Gaussian Mixture Penalization for Trajectory Optimization Problems

C. Rommel[1,2], J. F. Bonnans[1],
B. Gregorutti[2] and P. Martinon[1]

CMAP Ecole Polytechnique - INRIA[1]
Safety Line[2]

ISMP - July 2nd 2018

Motivation

- 20 000 airplanes — 80 000 flights per day,

Motivation

- 20 000 airplanes — 80 000 flights per day,
- Should double until 2033,

Motivation

- 20 000 airplanes — 80 000 flights per day,
- Should double until 2033,
- Responsible for 3% of CO_2 emissions,

Motivation

- 20 000 airplanes — 80 000 flights per day,
- Should double until 2033,
- Responsible for 3% of CO_2 emissions,
- Accounts for 30% of operational cost for an airline,

Motivation

- 20 000 airplanes — 80 000 flights per day,
- Should double until 2033,
- Responsible for 3% of CO_2 emissions,
- Accounts for 30% of operational cost for an airline,
- Rectilinear climb trajectories at full thrust.

Motivation

- 20 000 airplanes — 80 000 flights per day,
- Should double until 2033,
- Responsible for 3% of CO_2 emissions,
- Accounts for 30% of operational cost for an airline,
- Rectilinear climb trajectories at full thrust.

OPTI CLIMB
More savings, less CO_2

Optimal Control Problem

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$

$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = g(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t \in [0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t) \in U_{ad}, \quad \boldsymbol{x}(t) \in X_{ad}, \quad \text{for a.e. } t \in [0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, \quad \text{for all } t \in [0,t_f]. \end{cases} \tag{OCP}$$

Optimal Control Problem

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$

$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = g(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t\in[0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t)\in U_{ad}, \quad \boldsymbol{x}(t)\in X_{ad}, \quad \text{for a.e. } t\in[0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, \quad \text{for all } t\in[0,t_f]. \end{cases} \tag{OCP}$$

Optimal Control Problem

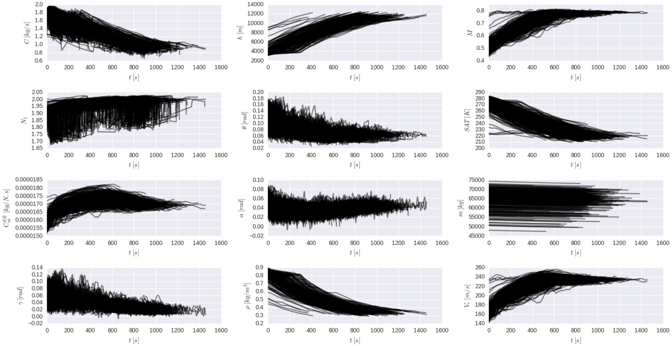
$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$

$$\text{s.t.} \begin{cases} \dot{\boldsymbol{x}} = \hat{g}(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t \in [0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t) \in U_{ad}, \quad \boldsymbol{x}(t) \in X_{ad}, \quad \text{for a.e. } t \in [0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \le 0, \quad \text{for all } t \in [0,t_f]. \end{cases} \tag{OCP}$$

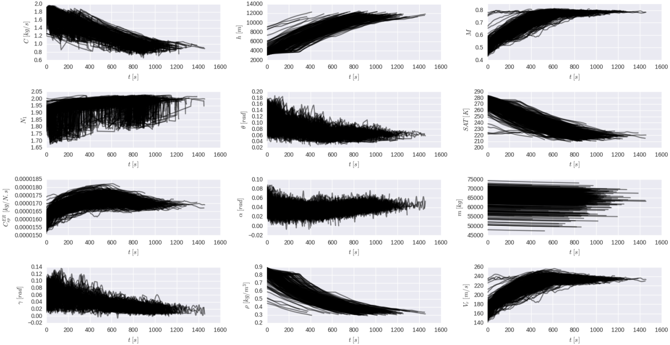
Dynamics are learned from QAR data

Dynamics are learned from QAR data

Dynamics are learned from QAR data



Dynamics are learned from QAR data



See e.g. [Rommel et al., 2017a] and [Rommel et al., 2017b]

Trajectory acceptability

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$

$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = \hat{g}(t,\boldsymbol{u},\boldsymbol{x}), & \text{for a.e. } t\in[0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t)\in U_{ad}, \quad \boldsymbol{x}(t)\in X_{ad}, & \text{for a.e. } t\in[0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \le 0, & \text{for all } t\in[0,t_f]. \end{cases} \tag{OCP}$$

Trajectory acceptability

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$

$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = \hat{g}(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t\in[0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t)\in U_{ad}, \quad \boldsymbol{x}(t)\in X_{ad}, \quad \text{for a.e. } t\in[0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, \quad \text{for all } t\in[0,t_f]. \end{cases} \tag{OCP}$$

$\Rightarrow \hat{\boldsymbol{z}} = (\hat{\boldsymbol{x}},\hat{\boldsymbol{u}})$ solution of (OCP).

Trajectory acceptability

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$

$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = \hat{g}(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t\in[0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t)\in U_{ad}, \quad \boldsymbol{x}(t)\in X_{ad}, \quad \text{for a.e. } t\in[0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, \quad \text{for all } t\in[0,t_f]. \end{cases} \tag{OCP}$$

$$\Rightarrow \hat{\boldsymbol{z}} = (\hat{\boldsymbol{x}},\hat{\boldsymbol{u}}) \text{ solution of (OCP).}$$

Is $\hat{z}$ inside the validity region of the dynamics model $\hat{g}$?

Trajectory acceptability

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$

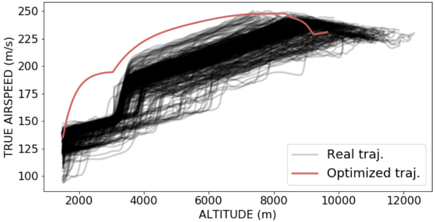
$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = \hat{g}(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t\in[0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t)\in U_{ad}, \quad \boldsymbol{x}(t)\in X_{ad}, \quad \text{for a.e. } t\in[0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, \quad \text{for all } t\in[0,t_f]. \end{cases} \tag{OCP}$$

$$\Rightarrow \hat{\boldsymbol{z}} = (\hat{\boldsymbol{x}},\hat{\boldsymbol{u}}) \text{ solution of (OCP).}$$

Is $\hat{z}$ inside the validity region of the dynamics model $\hat{g}$?

