defined on the cylinders of $\{0, 1\}^\mathbb{N}$. Intuitively, the notation $\mathcal{M}([011])$ indicates the probability that the last three events were: drop, received, received. The following is an explicit definition of $\mathcal{M}$ in the case of a Markov chain. Assume the transition matrix is $T = \left(\begin{smallmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{smallmatrix}\right)$. Then $\mathcal{M}([m_0 \cdots m_k]) = p(m_k) \prod_{i=0}^{k-1} p(m_i|m_{i+1})$, and for the single probabilities, $p(a) = (T^\infty)_{a,a}$, and $p(a|b) = T_{a,b}$.

B. Rényi β-expansions [11], [12]

Proposition 1: Choose a base $\beta > 1$. Every real number $x \in [0, \beta)$ can be represented as a power series $x = \sum_{i=0}^\infty b_i \beta^{-i}$, with $b_i \in \{0, 1, \ldots, \lfloor \beta \rfloor\}^\mathbb{N}$. The representation is unique for $\beta \geq 2$ (save for the usual identifications of the kind $0.\overline{9} = 1$), while for $\beta < 2$ almost every x has infinitely many representations. The "greedy" algorithm for obtaining the digit expansion is:

$$D_\beta : [0, \beta) \rightarrow \{0, 1, \ldots, \lfloor \beta \rfloor\}^\mathbb{N}$$
$$x \mapsto \lfloor x \rfloor \cdot D_\beta\left(\beta(x - \lfloor x \rfloor)\right) \tag{4}$$

C. Singular functions

Definition 1: A function $s(x) : [a, b] \rightarrow \mathbb{R}$ is called a *singular function* if 1) $s(x)$ is continuous, non decreasing on $[a, b]$, and $s(a) < s(b)$; and 2) there exists a set S of Lebesgue

measure 0 such that for all x outside of S the derivative $s'(x)$ exists and is 0.

A folklore name for these functions is "devil's slippery staircases": they are flat almost everywhere, they are continuous everywhere, yet somehow they manage to grow. If $s(a) = 0$ and $s(b) = 1$, then the singular function can be interpreted as the cdf of a random variable. However, note that in the elementary theory, a random variable is assumed to be either "continuous", with differentiable cdf, or "discrete", with constant cdf with steps, or a mix of the two; none of those can have a singular cdf.

D. The Cantor ternary set and Cantor function

The Cantor set C_0 is created by removing the open middle third $(1/3, 2/3)$ from the segment $[0, 1]$ and recursively repeating this operation. These are the first iterations for creating C_0:

By considering the ternary representation of a point $x \in [0, 1]$ then x belongs to C_0 if and only if there are no "1"s in his ternary representation. The Cantor set is a fractal: it is a compact, uncountable set with Lebesgue measure 0, and it does not contain any interval of finite length. Its Hausdorff dimension is $\log(2)/\log(3)$. It is homeomorphic to the set $\{0, 1\}^{\mathbb{N}}$ defined earlier. The Cantor function $C : C_0 \to \mathbb{R}$ is defined on the Cantor set by looking at the ternary representation a_k and translating it as a binary representation. In formulas:

$$C\left(\sum_{k=1}^{\infty} \frac{a_k}{3^k}\right) = \sum_{k=1}^{\infty} \frac{a_k/2}{2^k} \tag{5}$$

C can be thought as the cdf of C_0: when $C = 0.5$, half of C_0 is on the right, and half on the left (but note that the Lebesgue measure of either side is 0); C is an example of a singular function. The cdf for $\mathbf{Y}$ corresponds to the Cantor function in the scalar case for a particular choice of the parameters (Fig. 2); this will be proved as Proposition 3.

E. Iterated Function Systems

Definition 2: Consider a complete metric space S and a finite alphabet E. An *Iterated Function System* (IFS) *with probabilities* on S is a finite list of functions $f_e : S \to S$, $e \in E$; to each function is associated a probability p_e such that $p_e > 0$ and $\sum p_e = 1$. An IFS is *contracting* if all f_e are contractions. The *attractor* (or invariant measure) of an IFS is a probability measure $\mathcal{M}$ such that $\mathcal{M} = \sum_{e \in E} p_e f_{e*}(\mathcal{M})$, where $f_{e*}(\mathcal{M})$ is the push-forward measure of $\mathcal{M}$ through f_e. Equivalently, for all Borel sets A, the attractor satisfies:

$$\mathcal{M}(A) = \sum_{e \in E} p_e \mathcal{M}\left(f_e^{-1}[A]\right) \tag{6}$$

Proposition 2: A contracting IFS with probabilities has a unique tight probability measure as an attractor.

This is the basic result needed for the next section, but IFSs show convergence behavior with much weaker hypotheses: the functions f_e do not have to be *all* contractions on the *whole* space, but they must be contracting "on average", in a sense

which is made precise in [13]; the application of the functions can be governed by a Markov chain [14]; and the fact that the probabilities might be place-dependent [13] allows augmenting the state to use a Hidden Markov Model.

