Exploration de données pour l'optimisation de trajectoires aériennes

Cédric Rommel

Directeurs de thèse: Frédéric Bonnans, Pierre Martinon
Encadrant Safety Line: Baptiste Gregorutti

Soutenance de thèse, 26 octobre 2018

Context

MOTIVATION

FIGURE: World air traffic - source: www.flightradar24.com

Motivation

Figure: World air traffic - source: www.flightradar24.com

- 20 000 airplanes — 80 000 flights per day,

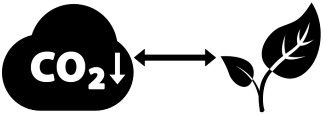
Motivation

Figure: World air traffic - source: www.flightradar24.com

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

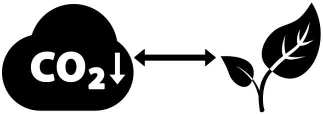
Motivation

- Most polluting means of transportation,



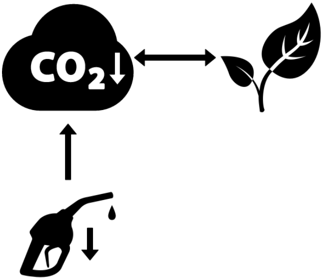
Motivation

- Most polluting means of transportation,
- Responsible for 3% of CO_2 emissions,



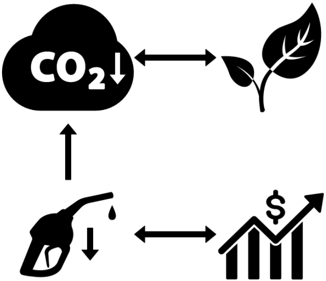
Motivation

- Most polluting means of transportation,
- Responsible for 3% of CO_2 emissions,



Motivation

- Most polluting means of transportation,
- Responsible for 3% of CO_2 emissions,
- Fuel $\simeq$ 30% of an airline operational cost,



MOTIVATION

How to tackle this problem ?

Motivation

How to tackle this problem ?

1. New hardware ?

Motivation

How to tackle this problem ?

1. New hardware ?
2. **Better use of existing fleet,**

Motivation

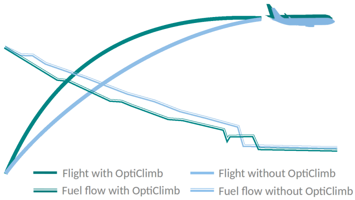
How to tackle this problem ?

1. New hardware ?
2. **Better use of existing fleet,**
   - Climb is the most consuming flight phase...

Motivation

How to tackle this problem ?

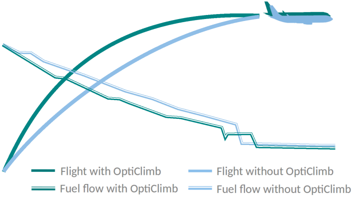
1. New hardware ?
2. **Better use of existing fleet,**
   - Climb is the most consuming flight phase...
   - Mostly rectilinear trajectories at full thrust,



MOTIVATION

How to tackle this problem ?

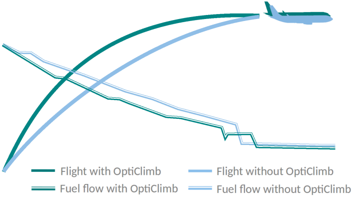
1. New hardware ?
2. **Better use of existing fleet,**
   - Climb is the most consuming flight phase...
   - Mostly rectilinear trajectories at full thrust,
   - Thousands of variables recorded every second,



Motivation

How to tackle this problem ?