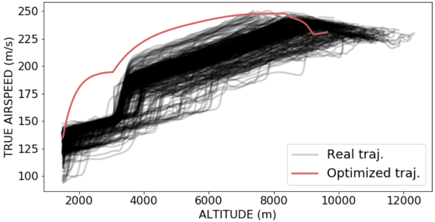
Does it look like a real aicraft trajectory ?

Trajectory acceptability



[1]NATS UK air traffic control

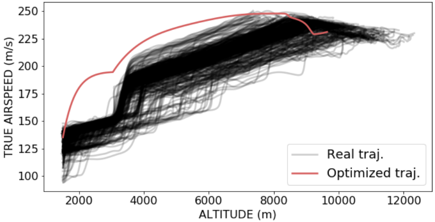
Trajectory acceptability



Pilots acceptance

[1]NATS UK air traffic control

Trajectory acceptability

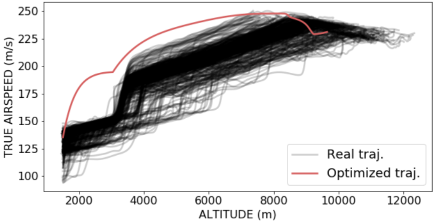


Pilots acceptance

Air Traffic Control[1]

[1]NATS UK air traffic control

Trajectory acceptability



Pilots acceptance Air Traffic Control[1]

How can we quantify the closeness from the optimized trajectory to the set of real flights?

[1]NATS UK air traffic control

Likelihood

Let X be a random variable following an absolutely continuous probability distribution with density function f depending on a parameter θ. Then the function

$$\mathcal{L}(\theta|x) = f_\theta(x) \quad (1)$$

considered as a function of θ, is the likelihood function of *theta*, given the outcome x of X.

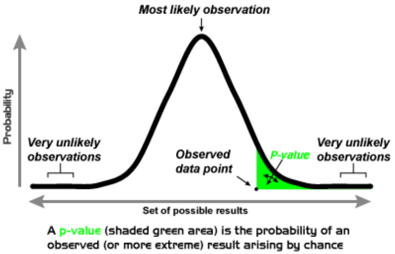
[0] Picture source: wikipedia, P-Value, author: Repapetilto CC.

Likelihood

Let X be a random variable following an absolutely continuous probability distribution with density function f depending on a parameter θ. Then the function

$$\mathcal{L}(\theta|x) = f_\theta(x) \tag{1}$$

considered as a function of θ, is the likelihood function of *theta*, given the outcome x of X.



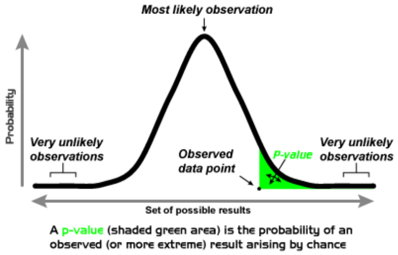
[0] Picture source: wikipedia, P-Value, author: Repapetilto CC.

Likelihood

Let X be a random variable following an absolutely continuous probability distribution with density function f depending on a parameter θ. Then the function

$$\mathcal{L}(\theta|x) = f_\theta(x) \quad (1)$$

considered as a function of θ, is the likelihood function of *theta*, given the outcome x of X.



In our case:

- the optimized trajectory plays the role of θ,

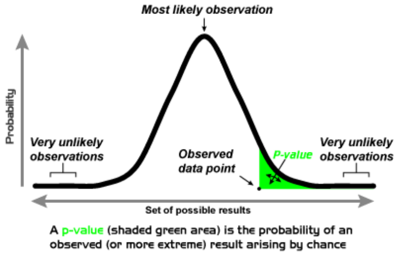
[0] Picture source: wikipedia, P-Value, author: Repapetilto CC.

Likelihood

Let X be a random variable following an absolutely continuous probability distribution with density function f depending on a parameter θ. Then the function

$$\mathcal{L}(\theta|x) = f_\theta(x) \tag{1}$$

considered as a function of θ, is the likelihood function of *theta*, given the outcome x of X.



In our case:

- the optimized trajectory plays the role of θ,
- the set of real flights plays the role of x,

[0]Picture source: wikipedia, P-Value, author: Repapetilto CC.

How to apply this to functional data?

Assumption: We suppose that the real flights are observations of the same functional random variable $Z = (Z_t)$ valued in $\mathcal{C}(\mathbb{T}, E)$, with E compact subset of $\mathbb{R}^d$ and $\mathbb{T} = [0, t_f]$.

How to apply this to functional data?

Assumption: We suppose that the real flights are observations of the same functional random variable $Z = (Z_t)$ valued in $\mathcal{C}(\mathbb{T}, E)$, with E compact subset of $\mathbb{R}^d$ and $\mathbb{T} = [0, t_f]$.

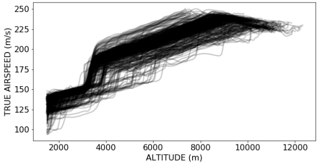
Problem: Computation of probability densities in infinite dimensional space is untractable...

How to apply this to functional data?

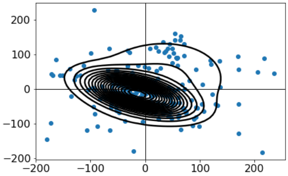
Assumption: We suppose that the real flights are observations of the same functional random variable $Z = (Z_t)$ valued in $\mathcal{C}(\mathbb{T}, E)$, with E compact subset of $\mathbb{R}^d$ and $\mathbb{T} = [0, t_f]$.

Problem: Computation of probability densities in infinite dimensional space is untractable...