III. THE SCALAR CASE

Consider the scalar system, with $\mathbf{A} = A > 1$, $\mathcal{I} = \mathcal{I} > 0$, and the particular case $\mathbf{Q} = 0$. The two maps h, g become:

$$h : Y \mapsto A^{-2}Y \qquad g : Y \mapsto A^{-2}Y + \mathcal{I}$$

If the packet drops are independent, these relations can be written as a LTI system $Y(k+1) = A^{-2}Y(k) + \mathcal{I}u(k)$ forced by Bernoulli noise. To compute the stationary distribution, let $b \in \{0, 1\}^{\mathbb{N}}$ be the infinite string describing the availability of observations. It is convenient to let b_0 be the last observation, b_1 the next-to-last, and so on. The current value of Y is given by $\varphi(b)$, defined as

$$\varphi : \{0, 1\}^{\mathbb{N}} \to [0, \mathcal{I}] \qquad \varphi : b \mapsto \mathcal{I}\sum_{i=0}^{\infty} \frac{b_i}{(A^2)^i} \tag{7}$$

What is the distribution of $\varphi(b)$? It is possible to give a complete answer in the case $A^2 \geq 2$, while the matter is delicate in the case $A^2 \in [1, 2)$. The following subsections are meant to build up intuition to the general case of Section IV, where all these arguments will be repeated, in a more precise and technical fashion, for the matrix case.

A. The case $A^2 \in [1, 2)$: Bernoulli convolutions

In a 1939 paper [15], Erdős studied the distribution of the infinite sum $\sum \pm \lambda^n$ where the signs + and - are chosen independently with probability $1/2$. (The function φ can be written in this form by appropriate shifting and scaling, and the relation between λ and the parameter A^2 is $\lambda = A^{-2}$). Mathematicians have been working on characterizing the properties of these "Bernoulli convolutions" since then [16]. Erdős showed that the cdf is singular for $\lambda < 1/2$, it is singular for some $\lambda \in [1/2, 1]$, and conjectured that it is *not* singular for almost all $\lambda \in [1/2, 1]$. This conjecture was eventually proved only in 1995 [17]. The case of a biased convolution, where the signs $\pm$ are chosen with a different probability (that is, $d \neq 1/2$) is not completely solved [18]; the cases of generic Markov chains or Hidden Markov Models have not been considered yet. Moreover, even in the case when the cdf is absolutely continuous, there is no known closed form for it.

B. The case $A^2 \geq 2$

1) The case $A^2 = 2$, $Q = 0$: Consider a point $y \in [0, \mathcal{I}]$: to compute the cdf at $F(y)$ one must find the probability measure of the set of strings such that $\varphi(b) \leq y$:

$$F(y) \triangleq \mathcal{M}(\{b \mid \varphi(b) \leq y\}) \tag{8}$$

Consider the binary expansion of y as $y = \mathcal{I}\sum m_i 2^{-i}$, with $m = D_2(\mathcal{I}^{-1}y)$ given by the Rényi expansion in (4). Substituting this for y in (8), one obtains:

$$F(y) = F\left(\mathcal{I}\sum_{i=0}^{\infty} \frac{m_i}{2^i}\right) = \mathcal{M}\left(\left\{b \mid \sum_{i=0}^{\infty} \frac{b_i}{2^i} \leq \sum_{i=0}^{\infty} \frac{m_i}{2^i}\right\}\right) \tag{9}$$

But note that $\sum b_i 2^{-i} \le \sum m_i 2^{-i}$ if and only if $b \trianglelefteq m$. Hence the cdf at y can be found by considering the measure of a certain set on the symbolic space:

$$F(y) = \mathcal{M}(\{b \mid b \trianglelefteq D_2(\mathcal{I}^{-1}y)\}) \qquad (10)$$

This is the main concept used in the paper: to find the cdf in the space of Y one must somehow go back to the space of sequences, where the measure $\mathcal{M}$, describing the packet drops, is defined. Evaluating (10) is easy once the set $\{0,1\}^{\mathbb{N}}$ is arranged in a tree (Fig 3a), starting from the root which symbolizes the empty string. To compute $\mathcal{M}(\{b \mid b \trianglelefteq m\})$, start at the root of the tree, and compute the sum of the probabilities associated to the strings to the "left" of m. For each digit $m_i = 1$, add the probability of the corresponding cylinder on the left. Assuming independent packet drops, the probability associated to the cylinder $[m_0, m_1, \ldots, m_{i-1}, 0]$ is $d\prod_{j=0}^{i-1} d^{(1-m_j)}(1-d)^{m_j}$. The overall closed-form expression for $F(y)$ is

$$F(y) = F\left(\mathcal{I}\sum_{i=0}^{\infty} \frac{m_i}{2^i}\right) = \sum_{i=0}^{\infty} m_i d \prod_{j=0}^{i-1} d^{(1-m_j)}(1-d)^{m_j}$$

2) The case $A^2 = 3$, $Q = 0$: Before going on to the case of generic A, it is instructive to consider the case $A = \sqrt{3}$ to build up the intuition, and because in the case $d = 1/2$, F is a scaled and translated version of the Cantor function in (5). The map φ is now $\varphi : b \mapsto \mathcal{I}\sum_{i=0}^{\infty} b_i 3^{-i}$. For evaluating $F(y)$, consider the 3-expansion $m = D_3(x)$, with $m \in \{0,1,2\}^{\mathbb{N}}$. Rewrite (9) as

$$F(y) = F\left(\mathcal{I}\sum_{i=0}^{\infty} \frac{m_i}{3^i}\right) = \mathcal{M}\left(\left\{b \mid \sum_{i=0}^{\infty} \frac{b_i}{3^i} \le \sum_{i=0}^{\infty} \frac{m_i}{3^i}\right\}\right)$$