1. New hardware ?
2. **Better use of existing fleet,**
   - Climb is the most consuming flight phase...
   - Mostly rectilinear trajectories at full thrust,
   - Thousands of variables recorded every second,

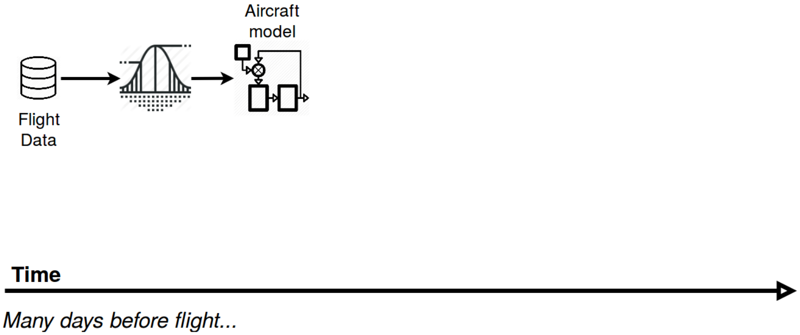


OPTI CLIMB

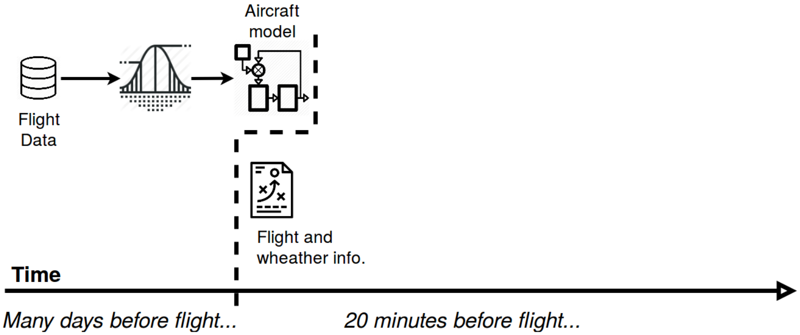
More savings, less CO_2

OptiClimb

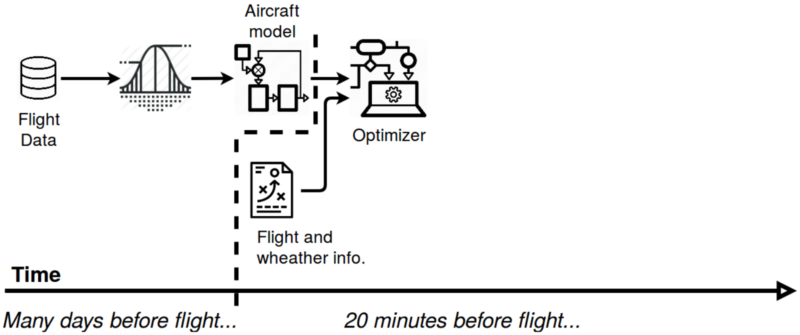
OPTICLIMB



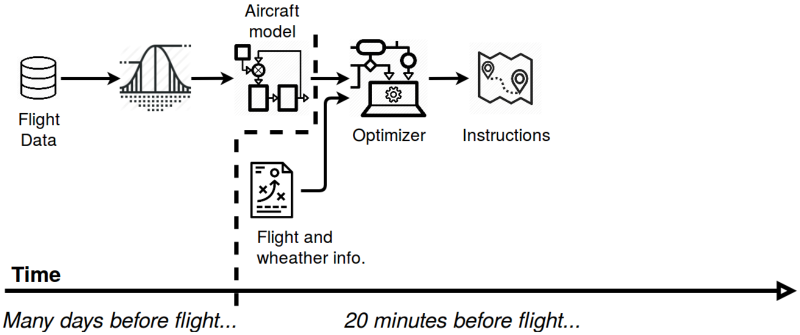
OptiClimb


Aircraft
model
Flight
Data
Flight and
wheather info.
Time
Many days before flight...
20 minutes before flight...