- Standard approach FDA: use FPCA to decompose the data in a small number of coefficients



$\Rightarrow$

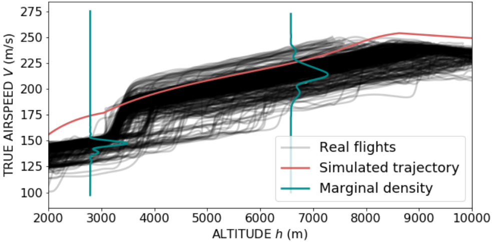


How to apply this to functional data?

Assumption: We suppose that the real flights are observations of the same functional random variable $Z = (Z_t)$ valued in $\mathcal{C}(\mathbb{T}, E)$, with E compact subset of $\mathbb{R}^d$ and $\mathbb{T} = [0, t_f]$.

Problem: Computation of probability densities in infinite dimensional space is untractable...

- Standard approach FDA: use FPCA to decompose the data in a small number of coefficients
- **Or: we can aggregate the marginal densities**



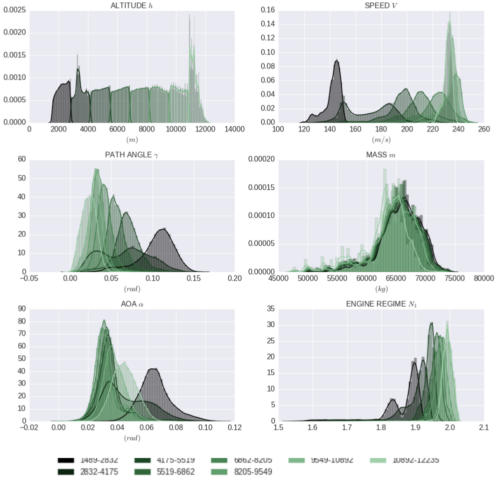
Why does it make sense for this type of data?

Why does it make sense for this type of data?

Likely values of flight variables during climb are strongly dependent on the altitude

Why does it make sense for this type of data?

Likely values of flight variables during climb are strongly dependent on the altitude



How do we aggregate the marginal likelihoods?

How do we aggregate the marginal likelihoods?

- f_t marginal density of Z, i.e. probability density function of Z_t,

How do we aggregate the marginal likelihoods?

- f_t marginal density of Z, i.e. probability density function of Z_t,
- $\boldsymbol{y}$ new trajectory,

How do we aggregate the marginal likelihoods?

- f_t marginal density of Z, i.e. probability density function of Z_t,
- $\boldsymbol{y}$ new trajectory,
- $f_t(\boldsymbol{y}(t))$ marginal likelihood of $\boldsymbol{y}$ at t, i.e. likelihood of observing $Z_t = \boldsymbol{y}(t)$.

How do we aggregate the marginal likelihoods?

- f_t marginal density of Z, i.e. probability density function of Z_t,
- $\boldsymbol{y}$ new trajectory,
- $f_t(\boldsymbol{y}(t))$ marginal likelihood of $\boldsymbol{y}$ at t, i.e. likelihood of observing $Z_t = \boldsymbol{y}(t)$.

Why not average over time ?...

$$\frac{1}{t_f} \int_0^{t_f} f_t(\boldsymbol{y}(t))dt$$

How do we aggregate the marginal likelihoods?

- f_t marginal density of Z, i.e. probability density function of Z_t,
- $\boldsymbol{y}$ new trajectory,
- $f_t(\boldsymbol{y}(t))$ marginal likelihood of $\boldsymbol{y}$ at t, i.e. likelihood of observing $Z_t = \boldsymbol{y}(t)$.

Why not average over time ?...

$$\frac{1}{t_f} \int_0^{t_f} f_t(\boldsymbol{y}(t))dt$$

Marginal densities may have really different shapes

How do we aggregate the marginal likelihoods?

- f_t marginal density of Z, i.e. probability density function of Z_t,
- $\boldsymbol{y}$ new trajectory,
- $f_t(\boldsymbol{y}(t))$ marginal likelihood of $\boldsymbol{y}$ at t, i.e. likelihood of observing $Z_t = \boldsymbol{y}(t)$.

Mean marginal likelihood [Rommel et al., 2018]

$$\mathrm{MML}(Z, \boldsymbol{y}) = \frac{1}{t_f} \int_0^{t_f} \psi[f_t, \boldsymbol{y}(t)] dt,$$

where $\psi : L^1(E, \mathbb{R}_+) \times \mathbb{R} \to [0; 1]$ is a continuous scaling map.

How do we aggregate the marginal likelihoods?

Possible scalings are the normalized density

$$\psi[f_t, \boldsymbol{y}(t)] := \frac{\boldsymbol{y}(t)}{\max_{z \in E} f_t(z)},$$

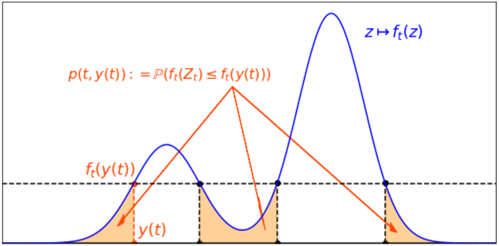
How do we aggregate the marginal likelihoods?

Possible scalings are the normalized density

$$\psi[f_t, \boldsymbol{y}(t)] := \frac{\boldsymbol{y}(t)}{\max\limits_{z \in E} f_t(z)},$$

or the confidence level

$$\psi[f_t, \boldsymbol{y}(t)] := \mathbb{P}\left(f_t(Z_t) \leq f_t(\boldsymbol{y}(t))\right).$$



How do we deal with sampled curves?

How do we deal with sampled curves?

In practice, the m trajectories are sampled at variable discrete times:

$$\mathcal{T}^D := \{(t_j^r, z_j^r)\}_{\substack{1\leq j\leq n \\ 1\leq r\leq m}} \subset \mathbb{T} \times E, \qquad z_j^r := \boldsymbol{z}^r(t_j^r),$$

$$\mathcal{Y} := \{(\tilde{t}_j, y_j)\}_{j=1}^{\tilde{n}} \subset \mathbb{T} \times E, \qquad y_j := \boldsymbol{y}(\tilde{t}_j).$$

How do we deal with sampled curves?

