Kalman filtering with intermittent observations: a geometric approach *

* ACC'09 paper plus other stuff

Andrea Censi

Ph.D. student,

Control & Dynamical Systems,

California Institute of Technology

Advisor: Richard Murray

Linear/Gaussian estimation

Consider the discrete-time linear dynamical system

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

with $\boldsymbol{\omega}(k)$ and $\boldsymbol{v}(k)$ white Gaussian sequences with covariance matrix equal to identity.

- Let $\mathbf{Q} \triangleq \mathbf{B}\mathbf{B}^*$ and $\boldsymbol{\mathcal{I}} \triangleq \mathbf{C}^*\mathbf{C}$. Then the <u>posterior</u> covariance matrix of the error

  $$\mathbf{P}(k) = \operatorname{cov}\left(\hat{\boldsymbol{x}}(k) - \boldsymbol{x}(k) | \boldsymbol{y}(1), \ldots, \boldsymbol{y}(k)\right)$$

  evolves according to the following deterministic map:

  $$g : \mathbf{P} \mapsto \left((\mathbf{A}\mathbf{P}\mathbf{A}^* + \mathbf{Q})^{-1} + \boldsymbol{\mathcal{I}} \right)^{-1}$$

  (a Riccati iteration "in compact form" for the lazy researcher)

- If $(\mathbf{A}, \mathbf{B})$ controllable and $(\mathbf{A}, \mathbf{C})$ detectable, $\mathbf{P}_\infty = \lim_{n\to\infty} g^n(\mathbf{P}(0))$

... with intermittent observations

- [Sinopoli'04] One can model packet drops as follows:

$$\begin{aligned} \boldsymbol{y}'(k) &= \gamma(k)\boldsymbol{y}(k) \\ \gamma(k) &\sim \text{Bernoulli, with probability } \overline{\gamma} \end{aligned}$$

- Evolution of $\mathbf{P}(k)$ as an **Iterated Function System** [Barnsley]**:**
  - execute g if the packet arrives
  - execute h if the packet does not arrive

$$\begin{aligned} g : \mathbf{P} &\mapsto \left((\mathbf{APA}^* + \mathbf{Q})^{-1} + \boldsymbol{\mathcal{I}} \right)^{-1}, & p_g &= \overline{\gamma} \\ h : \mathbf{P} &\mapsto \mathbf{APA}^* + \mathbf{Q}, & p_h &= 1 - \overline{\gamma} \end{aligned}$$

- The iteration of $\mathbf{P}(k)$ is not deterministic: rather than $\mathbf{P}_\infty$, we must speak of the *stationary distribution* of $\mathbf{P}$.
- What is the behavior as a function of the arrival probability $\overline{\gamma}$?

Three contributions

1. **Existence of the stationary distribution:**
   - **Literature:** Stationary distribution exists if $\overline{\gamma} > \overline{\gamma}_s$ [Kar'09].
   - **Contribution**: If $\mathbf{A}$ nonsingular, it always exists.

2. **Mean of the stationary distribution:**
   - **Literature:**
     - $\mathbb{E}\{\mathbf{P}\}$ exists if and only if $\overline{\gamma} > \overline{\gamma}_c$ [Sinopoli'04],
     - $\overline{\gamma}_c$ not precisely characterized yet. [Mo'08, Plarre'09,...].
   - **Contribution**: The *intrinsic Riemannian mean* always exists.

3. **CDF (performance bounds):**
   - **Literature:** Upper and lower bounds on $\mathbb{P}(\{\mathbf{P} \leq \mathbf{M}\})$. [Epstein'05]
   - **Contribution:** if a certain non-overlapping condition holds, then:
     - $p(\mathbf{P})$ has a fractal support.
     - $\mathbb{P}(\{\mathbf{P} \leq \mathbf{M}\})$ can be found in closed form.