OptiClimb

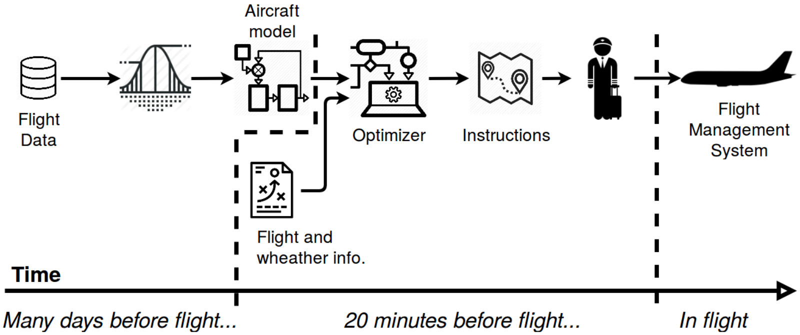

Aircraft model
Flight Data
Optimizer
Flight and wheather info.
Time
Many days before flight...
20 minutes before flight...

OPTICLIMB



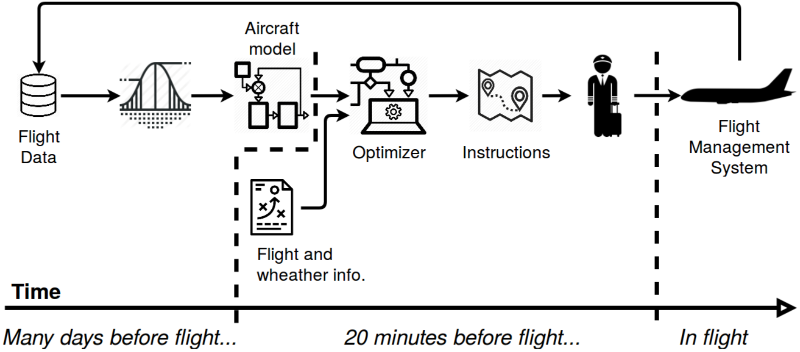
Aircraft
model
Flight
Data
Optimizer
Instructions
Flight and
wheather info.
Time
Many days before flight...
20 minutes before flight...

OptiClimb



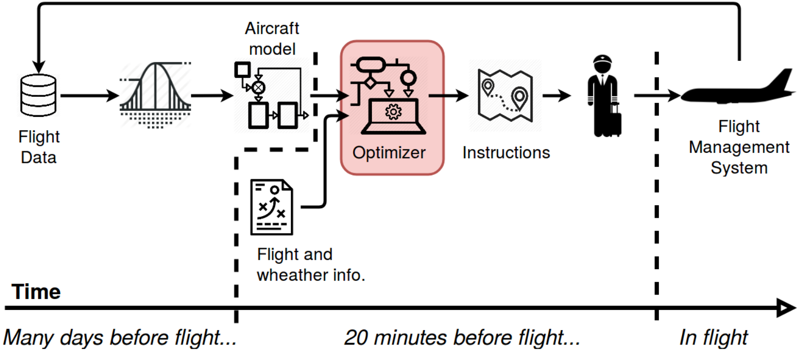
Aircraft model
Flight Data
Optimizer
Instructions
Flight Management System
Flight and wheather info.
Time
Many days before flight...
20 minutes before flight...
In flight

OptiClimb



Aircraft model
Flight Data
Optimizer
Instructions
Flight Management System
Flight and wheather info.
Time
Many days before flight...
20 minutes before flight...
In flight

OptiClimb

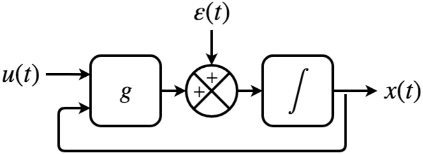


Aircraft model
Flight Data
Optimizer
Instructions
Flight Management System
Flight and wheather info.
Time
Many days before flight...
20 minutes before flight...
In flight

Trajectory Optimization

Dynamics:

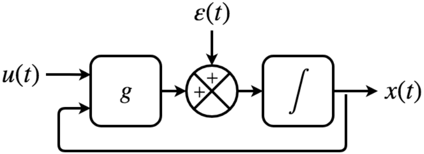
$$\dot{x}(t) = g(\boldsymbol{u}(t), \boldsymbol{x}(t)) + \varepsilon(t)$$