In practice, the m trajectories are sampled at variable discrete times:

$$\mathcal{T}^D := \{(t_j^r, z_j^r)\}_{\substack{1\leq j\leq n \\ 1\leq r\leq m}} \subset \mathbb{T} \times E, \qquad z_j^r := \boldsymbol{z}^r(t_j^r),$$

$$\mathcal{Y} := \{(\tilde{t}_j, y_j)\}_{j=1}^{\tilde{n}} \subset \mathbb{T} \times E, \qquad y_j := \boldsymbol{y}(\tilde{t}_j).$$

Hence, we approximate the MML using a Riemann sum which aggregates consistent estimators $\hat{f}_{\tilde{t}_j}^m$ of the marginal densities $f_{\tilde{t}_j}$:

$$\boxed{\mathrm{EMML}_m(\mathcal{T}^D, \mathcal{Y}) := \frac{1}{t_f} \sum_{j=1}^{\tilde{n}} \psi[\hat{f}_{\tilde{t}_j}^m, y_j] \Delta \tilde{t}_j.}$$

How can we estimate marginal densities?

- Suppose that sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ are i.i.d. sampled from r.v. T, indep. Z;

How can we estimate marginal densities?

- Suppose that sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ are i.i.d. sampled from r.v. T, indep. Z;
- f_t is the density of $Z_t = (Z_T | T = t) = (Y|X)$;

How can we estimate marginal densities?

- Suppose that sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ are i.i.d. sampled from r.v. T, indep. Z;
- f_t is the density of $Z_t = (Z_T | T = t) = (Y|X)$;
- Our problem can be seen as a conditional probability density learning problem with $(X, Y) = (T, Z_T)$.

How can we estimate marginal densities?

- Suppose that sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ are i.i.d. sampled from r.v. T, indep. Z;
- f_t is the density of $Z_t = (Z_T | T = t) = (Y|X)$;
- Our problem can be seen as a conditional probability density learning problem with $(X, Y) = (T, Z_T)$.

$\Rightarrow$ We could apply SOA conditional density estimation techniques, such as LS-CDE [Sugiyama et al., 2010],

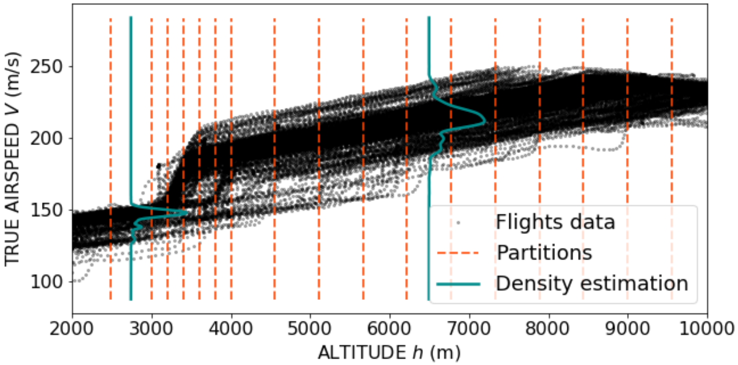
How can we estimate marginal densities?

- Suppose that sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ are i.i.d. sampled from r.v. T, indep. Z;
- f_t is the density of $Z_t = (Z_T | T = t) = (Y|X)$;
- Our problem can be seen as a conditional probability density learning problem with $(X, Y) = (T, Z_T)$.

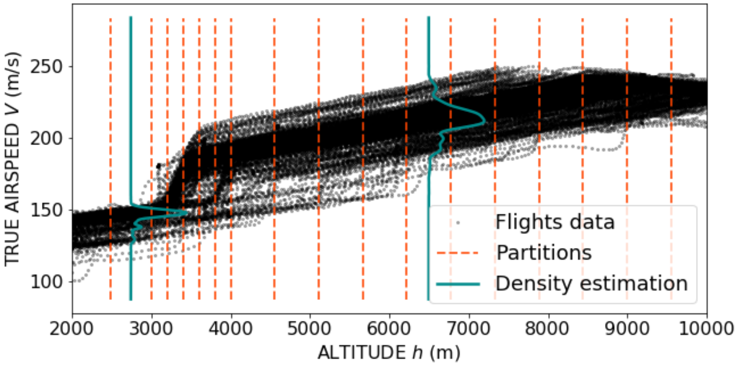
$\Rightarrow$ We could apply SOA conditional density estimation techniques, such as LS-CDE [Sugiyama et al., 2010],

$\Rightarrow$ **Instead, we choose to use a fine partitioning of the time domain.**

Partition based marginal density estimation



Partition based marginal density estimation



Idea: to average in time the marginal densities over small bins by applying classical multivariate density estimation techniques to each subset.

Consistency

Assumtion 1 - Positive time density

$\nu \in L^\infty(E, \mathbb{R}_+)$ density function of T, s.t.

$$\nu_+ := \operatorname*{ess\,sup}_{t \in \mathbb{T}} \nu(t) < \infty, \qquad \nu_- := \operatorname*{ess\,inf}_{t \in \mathbb{T}} \nu(t) > 0.$$

Consistency

Assumtion 1 - Positive time density

$\nu \in L^\infty(E, \mathbb{R}_+)$ density function of T, s.t.

$$\nu_+ := \operatorname*{ess\,sup}_{t \in \mathbb{T}} \nu(t) < \infty, \qquad \nu_- := \operatorname*{ess\,inf}_{t \in \mathbb{T}} \nu(t) > 0.$$

Assumtion 2 - Lipschitz in time

Function $(t, z) \in \mathbb{T} \times E \mapsto f_t(z)$ is continuous and

$$|f_{t_1}(z) - f_{t_2}(z)| \leq L|t_1 - t_2|, \qquad L > 0.$$

CONSISTENCY

ASSUMTION 1 - POSITIVE TIME DENSITY

$\nu \in L^\infty(E, \mathbb{R}_+)$ density function of T, s.t.

$$\nu_+ := \operatorname*{ess\,sup}_{t\in\mathbb{T}} \nu(t) < \infty, \qquad \nu_- := \operatorname*{ess\,inf}_{t\in\mathbb{T}} \nu(t) > 0.$$