A really useful metric

Let $\mathcal{P}$ be the set of positive definite matrices of order n. Define:

$$d(\mathbf{P}_1, \mathbf{P}_2) = \left[\sum_{i=1}^{n} \log^2(\lambda_i(\mathbf{P}_1^{-1}\mathbf{P}_2)) \right]^{1/2}$$

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

$$d(\mathbf{A}\mathbf{P}_1\mathbf{A}^*, \mathbf{A}\mathbf{P}_2\mathbf{A}^*) = d(\mathbf{P}_1, \mathbf{P}_2)$$

3. d is invariant to inversion:

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

4. For any two matrices $\mathbf{P}_1$, $\mathbf{P}_2$ in $\mathcal{P}$, and for any $\mathbf{Q} \geq 0$,

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

   where $\alpha = \max\{\lambda_{\max}(\mathbf{P}_1), \lambda_{\max}(\mathbf{P}_2)\}$ and $\beta = \lambda_{\min}(\mathbf{Q})$.

Contraction properties of g and h

- The maps g, h are compositions of: conjugations, addition, inversion.

$$\begin{aligned} g : \mathbf{P} &\mapsto \left((\mathbf{APA}^* + \mathbf{Q})^{-1} + \mathcal{I} \right)^{-1} \\ h : \mathbf{P} &\mapsto \mathbf{APA}^* + \mathbf{Q} \end{aligned}$$

- Therefore, **they are non-expansive** in this metric: (also h!)

$$\begin{aligned} d(g(\mathbf{P}_1), g(\mathbf{P}_2)) \leq d(\mathbf{P}_1, \mathbf{P}_2) \\ d(h(\mathbf{P}_1), h(\mathbf{P}_2)) \leq d(\mathbf{P}_1, \mathbf{P}_2) \end{aligned}$$

- [Bougerol '93]: If $\mathbf{A}$ nonsingular, $(\mathbf{A}, \mathbf{C})$ observable, $(\mathbf{A}, \mathbf{B})$ controllable, g^n **is a <u>strict</u> contraction**:

$$d(g^n(\mathbf{P}_1), g^n(\mathbf{P}_2)) \leq \rho \, d(\mathbf{P}_1, \mathbf{P}_2), \qquad \rho < 1$$

Existence of stationary distribution

- **Lemma** [Barnsley'88]. An iterated function system $\{f_i\}$ where all f_i are non-expansive, and *at least one is a strict contraction*, admits a unique attractive stationary distribution (convergence in distribution).

- **Proposition:** A stationary distribution for the covariance always exists if $\mathbf{A}$ nonsingular, $(\mathbf{A}, \mathbf{C})$ observable, $(\mathbf{A}, \mathbf{B})$ controllable.

  Proof: After n steps, there are 2^n combinations of g and h.

$$\begin{array}{ll} \overbrace{hhh\cdots h}^{n} & \text{non-expansive} \\ ghh\cdots h & \text{non-expansive} \\ & \vdots \\ hgg\cdots g & \text{non-expansive} \\ ggg\cdots g & \text{strict contraction} \end{array}$$

  Therefore the system satisfies the *average-contractivity* condition.

Existence of Riemannian mean(s)

- Does it not bother you that $\mathbb{E}\{\cdot\}$ is not invariant to change of coordinates?

$$\mathbb{E}\{\mathbf{P}\} \neq \mathbb{E}\{\sqrt{\mathbf{P}}\}^2 \neq \mathbb{E}\{\mathbf{P}^{-1}\}^{-1}$$

  Qualitative/quantitative results should be invariant of parametrizations.

- The manifold of positive definite matrices is not *flat*; we must be careful.
- The expectation can be generalized to a Riemannian manifolds $\mathcal{M}$ with distance d as follows:

$$\mathbb{M}\{X\} \triangleq \arg \inf_{y \in \mathcal{M}} \mathbb{E}\left\{d^2(X,y)\right\}$$

- In our case, different distances give different "critical probabilities":

	covariances	std devs	information	**Riemannian**
$d =$	$\lVert\mathbf{P}_1 - \mathbf{P}_2\rVert_F$	$\lVert\sqrt{\mathbf{P}}_1 - \sqrt{\mathbf{P}}_2\rVert_F$	$\lVert\mathbf{P}_1^{-1} - \mathbf{P}_2^{-1}\rVert_F$	$\left[\sum \log^2(\lambda_i(\mathbf{P}_1^{-1}\mathbf{P}_2)\right]^{\frac{1}{2}}$
$\overline{\gamma}_c^d =$	$\overline{\gamma}_c$	$< \overline{\gamma}_c$	none	none

Which is more natural?