Now it is clear that there are two different symbolic spaces being considered: b belongs to $\{0,1\}^{\mathbb{N}}$ while m belongs to $\{0,1,2\}^{\mathbb{N}}$. Nevertheless, b can be considered an element of $\{0,1,2\}^{\mathbb{N}}$ via the standard inclusion mapping i. Formally,

$$F(y) = i_*\mathcal{M}(\{b \mid b \trianglelefteq D_3(\mathcal{I}^{-1}y)\})$$

$\{x \mid x \triangleleft 101\cdots\}$
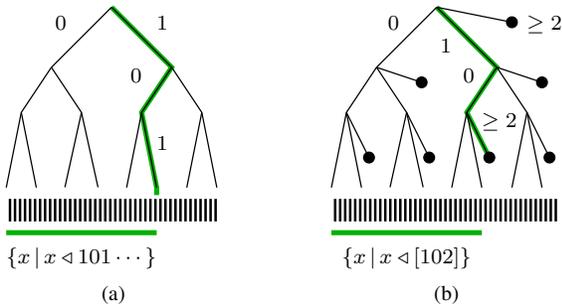
(a)

$\{x \mid x \triangleleft [102]\}$

(b)

Figure 3. Elements of symbolic spaces can be represented by the branches of an infinite tree whose root is the empty string. The lexicographical order "$\triangleleft$" indicates branches "on the left" of a particular path. Hence to evaluate the probability of the event $\{x \mid x \triangleleft y\}$ one walks down the path indicated by y and computes the sum of the probability mass associated to the branches on the left as indicated by the measure $\mathcal{M}$ (see Section III-B). In (a), the measure $\mathcal{M}$ is given on $\{0,1\}^{\mathbb{N}}$ and also y belongs to $\{0,1\}^{\mathbb{N}}$. In (b), x belongs to a larger set $\{0,1,2,\ldots,n\}^{\mathbb{N}}$. In this case, $\mathcal{M}$ is zero on cylinders containing "2" or higher digits, therefore the counting can stop at the first ≥ 2 digit of x.

where now $i_*\mathcal{M}$ is the push-forward of the measure $\mathcal{M}$ to $\{0,1,2\}^{\mathbb{N}}$.

The same reasoning with the tree can be repeated, with the difference that it represents $\{0,1,2\}^{\mathbb{N}}$, and the branches containing a 2 have zero weight. For each digit of m, the case analysis is as follows: if $m_i = 0$, there are no strings on the left; if $m_i = 1$, there is, as before, a branch whose weight is d; if $m_i = 2$, there are two branches on the left, with a weight $d + (1-d) = 1$. Moreover, the computation can be stopped at the first "2" digit, because there are no possible branches corresponding to the set $\{0,1\}^{\mathbb{N}}$ below it. In formulas:

$$F\left(\mathcal{I}\sum_{i=0}^{\infty} \frac{m_i}{3^i}\right) = \sum_{i=0}^{m_i=2} g(m_i) \prod_{j=0}^{i-1} d^{(1-m_j)}(1-d)^{m_j} \quad (11)$$

$$g(m_i) = \begin{cases} 0 & m_i = 0 \\ d & m_i = 1 \\ 1 & m_i = 2 \end{cases} \qquad (12)$$

Proposition 3: For $A^2 = 3$, $\mathcal{I} > 0$, $Q = 0$, $d = 1/2$, the cdf for Y is a scaled version of the Cantor function:

$$F(y) = C\left(\mathcal{I}^{-1}\frac{2}{3}y\right)$$

Proof: Recall that the definition of the Cantor function that we introduced is defined only on the points in the Cantor set. Let $z \in [0,1]$ be a point in the Cantor set; one has to show that $C(z) = F\left(\frac{3}{2}\mathcal{I}z\right)$. According to (11), write $\frac{3}{2}\mathcal{I}z$ as $\mathcal{I}\sum m_i 3^{-i}$, with m being a ternary expansion of $(3/2)z$. Let q be the base-3 representation of z. Multiplying z by 3 means shifting the digits of q one place to left. Division by 2 can be carried on digit-by-digit as well: because z is the Cantor set, there are no 1s in the expansion: a 0 remains a 0, a 2 becomes a 1. Hence $m_i = q_{i+1}/2$: note there are no 2 digits in m. Now notice that for $d = 1/2$, $d^{(1-m_j)}(1-d)^{m_j} = 1/2$, and (11) is simplified as follows:

$$F\left(\frac{3}{2}\mathcal{I}z\right) = F\left(\mathcal{I}\sum_{i=0}^{\infty} \frac{m_i}{3^i}\right) = \sum_{i=0}^{\infty} g(m_i)\frac{1}{2^i}$$

Because there are only 0 and 1 digits in the expansion, the sum goes up to infinity. Also, (12) can be simplified as $g(m_i) = m_i/2$. Finally:

$$F\left(\frac{3}{2}\mathcal{I}z\right) = \sum_{i=0}^{\infty} \frac{m_i}{2}\frac{1}{2^i} = \sum_{i=0}^{\infty} \frac{q_{i+1}}{4}\frac{1}{2^i} = \sum_{j=1}^{\infty} \frac{q_j/2}{2^j} = C(y)$$

$\blacksquare$

Note 1: The previous result implies that the stationary distribution of Y has support in a Cantor-like set, which is a set of Lebesgue measure 0. Note, however, that for all finite times k, the value of $Y(k)$ depends on the initial conditions of the filter, and thus can assume values outside the Cantor-like set.