ASSUMTION 2 - LIPSCHITZ IN TIME

Function $(t, z) \in \mathbb{T} \times E \mapsto f_t(z)$ is continuous and

$$|f_{t_1}(z) - f_{t_2}(z)| \leq L|t_1 - t_2|, \qquad L > 0.$$

ASSUMPTION 3 - SHRINKING BINS

The homogeneous partition $\{B_\ell^m\}_{\ell=1}^{q_m}$ of $[0; t_f]$, with binsize b_m, is s.t.

$$\lim_{m\to\infty} b_m = 0, \qquad \lim_{m\to\infty} m b_m = \infty.$$

Consistency

Assumption 4 - i.i.d. consistency

- $\mathcal{S} = \{(z_k)_{k=1}^N \in E^N : N \in \mathbb{N}^*\}$ set of finite sequences,
- $\Theta : \mathcal{S} \to L^1(E, \mathbb{R}_+)$ multivariate density estimation statistic,
- $\mathcal{G}$ arbitrary family of probability density functions on E, $\rho \in \mathcal{G}$,
- S_ρ^N i.i.d sample of size N drawn from ρ valued in $\mathcal{S}$.

The estimator obtained by applying Θ to S_ρ^N, denoted by

$$\hat{\rho}^N := \Theta[S_\rho^N] \in L^1(E, \mathbb{R}_+),$$

is a (pointwise) consistent density estimator, uniformly in ρ:

For all $z \in E, \varepsilon > 0, \alpha_1 > 0$, there is $N_{\varepsilon,\alpha_1} > 0$ such that, for any $\rho \in \mathcal{G}$,

$$N \geq N_{\varepsilon,\alpha_1} \Rightarrow \mathbb{P}\left(\left|\hat{\rho}^N(z) - \rho(z)\right| < \varepsilon\right) > 1 - \alpha_1.$$

Consistency

We denote by:

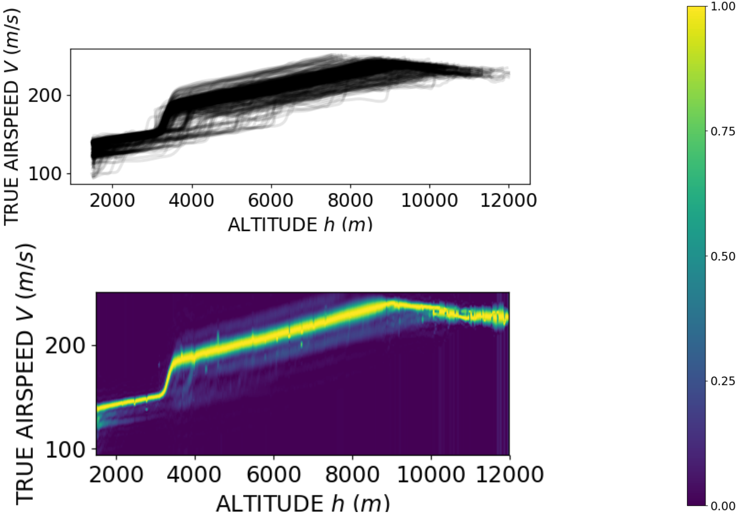
- $\ell^m(t) := \left\lceil \frac{t}{b_m} \right\rceil$ maps time to index of bin containing it;
- $\hat{f}^m_{\ell^m(t)} := \Theta[\mathcal{T}^m_{\ell^m(t)}]$ estimator trained using subset of data points $\mathcal{T}^m_{\ell^m(t)}$ whose sampling times fall in the bin containing t;

Theorem 1 - [Rommel et al., 2018]

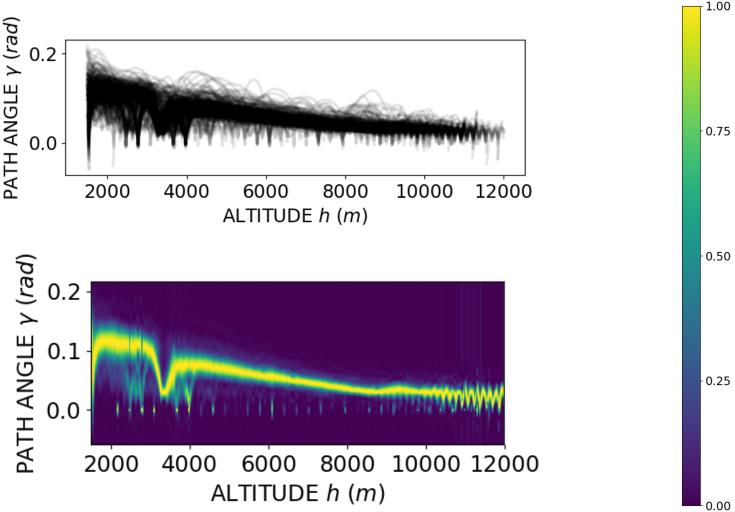
Under assumptions 1 to 4, for any $z \in E$ and $t \in \mathbb{T}$, $\hat{f}^m_{\ell^m(t)}(z)$ consistently approximates the marginal density $f_t(z)$ as the number of curves m grows:

$$\forall \varepsilon > 0, \quad \lim_{m \to \infty} \mathbb{P}\left(|\hat{f}^m_{\ell^m(t)}(z) - f_t(z)| < \varepsilon \right) = 1.$$

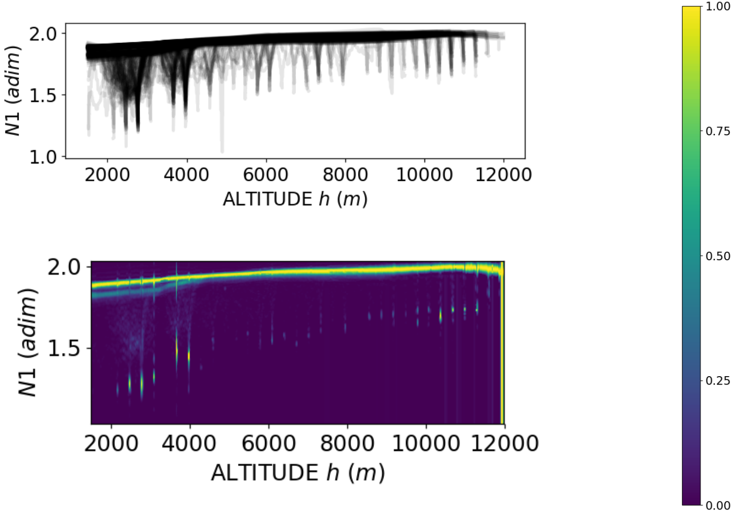
Marginal density estimation results



Marginal density estimation results



Marginal density estimation results



How good is it compared to other methods?

