Robust Orbit Determination: A Machine Learning Approach

Srinagesh Sharma, James Cutler

April 24, 2015

Table of Contents

INTRODUCTION

Introduction

The Following are true regarding state of CubeSat Development

- The number of cubesats per deployment has been steadily increasing in recent years, including cubesats to deep space.
- Autonomous Ground Station Networks with software-defined radios (SDRs) have become popular for decoding.
- Orbit Determination is difficult for large number of satellite deployments, and generally accomplished using linearized estimators and filters.

Introduction

However,

- Orbit Determination using ground stations require
  - Requires an initial estimate in the linear region of the orbit determination system
  - Hard to generalize for arbitrary gravitational potential maps
  - Stochastic parameters need to have "nice" probabilistic assumptions - such as gaussian probability distributions on process noise

- Autonomous orbit determination for multiple CubeSats over ground station networks is relatively unexplored ground

Motivation

With these considerations, would it be possible to

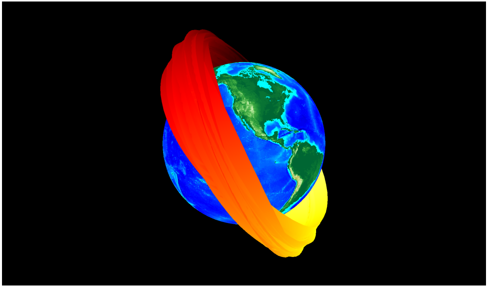
- Identify spacecraft whose communication characteristics are known a-priori
- Perform Orbit Determination whose orbit characteristics do not have nice initial estimates and whose stochastic parameters have distributions which may be non-parametric.
- Generalize such a system to operate over n_G ground stations, over N_P passes.

Essentially, can we build an autonomous system that "learns" to do orbit determination?

PROBLEM STATEMENT

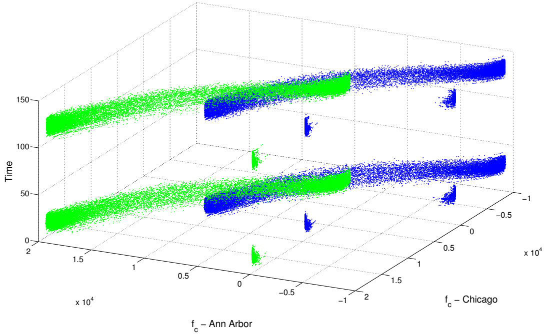
Problem Statement

- Consider Two Spacecraft, labeled $-1, 1$.
- Orbit parameters $\Gamma_j = \{a_j, e_j, \Omega_j, I_j, \omega_j, M_j\}$, $j = -1, 1$.
- Say we know that these orbit parameters are drawn from a probability distribution P_{Γ_j} and that Γ_j are bounded.



Problem Statement

- Spacecraft RF transmissions have certain characteristics such as modulation, center frequency instability, randomness of data, etc, which are known prior to launch.



- Observe transmissions from n_G ground stations over time period T_p, and extract "feature vectors" from the received signal.

Problem Statement

Assumptions

- The received RF signals are distinguishable almost everywhere.
- There exists a model of the system such that given an example orbit, example transmissions of these spacecraft can be generated (for example, a perturbation model and orbit propagator).

Given such a system, for any n_T test feature vectors, can we classify transmissions and estimate the orbits Γ_{-1}, Γ_1?

MACHINE LEARNING APPROACH

Learning Approach

- Generate n training datasets $\{\{x_{ij}, y_{ij}\}_{j=1}^{n_i}, \gamma_{i,-1}, \gamma_{i,1}\}_{i=1}^{n}$
  - $\gamma_{i,-1}$: example orbits of the first spacecraft
  - $\gamma_{i,1}$: example orbits of the second spacecraft
  - x_{ij}: RF Feature Vectors of a spacecraft in the ith orbit
  - y_{ij}: label of x_{ij}
- Training datasets can take into consideration very subtle variations in RF transmission characteristics, drifts in center frequency, variation of transmission rates etc.
- $P_{\Gamma_1}, P_{\Gamma_{-1}}$ induces a probability distribution on the space of probability distributions of the RF Feature Vectors.

Kernel Methods

Kernel Methods have become popular in learning theory due to their ability to estimate non-linear functions.

- $\mathcal{S}$: Compact Space
- $k : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$, a symmetric, positive definite function
- Reproducing Kernel Hilbert Space(RKHS): k generates a function space[1] $\mathcal{H}_k$ such that $\forall f \in \mathcal{H}_k, s \in \mathcal{S}$, $f(s) = \langle f, k(\cdot, s) \rangle$
- When universal kernels are used, the search space is a dense subset of $C(\mathcal{S})$, the space of all continuous bounded functions on $\mathcal{S}$.
- They can also be used to embed probability distributions on $\mathcal{S}$ onto $\mathcal{H}_k$, $\Phi(P_S) = \int_{\mathcal{S}} k(\cdot, s) dP_S$

[1] Nachman Aronszajn. "Theory of reproducing kernels". In: *Transactions of the American mathematical society* (1950), pp. 337–404.