3) The case $A^2 > 2$, $Q = 0$: It is easy to generalize the previous reasoning for every $A^2 > 2$, by using a β-expansion (Proposition 1) and $\beta = A^2$. The tree will now have $\lfloor \beta \rfloor + 1$ branches, and the formula (11) is generalized as follows:
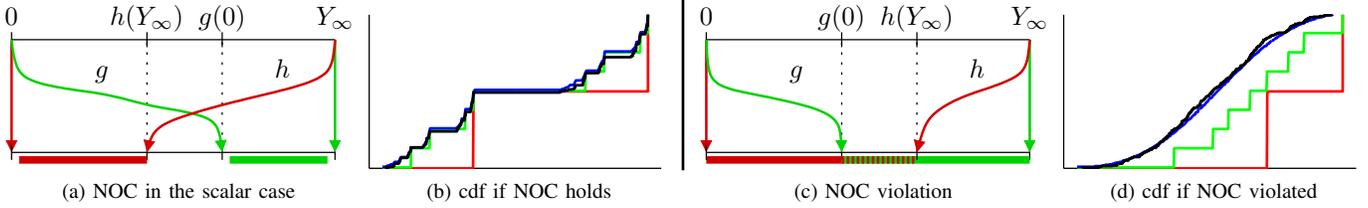
Figure 4. For a scalar system, the NOC is satisfied if and only if the images of h and g do not intersect, as in (a). In that case, the density is a fractal and the cdf is a singular function, as shown in (b). If the NOC is not satisfied, as in (c), then the density may be non-fractal. The lightly-colored graphs in (b) and (d) are the first iterations of the numerical computation of $F(y)$ according to the fixed-point method discussed in Section (III-C), while the black plot is the sample cdf.

$$F(y) = \sum_{i=0}^{m_i \geq 2} g(m_i) \prod_{j=0}^{i-1} d^{(1-m_j)}(1-d)^{m_j} \qquad (13)$$

with $m = D_\beta(\mathcal{I}^{-1}y)$, $\beta = A^2$, and

$$g(m_i) = \begin{cases} 0 & m_i = 0 \\ d & m_i = 1 \\ 1 & m_i \in [2, \lfloor \beta \rfloor] \end{cases} \qquad (14)$$

With this definition, it is easy to show that F is a singular function. F is continuous because D_β is continuous; and moreover:

Proposition 4: For $A^2 > 2$, F is differentiable almost anywhere with derivative 0.

Proof: Pick a point at random and consider its β-expansion. Almost all points have a digit $s \geq 2$ in their expansion, and a 1 somewhere after that. Hence the typical point x has representation $p_1 \cdot s \cdot p_2 \cdot 1 \cdot p_3$. Now consider the points x_a and x_b with representation $p_1 \cdot s \cdot \overline{0}$ and $p_1 \cdot s \cdot \overline{\lfloor \beta \rfloor}$; obviously, $x_a < x < x_b$. In such interval, F is constant, because the iteration of the formula for F stops at the digit s. Hence it is differentiable at x with vanishing derivative. ∎

### C. General case: numerical computation

In the case of independent packet drops, it is possible to compute the cdf $F(y)$ numerically, by computing the fixed point of (6). An example of the iteration is shown in Fig. 4b–4d. It is easy to show convergence in norm of such iterations, however, they cannot be converted in a closed-form expression, and, in general, they do not provide much intuition about the properties of the distribution.

IV. THE MATRIX CASE

The goal of this section is to obtain a closed-form expression for $\mathbb{P}(\mathbf{Y} \prec \mathbf{Z})$, for any matrix $\mathbf{Z}$, in the case that the NOC holds. This is equivalent to the case $A^2 > 2$ in the previous section. Of course, everything is suddenly more complicated and the presentation must be technical to be precise; hopefully, the previous section helped with providing the necessary intuition.

This section deals with several spaces and various transformations between them. $\mathcal{P}_*$ is considered as a metric space with the Riemannian distance, and as a lattice induced by the partial order $\succ$. The space $\{0,1\}^{\mathbb{N}}$ is still used for representing the packet dropping sequences, with the adequate measure $\mathcal{M}$ defined on its cylinders. Define also the space $\{0,1\}^{\mathbb{N}}_*$ as $\{0,1\}^{\mathbb{N}}$ with the sequences which end with an infinite number of 0s or 1s removed: these are the allowed sequences. Also the space $\{0,1,2\}^{\mathbb{N}}$ appears in a support role. The maps that will be defined on these spaces are summarized in the following figure.
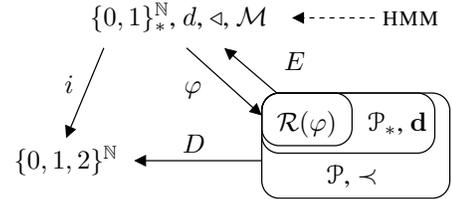
The first step is defining the metric properties of $\mathcal{P}_*$.