TRAJECTORY OPTIMIZATION

Dynamics:

$$\dot{x}(t) = g(\boldsymbol{u}(t), \boldsymbol{x}(t)) + \varepsilon(t)$$

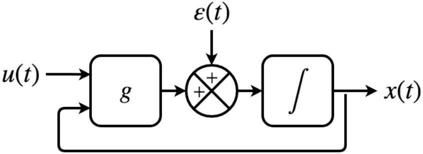


Optimization objective: $\int_0^{t_f} C(\boldsymbol{u}(t), \boldsymbol{x}(t))dt$

Trajectory Optimization

Dynamics:

$$\dot{x}(t) = g(\boldsymbol{u}(t), \boldsymbol{x}(t)) + \varepsilon(t)$$

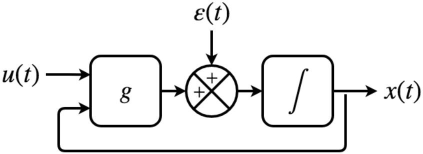


Optimization objective: $\int_0^{t_f} C(\boldsymbol{u}(t), \boldsymbol{x}(t))dt \Longleftarrow$  ,

Trajectory Optimization

Dynamics:

$$\dot{x}(t) = g(\boldsymbol{u}(t), \boldsymbol{x}(t)) + \varepsilon(t)$$

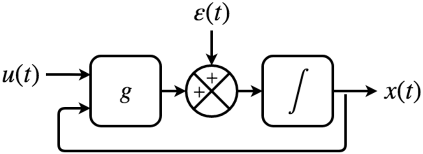


Optimization objective: $\int_0^{t_f} C(\boldsymbol{u}(t), \boldsymbol{x}(t))dt \Longleftarrow$ 

Trajectory Optimization

Dynamics:

$$\dot{x}(t) = g(\boldsymbol{u}(t), \boldsymbol{x}(t)) + \varepsilon(t)$$



Optimization objective: $\int_0^{t_f} C(\boldsymbol{u}(t), \boldsymbol{x}(t))dt \Longleftarrow$ 

Flight constraints:

$$\begin{cases} \Phi(\boldsymbol{x}(0), \boldsymbol{x}(t_f)) \in K_\Phi & \text{Initial and final conditions} \\ \boldsymbol{u}(t) \in U_{ad}, \quad \boldsymbol{x}(t) \in X_{ad}, & \text{Flight domain} \\ c(\boldsymbol{u}(t), \boldsymbol{x}(t)) \leq 0, & \text{Operational path constraints} \end{cases}$$