Classification

- Say every feature vector represents only one RF source.
- One of the characteristics of RF transmissions is that the center frequency varies due to Doppler Shift, and therefore their signals can overlap.
- When Transmitting Sources are identical in all aspects except $P_{\Gamma_{-1}}, P_{\Gamma_1}$, even to separate single RF transmissions, we consider all the RF transmissions received over the entire time period (i.e, the probabilistic embedding)
- This leads to **Transfer Learning**[2]
- The sources are separated by a hyperplane in this infinite dimensional space.

[2] Gilles Blanchard, Gyemin Lee, and Clayton Scott. "Generalizing from several related classification tasks to a new unlabeled sample". In: *Advances in neural information processing systems*. 2011, pp. 2178–2186.

Transfer Learning

- For $(\Gamma_{-1}, \Gamma_1) \in (\mathcal{J}^2, \mathcal{F}_{\mathcal{J}}^{\otimes 2})$, $(X, Y) \in (\mathcal{X} \times \mathcal{Y}, \mathcal{F}_{\mathcal{X}} \otimes \mathcal{F}_{\mathcal{Y}})$, A probability distribution P_{Γ_i} on $\Gamma_i, i = -1, 1$, induces λ on the space of probability distributions, $\mathcal{B}_{\mathcal{X}\times\mathcal{Y}}$ such that

$$P^{(i)}(X, Y|\Gamma_1 = \gamma_{i,1}, \Gamma_{-1} = \gamma_{i,-1}) \sim \lambda = (P_{\Gamma_1} \times P_{\Gamma_{-1}}) \circ \mu^{-1}$$

$$x_{ij}, y_{ij} \sim P^{(i)}(X, Y|\Gamma_1 = \gamma_{1,i}, \Gamma_{-1} = \gamma_{-1,i})$$

  μ is a function of Γ_{-1}, Γ_1 and a measure on $\mathcal{B}_{\mathcal{X}\times\mathcal{Y}}$. μ^{-1} is its pre-image.
- Due to this, Transfer Learning can be applied, therefore, for a given loss function l, it is possible to find a function $h : \mathcal{X} \times \mathcal{B}_{\mathcal{X}} \to \mathcal{Y}$ such that $y = \text{sign}(h(P_X, X)))$.
- Let $\mathcal{H}_k$ be the RKHS associated with kernel $k_1 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. For $\Phi(\mathcal{B}_{\mathcal{X}})$, the set of mean embeddings associated with $\mathcal{B}_{\mathcal{X}}$, let $\mathcal{H}_{k_P}$ be the RKHS associated with the kernel $k_P : \Phi(\mathcal{B}_{\mathcal{X}}) \times \Phi(\mathcal{B}_{\mathcal{X}}) \to \mathbb{R}$ (The mean embedding is defined as $\Phi(P_X) = \int_{\mathcal{X}} k_2(\cdot, x) dP_X$). We seek an estimate $h_{\mathcal{H}}$ of h such that the following criteria is satisfied

$$h_{\mathcal{H}} = \arg\min_{h \in \mathcal{H}_{\bar{k}}} \mathbb{E}_{P_{XY} \sim \lambda, (X,Y) \sim P_{XY}} [L(h(\Phi(P_X), X), Y)] \quad (2)$$

Orbit Determination

- When the following exist
  - An invertible mapping between the distribution of the feature vectors and the distribution of the center frequency in T_P.
  - A unique mapping between the orbit parameters in the support of P_Γ and the center frequency, given a communication system.

  Then, there exists a mapping between the distribution of the feature vectors and the orbit parameters.
- When the kernel used in an embedding is universal, the mapping between the RKHS and the Probability distribution is unique
- Therefore, when the above two criteria are satisfied, there exists a mapping between the kernel embedding of the RF feature vectors and the orbit parameters.
- Use a second universal kernel operator over the space of embeddings to estimate this function!