Lemma 1: Let $\mathbf{d}(\mathbf{Y}_1, \mathbf{Y}_2)$ be the Riemannian distance on $\mathcal{P}_*$, defined as $\mathbf{d}(\mathbf{Y}_1, \mathbf{Y}_2) = \left[\sum_{i=1}^{n} \log^2(\lambda_i)\right]^{1/2}$, with λ_i being the eigenvalues of the matrix $\mathbf{Y}_1 \mathbf{Y}_2^{-1}$. Then:

1) $(\mathcal{P}_*, \mathbf{d})$ is a complete metric space.
2) $\mathbf{d}$ is invariant to conjugacy. For any invertible matrix $\mathbf{A}$:

$$\mathbf{d}(\mathbf{A}\mathbf{Y}_1\mathbf{A}^T, \mathbf{A}\mathbf{Y}_2\mathbf{A}^T) = \mathbf{d}(\mathbf{Y}_1, \mathbf{Y}_2)$$

3) $\mathbf{d}$ is invariant to inversion:

$$\mathbf{d}(\mathbf{Y}_1^{-1}, \mathbf{Y}_2^{-1}) = \mathbf{d}(\mathbf{Y}_1, \mathbf{Y}_2)$$

4) For any two matrices $\mathbf{Y}_1, \mathbf{Y}_2$ in $\mathcal{P}_*$, and for any $\mathbf{Q} \in \mathcal{P}$,

$$\mathbf{d}(\mathbf{Y}_1 + \mathbf{Q}, \mathbf{Y}_2 + \mathbf{Q}) \leq \frac{\alpha}{\alpha + \beta} \mathbf{d}(\mathbf{Y}_1, \mathbf{Y}_2)$$

where $\alpha = \max\{\|\mathbf{Y}_1\|_2, \|\mathbf{Y}_2\|_2\}$ and $\beta = \inf\{\langle \mathbf{Q}x, x \rangle; \|x\|_2 = 1\}$.

These properties should convince you that the Riemannian distance is the natural one to consider: considering information matrices instead of covariances maintains the same distance. Now look back at the equations (2)-(1): h and g are compositions of three operations: 1) conjugation through $\mathbf{A}$; 2) inversion; and 3) addition of $\mathbf{Q}$ and $\mathcal{I}$. These are all natural operations on $\mathcal{P}_*$ (while the subtraction in the algebraic Riccati equation is not an operation on $\mathcal{P}_*$) and the Riemannian distance is invariant to two of them.

The addition of noise ($\mathbf{Q}$) or information ($\mathcal{I}$) is what makes h and g contractions. But notice h and g are *strict* contractions only if $\mathbf{Q} \succ 0$ and $\mathcal{I} \succ 0$. However, it is possible to prove the following:

Lemma 2: If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible and $(\mathbf{A}, \mathbf{B})$ is controllable, there exists $k \leq n$ such that any iterated composition of the functions h, g of length at least k is a strict contraction mapping.

Proof: See Proposition 1.3 and the case (ii) of Theorem 1.7 in [2]. The step k is the minimum integer such that $[\mathbf{B} \mid \mathbf{A}\mathbf{B} \mid \ \ldots \ \mid \mathbf{A}^{k-1}\mathbf{B} \ ]$ is full rank. ∎

Given this lemma, it is possible to show, using Proposition 2, that there exists an invariant measure in the case of independent packet drops.

Proposition 5: If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible and $(\mathbf{A}, \mathbf{B})$ is controllable, there exists a unique stationary measure for $\mathbf{Y}$ (and $\mathbf{P}$).

Proof: Define a set of 2^k functions corresponding to the possible compositions of g, h of length k, with corresponding probabilities. By Lemma 2, these functions are contractions, therefore they constitute a contracting IFS in the complete metric space $(\mathcal{P}_*, \mathbf{d})$. By Proposition 2 this IFS generates a unique stationary measure. ∎

The uniqueness case in case of arbitrary Hidden Markov Models requires more general theorems than Proposition 2 and a construction for which there is no space here; note, however, that the closed-form for the measure in Theorem 1 below is also a constructive proof of uniqueness.

As already pointed out by [2], considering these contraction properties leads to elementary proofs for the convergence of the regular Kalman filter (one just needs to note that g^n is a contraction).

Towards the computation of the cdf of the stationary distribution, define the function φ that takes the infinite arrival sequence as input and produces the current information matrix; φ is the analogous of (7).

Definition 3: Define the function $\varphi : \{0,1\}_*^{\mathbb{N}} \to \mathcal{P}_*$ as

$$\begin{cases} \varphi(0 \cdot s) & = h(\varphi(s)) \\ \varphi(1 \cdot s) & = g(\varphi(s)) \end{cases} \tag{15}$$

It might seem fishy to define a function in such a recursive way. However, this definition is well-posed in the sense that there exists only one φ which satisfies these two conditions, according to the String Model theorem in [10] (the same construction of the 2^k mappings used in the proof of Proposition 5 must be used). The same source proves the following.

Lemma 3: φ is continuous.

The next step is defining the inverse of the function φ. This is the matrix analogous of a β-expansion (compare Proposition 1). From now on, assume that the NOC holds.

Definition 4: Let $\mathcal{R}(\varphi) \subset \mathcal{P}_*$ be the image of φ. Let $\mathbf{Y}_\infty \notin \mathcal{R}(\varphi)$ be the fixed point of g, and let $\mathbf{Y}_0 \notin \mathcal{R}(\varphi)$ be the fixed point of h.

Lemma 4: For every $\mathbf{Y} \in \mathcal{R}(\varphi)$, $\mathbf{Y} \succ \mathbf{Y}_0$ and $\mathbf{Y} \prec \mathbf{Y}_\infty$. Either $\mathbf{Y} \succ g(\mathbf{Y}_0)$ or $\mathbf{Y} \prec h(\mathbf{Y}_\infty)$.