How good is it compared to other methods?

- Training set of $m = 424$ flights $\simeq$ 334 531 point observations,

How good is it compared to other methods?

- Training set of $m = 424$ flights $\simeq$ 334 531 point observations,
- Test set of 150 flights

How good is it compared to other methods?

- Training set of $m = 424$ flights $\simeq 334\ 531$ point observations,
- Test set of 150 flights = 50 real flights (*Real*), 50 optimized flights with operational constraints (*Opt1*) and 50 optimized flights without constraints (*Opt2*);

How good is it compared to other methods?

- Training set of $m = 424$ flights $\simeq$ 334 531 point observations,
- Test set of 150 flights = 50 real flights (*Real*), 50 optimized flights with operational constraints (*Opt1*) and 50 optimized flights without constraints (*Opt2*);
- Discrimination power comparison with (gmm-)FPCA and (integrated) LS-CDE:

How good is it compared to other methods?

- Training set of $m = 424$ flights $\simeq$ 334 531 point observations,
- Test set of 150 flights = 50 real flights (*Real*), 50 optimized flights with operational constraints (*Opt1*) and 50 optimized flights without constraints (*Opt2*);
- Discrimination power comparison with (gmm-)FPCA and (integrated) LS-CDE:

Var.	Estimated Likelihoods		
	Real	Opt1	Opt2
MML	**0.63** $\pm$ **0.07**	**0.43** $\pm$ **0.08**	**0.13** $\pm$ **0.02**
FPCA	0.16 $\pm$ 0.12	6.4e-03 $\pm$ 3.8e-03	3.6e-03 $\pm$ 5.4e-03
LS-CDE	0.77 $\pm$ 0.05	0.68 $\pm$ 0.04	0.49 $\pm$ 0.06

MML penalty

The MML can be used not only to assess the optimization solutions, but also to penalize the optimization itself:

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt$$

$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = g(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t\in[0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t)\in U_{ad}, \quad \boldsymbol{x}(t)\in X_{ad}, \quad \text{for a.e. } t\in[0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, \quad \text{for all } t\in[0,t_f]. \end{cases} \tag{OCP}$$

MML PENALTY

The MML can be used not only to assess the optimization solutions, but also to penalize the optimization itself:

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt - \lambda\,\mathrm{MML}(Z,\boldsymbol{x}),$$

$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = g(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t \in [0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t) \in U_{ad}, \quad \boldsymbol{x}(t) \in X_{ad}, \quad \text{for a.e. } t \in [0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, \quad \text{for all } t \in [0,t_f]. \end{cases} \tag{OCP}$$

MML PENALTY

The MML can be used not only to assess the optimization solutions, but also to penalize the optimization itself:

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt - \lambda\,\mathrm{MML}(Z,\boldsymbol{x}),$$

$$\text{s.t.} \begin{cases} \dot{\boldsymbol{x}} = g(t,\boldsymbol{u},\boldsymbol{x}), & \text{for a.e. } t\in[0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t)\in U_{ad}, \quad \boldsymbol{x}(t)\in X_{ad}, & \text{for a.e. } t\in[0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, & \text{for all } t\in[0,t_f]. \end{cases} \tag{OCP}$$

- λ sets trade-off between a fuel minimization and a likelihood maximization,

MML PENALTY

The MML can be used not only to assess the optimization solutions, but also to penalize the optimization itself:

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(t,\boldsymbol{u}(t),\boldsymbol{x}(t))dt - \lambda\,\mathrm{MML}(Z,\boldsymbol{x}),$$

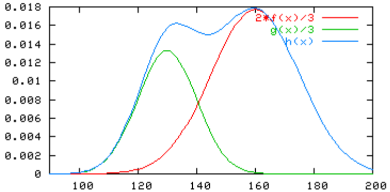
$$\text{s.t.} \quad \begin{cases} \dot{\boldsymbol{x}} = g(t,\boldsymbol{u},\boldsymbol{x}), \quad \text{for a.e. } t\in[0,t_f], \\ \Phi(\boldsymbol{x}(0),\boldsymbol{x}(t_f)) \in K_\Phi, \\ \boldsymbol{u}(t)\in U_{ad}, \quad \boldsymbol{x}(t)\in X_{ad}, \quad \text{for a.e. } t\in[0,t_f], \\ c(\boldsymbol{u}(t),\boldsymbol{x}(t)) \leq 0, \quad \text{for all } t\in[0,t_f]. \end{cases} \tag{OCP}$$

- λ sets trade-off between a fuel minimization and a likelihood maximization,
- If (OCP) is solved using NLP techniques, parametric estimator of MML is needed.

Gaussian mixture model for marginal densities

Gaussian mixture model for marginal densities

$$f_t(z) = \sum_{k=1}^{K} w_{t,k} \phi(z, \mu_{t,k}, \Sigma_{t,k}),$$

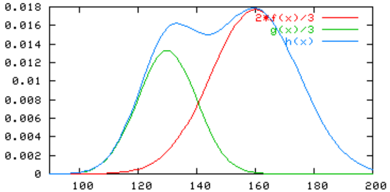


$$\sum_{k=1}^{K} w_{t,k} = 1, \qquad w_{t,k} \geq 0,$$

$$\phi(z, \mu, \Sigma) := \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} e^{-\frac{1}{2}(z-\mu)^\top \Sigma^{-1}(z-\mu)}.$$

Gaussian mixture model for marginal densities

$$\boxed{f_t(z) = \sum_{k=1}^{K} w_{t,k}\phi(z, \mu_{t,k}, \Sigma_{t,k}),}$$



$$\sum_{k=1}^{K} w_{t,k} = 1, \qquad w_{t,k} \geq 0,$$

$$\phi(z, \mu, \Sigma) := \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} e^{-\frac{1}{2}(z-\mu)^\top \Sigma^{-1}(z-\mu)}.$$

Assuming that the number of components is known, the weights $w_{t,k}$, means $\mu_{t,k}$ and covariance matrices $\Sigma_{t,k}$ need to be estimated.