Distribution Regression

- P_Γ induces a probability distribution $\nu = P_{\Gamma_i} \otimes \boldsymbol{\delta}$ on $(\Gamma_{-1}, P_{X|y=-1})$ and on $(\Gamma_1, P_{X|Y=1})$.
- We estimate a function[3] f_l such that, for the mean embedding $\Psi(P_X) = \int_{\mathcal{X}} k(\cdot, x), dP_X$, and for the kernel operator $K : \Psi(P_X) \times \Psi(P_X) \to \mathcal{J}$,

$$\widehat{f}_{\xi_2,l} = \arg\min_{f_l \in \mathcal{H}_{k_P}} \mathbb{E}(\|f_l(\Psi(\widehat{P}_X)) - \gamma_l\|^2_{\mathcal{J}}) + \xi_2 \|f_l\|^2_{\mathcal{H}_{k_P}} \tag{3}$$

For $l = -1, 1$

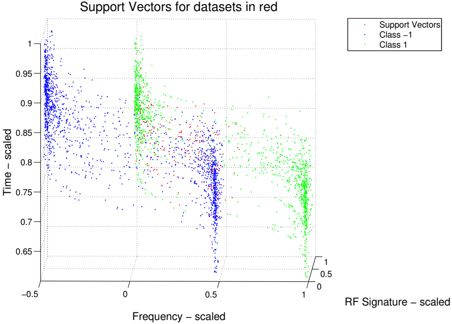
- This is an estimate of the function which maps the embedding of RF Feature Vectors to the orbit parameters!
- Applied to the test dataset to determine orbits!

[3] Zoltán Szabó et al. "Consistent, two-stage sampled distribution regression via mean embedding". In: *arXiv preprint arXiv:1402.1754* (2014).

RESULTS

Results - Classification

Two Spacecraft with BPSK transmissions, 10Kbps bandwidth, center frequencies separated by 10 Khz at 437.5MHz, signals received at 10dB SNR at one ground station over one pass. Synthetic Data, feature vectors were RF Signatures (3 dimensional) by Bkassiny et. al[4]



Classification Method	% Error
Transfer Learning	2.93
Pooled Classification	6.83

Training: 40 orbits
Testing: 10 orbits
SVM with bias

[4] Mario Bkassiny et al. "Blind cyclostationary feature detection based spectrum sensing for autonomous self-learning cognitive radios". In: *Communications (ICC), 2012 IEEE International Conference on*. IEEE. 2012, pp. 1507–1511.

Results - Orbit Determination

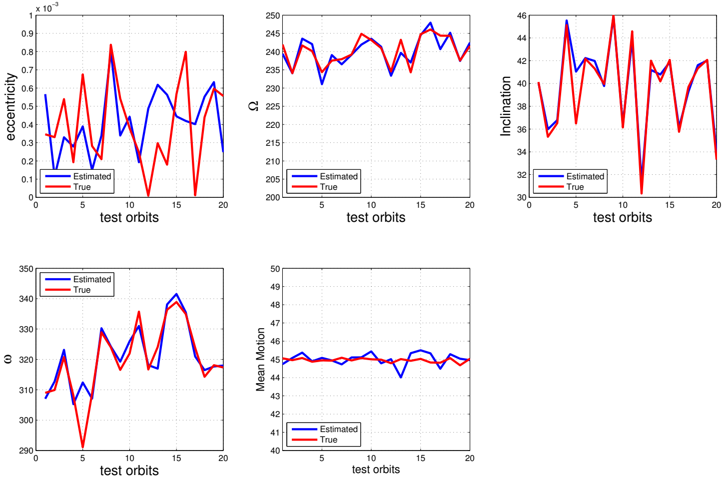
One Spacecraft with BPSK transmissions, Two Ground Stations (Ann Arbor and Chicago) $\sim 10dB$ SNR at receiver in an AWGN channel, Receiving over two passes (2 hour interval), Synthetic Data. Probability of transmission 0.03, uniform over the passes in T_P. Mean Motion was kept constant at 14.7732. Priors:

$$
\begin{aligned}
e &\sim \mathbf{1}_{0 \le e \le 1\times 10^{-3}}(\mathcal{N}(4\times 10^{-4}, 1\times 10^{-8}) + \mathcal{N}(4\times 10^{-4}, 1\times 10^{-8})) \\
\Omega &\sim \mathcal{N}(3\pi/2, (\pi/36)^2) \\
I &\sim \mathcal{N}(\pi/4, (\pi/18^2)) \\
\omega &\sim \mathcal{N}(16\pi/9, (\pi/18)^2) \\
M &\sim \mathcal{N}(\pi/4, (\pi/1800)^2)
\end{aligned}
$$

Training with 1080 orbit insertions, testing with 20 orbits.

Results

Results based on a small cross validation set indicates initial convergence



Future Work

- Upper and Lower bounds on performance of successive classification and orbit determination.
- Integration of two loss functions into one single machine learning system.
- Experimental Implementation into networked global ground stations.
- Extension to Geolocation systems.

Acknowledgements

- JPL Strategic University Research Partnership FY2015

Thank You

Questions?