Trajectory Optimization

Optimal Control Problem

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

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

Optimal Control Problem

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

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

Trajectory Optimization

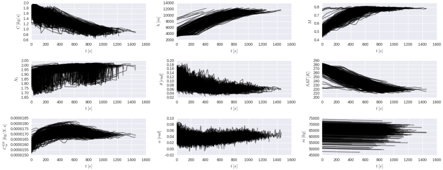
Optimal Control Problem

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

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

System Identification

Black box



QAR data

Trajectory Optimization

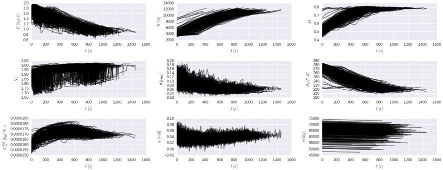
Approximate Optimal Control Problem

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

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

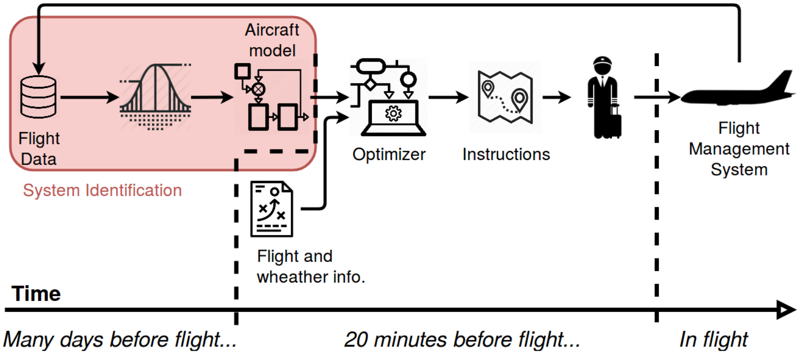
System Identification

Black box



QAR data

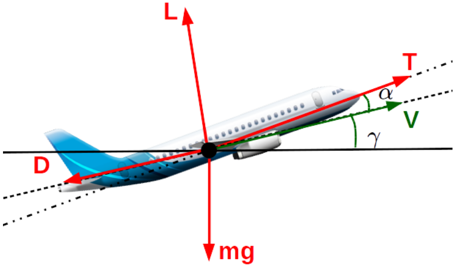
System Identification



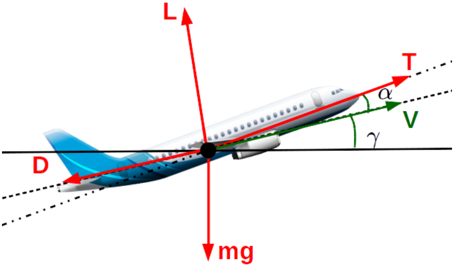
1. Context - *Chapter 1*
2. System Identification - *Chapter 4*
3. Trajectory Acceptability - *Chapters 5 and 6*

System Identification

Flight dynamics

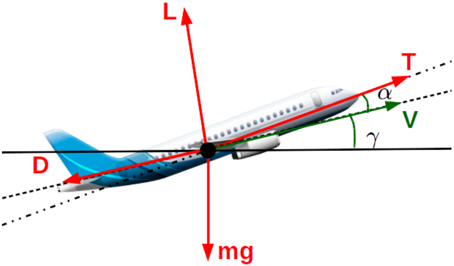


Flight dynamics



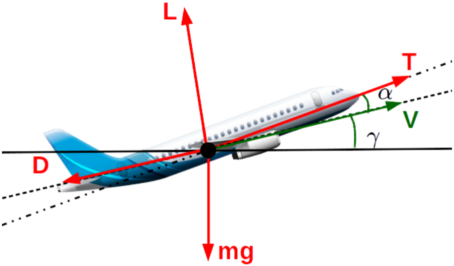
$$\begin{cases} \dot{h} = & V \sin \gamma \\ \dot{V} = & \dfrac{T \cos \alpha - D - mg \sin \gamma}{m} \\ \dot{\gamma} = & \dfrac{T \sin \alpha + L \qquad - mg \cos \gamma}{mV} \\ \dot{m} = & -\dfrac{T}{I_{sp}} \end{cases}$$

Flight dynamics



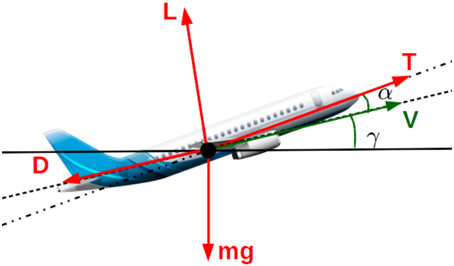
$$
\begin{cases}
\dot{h} = & V \sin\gamma + \dot{W}_z \\
\dot{V} = & \dfrac{T\cos\alpha - D - mg\sin\gamma - m\dot{W}_{xv}}{m} \\
\dot{\gamma} = & \dfrac{(T\sin\alpha + L)\cos\mu - mg\cos\gamma - m\dot{W}_{zv}}{mV} \\
\dot{m} = & -\dfrac{T}{I_{sp}}
\end{cases}
$$

Flight dynamics



$$\begin{cases} \dot{h} = & V \sin\gamma \\ \dot{V} = & \dfrac{T\cos\alpha - D - mg\sin\gamma}{m} \\ \dot{\gamma} = & \dfrac{T\sin\alpha + L \quad - mg\cos\gamma}{mV} \\ \dot{m} = & -\dfrac{T}{I_{sp}} \end{cases}$$