Definition 5: Define the function $E : \mathcal{R}(\varphi) \to \{0,1\}^{\mathbb{N}}$ as

$$E : \mathbf{Y} \mapsto \begin{cases} 0 \cdot E(h^{-1}(\mathbf{Y})) & \mathbf{Y} \prec h(\mathbf{Y}_\infty) \\ 1 \cdot E(g^{-1}(\mathbf{Y})) & \mathbf{Y} \succ g(\mathbf{Y}_0) \end{cases}$$

Note that this definition is well posed, as for every $\mathbf{Y} \in \mathcal{R}(\varphi)$, either $\mathbf{Y} \prec g(\mathbf{Y}_0)$ xor $\mathbf{Y} \succ h(\mathbf{Y}_\infty)$.

Proposition 6: E is continuous.

Proof: E is continuous if the preimage of open sets are open sets. Open sets for $\{0,1\}^{\mathbb{N}}$ are the cylinders: consider for example the cylinder $c = [010]$. The preimage of c is the set of $\mathbf{Y}$ such that $\mathbf{Y} \prec g(\mathbf{Y}_0)$, $h^{-1}(\mathbf{Y}) \succ h(\mathbf{Y}_\infty)$, $g^{-1} \cdot h^{-1}(\mathbf{Y}) \prec g(\mathbf{Y}_0)$. That is, $E^{-1}(c)$ is the intersection of a finite number (equal to the length of the cylinder) of open sets, and therefore is open. ∎

Proposition 7: $E \circ \varphi = \mathrm{Id}$

Proof: Assume $x \in \{0,1\}^{\mathbb{N}}$ such that $x = 0 \cdot s$. Assume by contradiction that $E(\varphi(0 \cdot s)) = 1 \cdot t$. If E assigned the first digit 1, then $\varphi(0 \cdot s) = h(\varphi(s)) \succ h(\mathbf{Y}_\infty)$ which implies $\varphi(s) \succ \mathbf{Y}_\infty$ (contradiction). A similar reasoning with 0 and 1 exchanged allows to conclude that x and $E(\varphi(x))$ agree on the first digit: $E(\varphi(d \cdot r)) = d \cdot E(\varphi(r))$. By induction, they agree on the whole sequence. ∎

Proposition 8: $\mathcal{R}(\varphi)$ is totally disconnected and nowhere dense.

Proof: The previous three propositions imply that $\mathcal{R}(\varphi)$ is homeomorphic to $\{0,1\}_*^{\mathbb{N}}$ which is a subset of the Cantor set which is totally disconnected and nowhere dense. Note however that $\mathcal{R}(\varphi)$ is not closed as it lacks accumulation points such as $\mathbf{Y}_\infty$ and $\mathbf{Y}_0$. ∎

The following propositions establish that the lexicographical order on the arrival sequences is compatible with the Loewner order on matrices through the mapping φ.

Proposition 9: For all $x, y \in \{0,1\}_*^{\mathbb{N}}$:

$$x \triangleleft y \quad \Leftrightarrow \quad \varphi(x) \prec \varphi(y)$$

Proof: ($\Rightarrow$) Assume that $x \triangleleft y$: then $x = p \cdot 0 \cdot q$ and $y = p \cdot 1 \cdot r$. Then $(\varphi(x) \prec \varphi(y)) \Leftrightarrow \left(\varphi_{(p)}^{-1} \circ \varphi(x) \prec \varphi_{(p)}^{-1} \circ \varphi(y) \right) \Leftrightarrow (\varphi(0 \cdot q) \prec \varphi(1 \cdot r)) \Leftrightarrow (h(\varphi(q)) \prec g(\varphi(r)))$ which is always true due to the NOC. ($\Leftarrow$) Note that $\varphi(x) \prec \varphi(y)$ implies $x \neq y$. Because "$\triangleleft$" is a strict total order, either $x \triangleleft y$ or $y \triangleleft x$. But the latter would imply (given the first part of this proposition) that $\varphi(x) \succ \varphi(y)$ which would be a contradiction. ∎

Proposition 10: "$\prec$" is a strict total order on $\mathcal{R}(\varphi)$.

Proof: Apply the previous proposition after noticing that each couple of elements $\mathbf{Y}_1 \neq \mathbf{Y}_2$ in $\mathcal{R}(\varphi)$ can be written as $\varphi(x)$ and $\varphi(y)$ with $x = E(\mathbf{Y}_1)$ and $y = E(\mathbf{Y}_2)$. ∎

The consequence of $\prec$ being a strict total order is that every $\mathbf{Y} \in \mathcal{R}(\varphi)$ is comparable with every other; hence $\mathbb{P}(\{\mathbf{Y} \prec \overline{\mathbf{Y}}\}) + \mathbb{P}(\{\mathbf{Y} \succ \overline{\mathbf{Y}}\}) = 1$, and therefore this is a way to easily describe the cdf in a multi-dimensional space (and in a coordinate-free manner). At this point, all would be ready for writing a closed form expression for $\mathbb{P}(\{\mathbf{Y} \prec \overline{\mathbf{Y}}\})$, with $\overline{\mathbf{Y}} \in \mathcal{R}(\varphi)$. This is skipped and the following propositions go towards an expression for $\mathbb{P}(\{\mathbf{Y} \prec \mathbf{Z}\})$, with $\mathbf{Z}$ possibly not in $\mathcal{R}(\varphi)$.