Maximum likelihood parameters estimation

For $K = 1$, maximum likelihood estimates have closed form:

Maximum likelihood parameters estimation

For $K = 1$, maximum likelihood estimates have closed form:

$$\mathcal{L}(\mu_{t,1}, \Sigma_{t,1} | z_1, \ldots, z_N) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^d \det \Sigma_{t,1}}} e^{-\frac{1}{2}(z-\mu_{t,1})^\top \Sigma_{t,1}^{-1}(z-\mu_{t,1})}$$

Maximum likelihood parameters estimation

For $K = 1$, maximum likelihood estimates have closed form:

$$\mathcal{L}(\mu_{t,1}, \Sigma_{t,1} | z_1, \ldots, z_N) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^d \det \Sigma_{t,1}}} e^{-\frac{1}{2}(z-\mu_{t,1})^\top \Sigma_{t,1}^{-1}(z-\mu_{t,1})}$$

$$\hat{\theta} := (\hat{\mu}_{t,1}, \hat{\Sigma}_{t,1}) = \underset{(\mu_{t,1}, \Sigma_{t,1})}{\arg\min} \sum_{i=1}^{N} \Big(\log \det \Sigma_{t,1} + (z_i - \mu_{t,1})^\top \Sigma_{t,1}^{-1} (z_i - \mu_{t,1}) \Big)$$

Maximum likelihood parameters estimation

For $K = 1$, maximum likelihood estimates have closed form:

$$\mathcal{L}(\mu_{t,1}, \Sigma_{t,1} | z_1, \ldots, z_N) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^d \det \Sigma_{t,1}}} e^{-\frac{1}{2}(z-\mu_{t,1})^\top \Sigma_{t,1}^{-1}(z-\mu_{t,1})}$$

$$\hat{\theta} := (\hat{\mu}_{t,1}, \hat{\Sigma}_{t,1}) = \underset{(\mu_{t,1}, \Sigma_{t,1})}{\arg\min} \sum_{i=1}^{N} \Big(\log\det \Sigma_{t,1} + (z_i - \mu_{t,1})^\top \Sigma_{t,1}^{-1} (z_i - \mu_{t,1}) \Big)$$

$$\hat{\mu}_{t,1} = \frac{1}{N} \sum_{i=1}^{N} z_i, \qquad \hat{\Sigma}_{t,1} = \frac{1}{N} \sum_{i=1}^{N} (z_i - \hat{\mu}_{t,1})(z_i - \hat{\mu}_{t,1})^\top.$$

EM Algorithm

EM Algorithm

- Hidden random variable J valued on $\{1, \dots, K\}$,

EM Algorithm

- Hidden random variable J valued on $\{1, \ldots, K\}$,
- If i^{th} observation $J_i = k$, then z_i was drawn from the k^{th} component,

EM Algorithm

- Hidden random variable J valued on $\{1, \ldots, K\}$,
- If i^{th} observation $J_i = k$, then z_i was drawn from the k^{th} component,
- Group observations by component and compute $(\hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})$ with $K = 1$ maximum likelihood formulas.

EM Algorithm

- Hidden random variable J valued on $\{1, \ldots, K\}$,
- If i^{th} observation $J_i = k$, then z_i was drawn from the k^{th} component,
- Group observations by component and compute $(\hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})$ with $K = 1$ maximum likelihood formulas.

Expectation-Maximization - [Dempster et al., 1977]

Initialization: $\hat{\theta} = (\hat{w}_{t,k}, \hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})_{k=1}^{K} = (w_{t,k}^{0}, \mu_{t,k}^{0}, \Sigma_{t,k}^{0})_{k=1}^{K}$,

EM Algorithm

- Hidden random variable J valued on $\{1, \ldots, K\}$,
- If i^{th} observation $J_i = k$, then z_i was drawn from the k^{th} component,
- Group observations by component and compute $(\hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})$ with $K = 1$ maximum likelihood formulas.

Expectation-Maximization - [Dempster et al., 1977]

Initialization: $\hat{\theta} = (\hat{w}_{t,k}, \hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})_{k=1}^{K} = (w_{t,k}^{0}, \mu_{t,k}^{0}, \Sigma_{t,k}^{0})_{k=1}^{K}$,
Expectation: For $k = 1, \ldots, K$ and $i = 1, \ldots, N$,

$$\boxed{\hat{w}_{t,k} = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_{k,i},} \quad \hat{\pi}_{k,i} := \mathbb{P}(J_i = k | \hat{\theta}_t, Z_h) = \frac{\hat{\mu}_{t,k}\phi(z_i, \hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})}{\sum_{j=1}^{N} \hat{w}_{t,k}\phi(z_j, \hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})}.$$

EM Algorithm

- Hidden random variable J valued on $\{1, \dots, K\}$,
- If i^{th} observation $J_i = k$, then z_i was drawn from the k^{th} component,
- Group observations by component and compute $(\hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})$ with $K = 1$ maximum likelihood formulas.