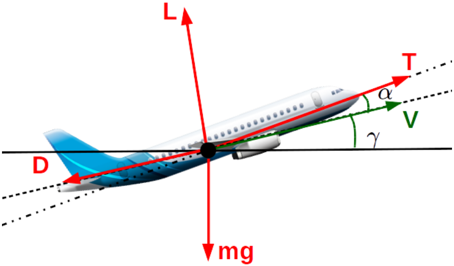
Flight dynamics



States: $\boldsymbol{x} = (h, V, \gamma, m)$

$$
\begin{cases}
\dot{h} = V \sin \gamma \\
\dot{V} = \dfrac{T \cos \alpha - D - mg \sin \gamma}{m} \\
\dot{\gamma} = \dfrac{T \sin \alpha + L - mg \cos \gamma}{mV} \\
\dot{m} = -\dfrac{T}{I_{sp}}
\end{cases}
$$

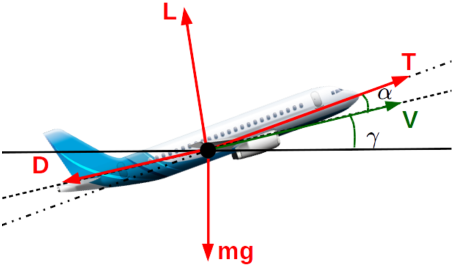
Flight dynamics



States: $\boldsymbol{x} = (h, V, \gamma, m)$
Controls: $\boldsymbol{u} = (\alpha, N_1)$

$$\begin{cases} \dot{h} = V \sin\gamma \\ \dot{V} = \dfrac{T\cos\alpha - D - mg\sin\gamma}{m} \\ \dot{\gamma} = \dfrac{T\sin\alpha + L - mg\cos\gamma}{mV} \\ \dot{m} = -\dfrac{T}{I_{sp}} \end{cases}$$

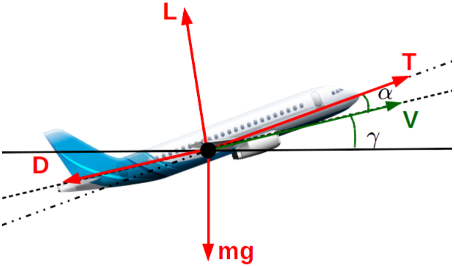
Flight dynamics



States: $\boldsymbol{x} = (h, V, \gamma, m)$
Controls: $\boldsymbol{u} = (\alpha, N_1)$
Unknown functions of $\boldsymbol{x}, \boldsymbol{u}$

$$\begin{cases} \dot{h} = & V \sin\gamma \\ \dot{V} = & \dfrac{T\cos\alpha - D - mg\sin\gamma}{m} \\ \dot{\gamma} = & \dfrac{T\sin\alpha + L \qquad - mg\cos\gamma}{mV} \\ \dot{m} = & -\dfrac{T}{I_{sp}} \end{cases}$$

Flight dynamics



States: $\boldsymbol{x} = (h, V, \gamma, m)$
Controls: $\boldsymbol{u} = (\alpha, N_1)$
Unknown functions of $\boldsymbol{x}, \boldsymbol{u}$

$$\begin{cases} \dot{h} = & V \sin\gamma \\ \dot{V} = & \dfrac{T(\boldsymbol{u}, \boldsymbol{x})\cos\alpha - D(\boldsymbol{u}, \boldsymbol{x}) - mg\sin\gamma}{m} \\ \dot{\gamma} = & \dfrac{T(\boldsymbol{u}, \boldsymbol{x})\sin\alpha + L(\boldsymbol{u}, \boldsymbol{x}) - mg\cos\gamma}{mV} \\ \dot{m} = & -\dfrac{T(\boldsymbol{u}, \boldsymbol{x})}{I_{sp}(\boldsymbol{u}, \boldsymbol{x})} \end{cases}$$