Definition 6: Define the function $D : \mathcal{P} \to \{0,1,2\}^{\mathbb{N}}$ as

$$D : \mathbf{Y} \mapsto \begin{cases} 2 \cdot \overline{0} & \mathbf{Y} \succ \mathbf{Y}_\infty \\ 1 \cdot D(g^{-1}(\mathbf{Y})) & \neg(\mathbf{Y} \succ \mathbf{Y}_\infty) \wedge (\mathbf{Y} \succeq g(\mathbf{Y}_0)) \\ 0 \cdot D(h^{-1}(\mathbf{Y})) & \neg(\mathbf{Y} \succ \mathbf{Y}_\infty) \wedge \neg(\mathbf{Y} \succeq g(\mathbf{Y}_0)) \end{cases} \tag{16}$$

The following trivial lemmas are stated without proof:

Lemma 5: The restriction of D to $\mathcal{R}(\varphi)$ is equal[1] to E.

Lemma 6: $D \circ \varphi = \text{Id}$

Lemma 7: If $\mathbf{Z} \notin \mathcal{R}(\varphi)$, then $D(\mathbf{Z}) = p \cdot 2 \cdot \bar{0}$ for some p.

The following is generalization of Proposition 9 to $\{0, 1, 2\}^{\mathbb{N}}$ and D.

Proposition 11: For $\mathbf{Y} \in \mathcal{R}(\varphi)$ and $\mathbf{Z} \in \mathcal{P}$:

$$D(\mathbf{Y}) \triangleleft D(\mathbf{Z}) \Leftrightarrow \mathbf{Y} \prec \mathbf{Z}$$

Proof: ($\Leftarrow$) Assume $\mathbf{Y} \prec \mathbf{Z}$. Note that $D(\mathbf{Y}) \neq D(\mathbf{Z})$ because the latter contains at least a 2 (by Lemma 7). Because $\triangleleft$ is a strict total order on $\{0, 1, 2\}^{\mathbb{N}}$, to show $D(\mathbf{Y}) \triangleleft D(\mathbf{Z})$ it is sufficient to assume $D(\mathbf{Y}) \triangleright D(\mathbf{Z})$ and arrive at a contradiction. If $D(\mathbf{Y}) \triangleright D(\mathbf{Z})$, then $D(\mathbf{Y}) = p \cdot 1 \cdot r$ (because $\mathbf{Y} \in \mathcal{R}(\varphi)$, $D(\mathbf{Y})$ cannot contain any 2 by Lemma 5) and $D(\mathbf{Z}) = p \cdot 0 \cdot s$ for some p, r, s. Let $\mathbf{Y}' = \varphi_{(p)}^{-1}(\mathbf{Y})$ and $\mathbf{Z}' = \varphi_{(p)}^{-1}(\mathbf{Z})$. Then $D(\mathbf{Y}') = 1 \cdot r$ and $D(\mathbf{Z}') = 0 \cdot s$. From the analysis of (16) one concludes that $\mathbf{Y}' \succeq g(\mathbf{Y}_0)$ and $\neg(\mathbf{Z} \succeq g(\mathbf{Y}_0))$, but the assumption $\mathbf{Y} \prec \mathbf{Z}$ leads to a contradiction. ($\Rightarrow$) Assume $D(\mathbf{Y}) \triangleleft D(\mathbf{Z})$ such that $D(\mathbf{Y}) = p \cdot 0 \cdot r$ and $D(\mathbf{Z}) = p \cdot 1 \cdot r$. Similarly as before, let $\mathbf{Y}' = \varphi_{(p)}^{-1}(\mathbf{Y})$ and $\mathbf{Z}' = \varphi_{(p)}^{-1}(\mathbf{Z})$. Because D assigns a 0 for the first digit of $\mathbf{Y}'$, then $\neg(\mathbf{Y}' \succeq g(\mathbf{Y}_0))$. Because $\mathbf{Y}' \in \mathcal{R}(\varphi)$ (this is a necessary condition), by Lemma 4 $\mathbf{Y}' \prec h(\mathbf{Y}_\infty)$. Because D assigns a 1 to $\mathbf{Z}'$, $\mathbf{Z}' \succeq g(\mathbf{Y}_0) \succ h(\mathbf{Y}_\infty) \succ \mathbf{Y}'$ which implies the thesis. ∎

Theorem 1: For any matrix $\mathbf{Z} \in \mathcal{P}$,

$$\mathbb{P}(\{\mathbf{Y} \prec \mathbf{Z}\}) = i_* \mathcal{M}(\{x \mid x \triangleleft D(\mathbf{Z})\}),$$

which, by letting $m = D(\mathbf{Z})$, can be computed as follows:

$$\mathbb{P}(\{\mathbf{Y} \prec \mathbf{Z}\}) = \sum_{i=0}^{\infty} \begin{cases} 0 & m_i = 0 \\ \mathcal{M}([m_{0:i-1} \cdot 0]) & m_i = 1 \\ \mathcal{M}([m_{0:i-1}]) & m_i = 2 \end{cases} \quad (17)$$

Moreover, $\mathbb{P}(\mathbf{Y} \prec \mathbf{Z})$ is a singular function of $\mathbf{Z}$.