	covariances	std devs	information	**Riemannian**
$d =$	$\|\|\mathbf{P}_1 - \mathbf{P}_2\|\|_F$	$\|\|\sqrt{\mathbf{P}}_1 - \sqrt{\mathbf{P}}_2\|\|_F$	$\|\|\mathbf{P}_1^{-1} - \mathbf{P}_2^{-1}\|\|_F$	$\left[\sum \log^2(\lambda_i(\mathbf{P}_1^{-1}\mathbf{P}_2)\right]^{\frac{1}{2}}$
$\overline{\gamma}_c^d =$	$\overline{\gamma}_c$	$< \overline{\gamma}_c$	none	none

- **Covariances:** Corresponds to the average squared error $\mathbb{E}\{\|\|\boldsymbol{e}\|\|^2\}$.
- **Standard deviations**
  - Corresponds to the average error $\mathbb{E}\{\|\|\boldsymbol{e}\|\|\}$ (physical meaning).
- **Information matrices**
  - $\mathbf{P}^{-1}$ is the natural parametrization for Gaussians.
- **Riemannian distance**
  - Information Geometry interpretation: the natural distance from the Fisher Information Metric on the manifodl of Gaussian distributions.
  - $d(\mathbf{P}_1, \mathbf{P}_2)$ can be linked to the probability of distinguishing the two distributions from the samples [Amari'00].

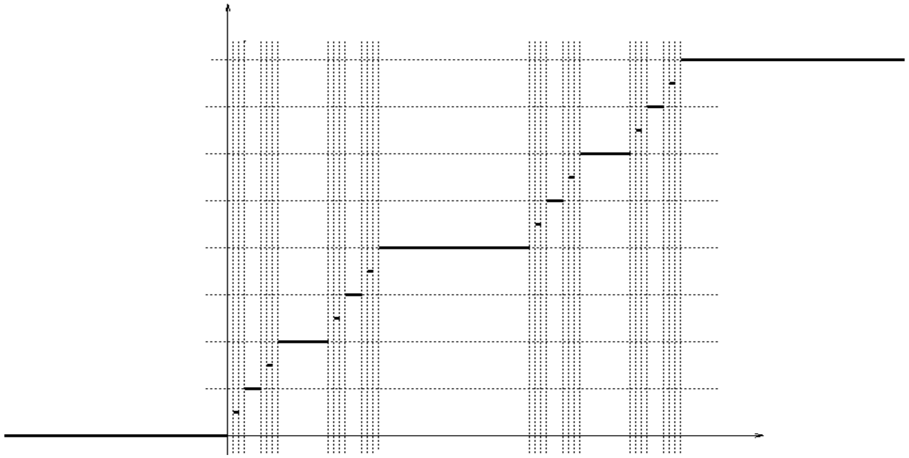
Finally the fractals

Cantor set and Cantor function

- Instructions: take a segment, remove the middle third, repeat.

↑ Cantor Set

The CDF is the Cantor Function →

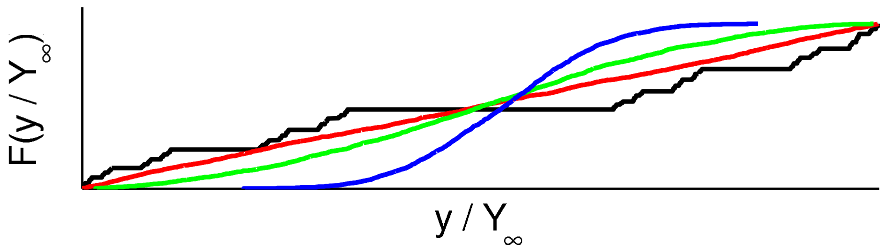


- The Cantor set is a “fractal”:
  - it is totally disconnected
  - it is self-similar
  - it has non-integer dimension
- The Cantor function is a *singular* function (“devil’s staircase”)
  - continuous
  - differentiable almost everywhere, with derivative 0

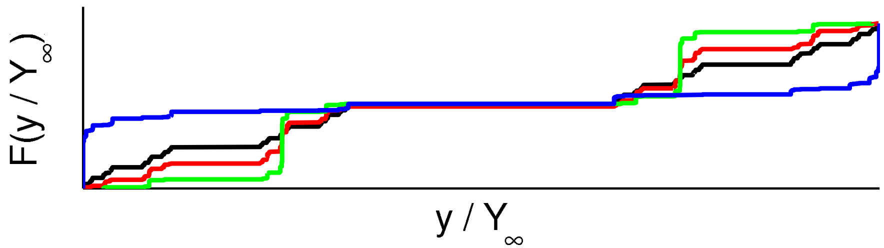
Cantor set

- **Proposition**: For some value of the parameters, the distribution of $\mathbf{P}^{-1}$ is exactly the (scaled/translated) Cantor set.