Expectation-Maximization - [Dempster et al., 1977]

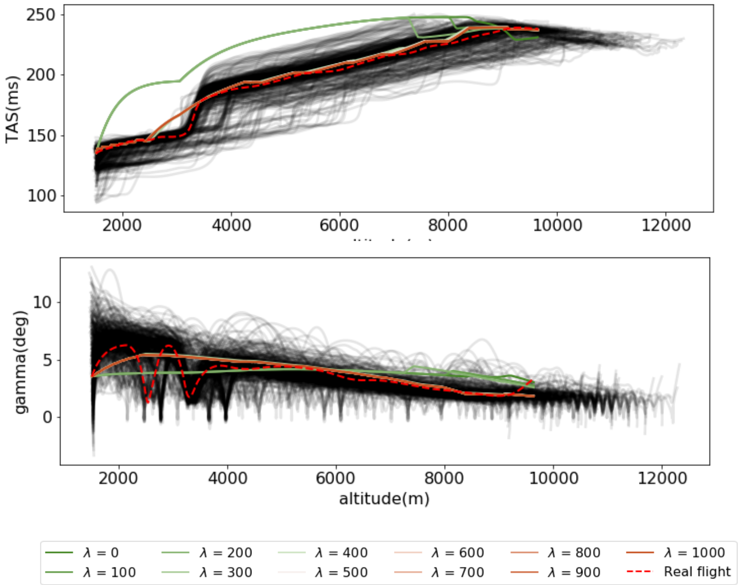
Initialization: $\hat{\theta} = (\hat{w}_{t,k}, \hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})_{k=1}^{K} = (w_{t,k}^0, \mu_{t,k}^0, \Sigma_{t,k}^0)_{k=1}^{K}$,
Expectation: For $k = 1, \dots, K$ and $i = 1, \dots, N$,

$$\boxed{\hat{w}_{t,k} = \frac{1}{N} \sum_{i=1}^{N} \hat{\pi}_{k,i},} \quad \hat{\pi}_{k,i} := \mathbb{P}(J_i = k | \hat{\theta}_t, Z_h) = \frac{\hat{\mu}_{t,k} \phi(z_i, \hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})}{\sum_{j=1}^{N} \hat{w}_{t,k} \phi(z_j, \hat{\mu}_{t,k}, \hat{\Sigma}_{t,k})}.$$

Maximization:

$$\boxed{\hat{\mu}_{t,k} = \frac{\sum_{i=1}^{N} \hat{\pi}_{k,i} z_i}{\sum_{i=1}^{N} \hat{\pi}_{k,i}},} \qquad \boxed{\hat{\Sigma}_{t,k} = \frac{\sum_{i=1}^{N} \hat{\pi}_{k,i} (z_i - \hat{\mu}_{t,k})(z_i - \hat{\mu}_{t,k})^\top}{\sum_{i=1}^{N} \hat{\pi}_{k,i}}.}$$

Penalty effect



Consumption x Acceptability trade-off

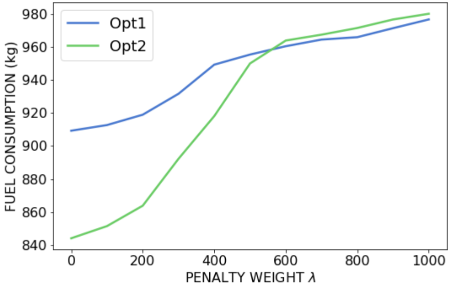


Figure: Average over 20 flights of the fuel consumption and MML score (called acceptability here) of optimized trajectories with varying MML-penalty weight λ.

Conclusion

1. General probabilistic criterion for quantifying the closeness between a curve and a set random trajectories,

Conclusion

1. General probabilistic criterion for quantifying the closeness between a curve and a set random trajectories,
2. Class of consistent plug-in estimators, based on “histogram” of multivariate density estimators,

Conclusion

1. General probabilistic criterion for quantifying the closeness between a curve and a set random trajectories,
2. Class of consistent plug-in estimators, based on “histogram” of multivariate density estimators,
3. Applicable to the case of aircraft climb trajectories,

Conclusion

1. General probabilistic criterion for quantifying the closeness between a curve and a set random trajectories,
2. Class of consistent plug-in estimators, based on "histogram" of multivariate density estimators,
3. Applicable to the case of aircraft climb trajectories,
   - Competitive with other well-established SOA approaches,

Conclusion

1. General probabilistic criterion for quantifying the closeness between a curve and a set random trajectories,
2. Class of consistent plug-in estimators, based on "histogram" of multivariate density estimators,
3. Applicable to the case of aircraft climb trajectories,
   - Competitive with other well-established SOA approaches,
4. Particular Gaussian mixture model implementation,

Conclusion

1. General probabilistic criterion for quantifying the closeness between a curve and a set random trajectories,
2. Class of consistent plug-in estimators, based on "histogram" of multivariate density estimators,
3. Applicable to the case of aircraft climb trajectories,
   - Competitive with other well-established SOA approaches,
4. Particular Gaussian mixture model implementation,
   - Showed that it can be used in optimal control problems to obtain solutions close to optimal, and still realistic.

Conclusion

1. General probabilistic criterion for quantifying the closeness between a curve and a set random trajectories,
2. Class of consistent plug-in estimators, based on "histogram" of multivariate density estimators,
3. Applicable to the case of aircraft climb trajectories,
   - Competitive with other well-established SOA approaches,
4. Particular Gaussian mixture model implementation,
   - Showed that it can be used in optimal control problems to obtain solutions close to optimal, and still realistic.

$\Rightarrow$ How could we automatically set the trade-off ?...

THANK YOU FOR YOUR ATTENTION !!

REFERENCES

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (methodological), pages 1–38.

Rommel, C., Bonnans, J. F., Gregorutti, B., and Martinon, P. (2017a). Aircraft dynamics identification for optimal control. In Proceedings of the 7th European Conference for Aeronautics and Aerospace Sciences.

Rommel, C., Bonnans, J. F., Gregorutti, B., and Martinon, P. (2017b). Block sparse linear models for learning structured dynamical systems in aeronautics. HAL report hal-01816400.

Rommel, C., Bonnans, J. F., Gregorutti, B., and Martinon, P. (2018). Quantifying the closeness to a set of random curves via the mean marginal likelihood. HAL report hal-01816407.

Sugiyama, M., Takeuchi, I., Suzuki, T., Kanamori, T., Hachiya, H., and Okanohara, D. (2010). Conditional density estimation via least-squares density ratio estimation. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pages 781–788.