Proof: The proof builds on the previous results:

$$\begin{aligned} \mathbb{P}(\{\mathbf{Y} \prec \mathbf{Z}\}) &\triangleq \mathbb{P}(\{x \mid \varphi(x) \prec \mathbf{Z}\}) \\ \{\text{Prop. 11} \Rightarrow\} &= \mathbb{P}(\{x \mid D(\varphi(x)) \triangleleft D(\mathbf{Z})\}) \\ \{\text{Lemma 6} \Rightarrow\} &= \mathbb{P}(\{x \mid x \triangleleft D(\mathbf{Z})\}) \\ &= i_* \mathcal{M}(\{x \mid x \triangleleft D(\mathbf{Z})\}) \end{aligned}$$

$\mathbb{P}(\mathbf{Y} \prec \mathbf{Z})$ is a singular function because $\mathcal{R}(\varphi)$ is a totally disconnected set (Proposition 8). The final expression in terms of $\mathcal{M}$ can be found by a generalization of the tree-walking reasoning used for deriving (13). ∎

This result takes into account arbitrary generation models for the packet drops. Note also that, if one wants to compute $\mathbb{P}(\{\mathbf{Y} \prec \mathbf{Z}\})$, it is not necessary to compute all the digits of $m = D(\mathbf{Z})$: for example, in the Bernoulli case, it is easy to see that (17) has exponential convergence in the number of digits, and the recursive function D (equation (16)) can be used to compute only the digits which are needed.

V. CONCLUSIONS AND FUTURE WORK

This paper provided a geometric analysis of the stationary distribution of the covariance for the Kalman filter with intermittent observations, from a geometric point of view. If a certain non-overlapping condition (NOC) holds, the support of the stationary distribution of the *a posteriori* covariance is a one-dimensional fractal curve residing in the space of positive semidefinite matrices, and provided a means to compute the cumulative distribution function in closed form, improving on previous work that only provided upper and lower bounds. The NOC depends only on the parameters of the system ($\mathbf{A}$, $\mathcal{I}$, $\mathbf{Q}$) and not on the statistics of the observation availability (the measure $\mathcal{M}$). It is a fairly strong assumption as it implies that $\mathbf{C}$ has full rank. Future work consists in trying to relax the NOC, perhaps by considering the weaker condition that the images of $\{\mathbf{Y} \mid \mathbf{0} \preceq \mathbf{Y} \preceq \mathbf{P}_\infty^{-1}\}$ through g and h are disjoint.

REFERENCES

[1] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1453–1464, Sept. 2004.

[2] P. Bougerol, "Kalman filtering with random coefficients and contractions," *SIAM J. Control and Optimization*, vol. 31, no. 4, 1993.

[3] M. Huang and S. Dey, "Stability of Kalman filtering with Markovian packet losses," *Automatica*, vol. 43, no. 4, pp. 598 – 607, 2007.

[4] Y. Mo and B. Sinopoli, "A characterization of the critical value for Kalman filtering with intermittent observations," in *IEEE Conference on Decision and Control (CDC)*, pp. 2692–2697, Dec. 2008.

[5] K. Plarre and F. Bullo, "On Kalman filtering for detectable systems with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 54, pp. 386–390, Feb. 2009.

[6] L. Shi, M. Epstein, A. Tiwari, and R. M. Murray, "Estimation with information loss: Asymptotic analysis and error bounds," in *IEEE Conference on Decision and Control*, 2005.

[7] M. Epstein, L. Shi, A. Tiwari, and R. M. Murray, "Probabilistic performance of state estimation across a lossy network," *Automatica*, vol. 44, no. 12, pp. 3046 – 3053, 2008.

[8] B. Mandelbrot, *Gaussian self-affinity and fractals*. Springer, 2001.

[9] G. A. Edgar, *Integral, probability, and fractal measures*. Springer, 1998.

[10] G. A. Edgar, *Measure, topology and fractal geometry*. Springer, 1995.

[11] A. Rényi, "Representations for real numbers and their ergodic properties.," *Acta Math. Acad. Sci. Hungar.*, vol. 8, pp. 477–493, 1957.

[12] W. Parry, "On the β-expansions of real numbers," *Acta Math. Acad. Sci. Hungar.*, vol. 11, pp. 401–416, 1957.

[13] M. F. Barnsley, S. G. Demko, J. H. Elton, and J. S. Geronimo, "Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities," *Annales de l'Institut Henri Poincaré – Probabilité et Statistiques*, vol. 24, no. 3, 1988.

[14] M. F. Barnsley, J. M. Elton, and D. P. Hardin, "Recurrent Iterated Function Systems," *Constructive Approximation*, vol. 5, 1989.

[15] P. Erdős, "On a family of symmetric Bernoulli convolutions," *American Journal of Mathematics*, vol. 61, no. 4, pp. 974–976, 1939.

[16] B. S. Yuval Peres, Wilhelm Schlag, "Sixty years of Bernoulli convolutions," in *Fractals and Stochastics, II, Proceedings of the Greifswald 1998 Conference*, 1998.

[17] B. Solomyak, "On the random series $\sum \pm \lambda^n$ (an Erdős problem)," *Annals of Mathematics*, vol. 142, no. 3, 1995.

[18] Y. Peres and B. Solomyak, "Self-similar measures and intersections of Cantor sets," *Transactions of the American Mathematical Society*, vol. 350, pp. 4065–4087, 1998.

---

[1] E was introduced as an intermediate step because it was expressed with open sets and that was convenient for the proofs of Propositions 6–7.