- Black plot: nominal system $A = 1/\sqrt{3}$, $Q = 0$, $\mathcal{I} = 1$ and packet dropping governed by a Markov chain with transition matrix $T = [\alpha, 1-\alpha; 1-\beta, \beta]$, $\alpha = \beta = 0.5$.



Varying Q:
■ $Q = 0$; ■ $Q = 1$; ■ $Q = 2$; ■ $Q = 3$.



Varying the parameters of the Markov Chain:
■ $\alpha = 0.5$, $\beta = 0.5$; ■ $\alpha = 0.3$, $\beta = 0.3$; ■ $\alpha = 0.1$, $\beta = 0.1$; ■ $\alpha = 0.9$, $\beta = 0.9$.

Fractal properties

- **Proposition:** If h and g have disjoint range, then:
  - the stationary distribution is homeomorphic to the Cantor set,
  - it is a compact, totally disconnected set
  - the cumulative distribution function is a singular function
- If h and g have non-disjoint range, then it **might** be still fractal.
  - This is an open problem in number theory.
- **Proposition:** If h and g satisfy the stronger condition:

$$h(\mathbf{P}_\infty) \geq g(\mathbf{0})$$

then one can find a closed form solution for

$$\mathbb{P}(\{\mathbf{P} \leq \mathbf{M}\}) = \text{"ugly" formula involving the "digit representation" of } \mathbf{M}$$

  - This implies that **C** is invertible (strong condition).
  - This is also valid for Markov Chain driving the packet drops.

Conclusions

Three contributions:

- The stationary distribution always exists.
- The *intrinsic Riemannian mean* always exists.
- $\mathbb{P}(\{\mathbf{P} \leq \mathbf{M}\})$ in closed form with a (strong) non-overlapping condition.

Main ideas:

- Theory of **Iterated Function Systems** — many ready-to-use results in books with very nice fractal illustrations [Barnsley].
- A **useful Riemannian metric** for positive definite matrices — natural Information Geometry metric for manifold of Gaussian distributions
- **Contraction properties of Riccati recursions** in this metric.

Open problems:

- Case of a singular $\mathbf{A}$.
- Computing the Riemannian mean (only proved existence).
- Computing the CDF $\mathbb{P}(\{\mathbf{P} \leq \mathbf{M}\})$ without non-overlapping condition.

Metric spaces properties

- $(\mathcal{P}, d)$ is a *complete metric space*: every Cauchy sequence has a limit in $\mathcal{P}$.
- Fixed Point Theorem: If f strict contraction mapping:

$$\sup_{x,y} \frac{d(f(x), f(y))}{d(x,y)} = q < 1$$

  and complete metric space, then $\lim_{n\to\infty} f^n(x) = x_0$.

Cantor set, more formally

- Let $\{0,1\}^{\mathbb{N}}$ the Cantor space with the following metric:

$$d(x,y) = 2^{-k}; \quad k \text{ first digit for which } x_k \neq y_k$$

  Note that with this metric, $0.11111\cdots \neq 1.0000\ldots$ because $d(0.\overline{1}, 1.\overline{0}) \neq 0$.

- Any set is called "Cantor set" if it is homeomorphic to the Cantor space.

Cantor set, more formally

- Let $\{0,1\}^{\mathbb{N}}$ represent the arrival sequence. You can write the final covariance as:

$$\begin{array}{rcl} \varphi : \{0,1\}^{\mathbb{N}} & \to & \mathcal{P} \\ \text{arrival sequence} & \mapsto & \text{final covariance} \end{array}$$

- If you can prove that φ is invertible, then the range of φ is a Cantor set.
- **Non-overlapping condition:** If the ranges of h and g are non-overlapping, then φ is invertible.
- Morever:

$$\begin{array}{rcl} \varphi : \{0,1\}^{\mathbb{N}} & & \mathcal{P} \\ \text{statistics of packet arrival} & \overset{\varphi}{\mapsto} & \text{probability distributions} \end{array}$$