Physical models of nested functions

$$\begin{cases} T & \text{function of} \quad (N_1, M, \rho) \\ D & \text{function of} \quad (q, M, \alpha) \\ L & \text{function of} \quad (q, M, \alpha) \\ I_{sp} & \text{function of} \quad (SAT, M, h) \end{cases}$$

Physical models of nested functions

$$\begin{cases} T & \text{function of } \; (N_1, M, \rho) = \varphi_T(\boldsymbol{x}, \boldsymbol{u}) \\ D & \text{function of } \; (q, M, \alpha) = \varphi_D(\boldsymbol{x}, \boldsymbol{u}) \\ L & \text{function of } \; (q, M, \alpha) = \varphi_L(\boldsymbol{x}, \boldsymbol{u}) \\ I_{sp} & \text{function of } \; (SAT, M, h) = \varphi_{Isp}(\boldsymbol{x}, \boldsymbol{u}) \end{cases}$$

Physical models of nested functions

$$\begin{cases} T(\boldsymbol{x},\boldsymbol{u},\quad) &= N_1 \times P_T(\rho, M) \\ D(\boldsymbol{x},\boldsymbol{u},\quad) &= q \times P_D(\alpha, M) \\ L(\boldsymbol{x},\boldsymbol{u},\quad) &= q \times P_L(\alpha, M) \\ I_{sp}(\boldsymbol{x},\boldsymbol{u},\quad) &= SAT \times P_{Isp}(h, M) \end{cases}$$

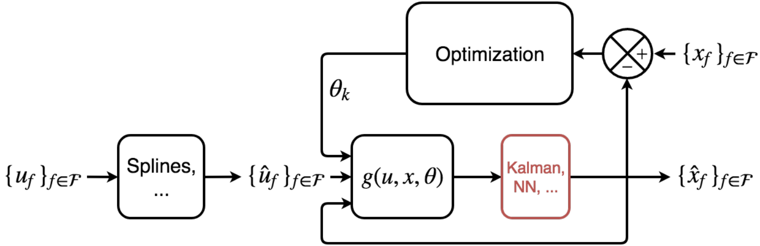
Physical models of nested functions

$$\begin{cases} T(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\theta}_T) &= N_1 \times P_T(\rho, M) &= X_T \cdot \boldsymbol{\theta}_T \\ D(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\theta}_D) &= q \times P_D(\alpha, M) &= X_D \cdot \boldsymbol{\theta}_D \\ L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\theta}_L) &= q \times P_L(\alpha, M) &= X_L \cdot \boldsymbol{\theta}_L \\ I_{sp}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\theta}_{Isp}) &= SAT \times P_{Isp}(h, M) &= X_{Isp} \cdot \boldsymbol{\theta}_{Isp} \end{cases}$$

$$X_T = N_1 \begin{pmatrix} 1 \\ \rho \\ M \\ \rho^2 \\ \rho M \\ M^2 \\ \vdots \end{pmatrix}, X_D = X_L = q \begin{pmatrix} 1 \\ \alpha \\ M \\ \alpha^2 \\ \alpha M \\ M^2 \\ \vdots \end{pmatrix}, X_{Isp} = SAT \begin{pmatrix} 1 \\ h \\ M \\ h^2 \\ hM \\ M^2 \\ \vdots \end{pmatrix}.$$

State-of-the-art - [Jategaonkar, 2006]

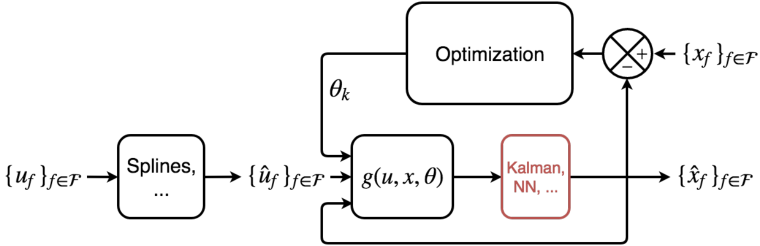
- Output-Error Method
- Filter-Error Method



State-of-the-art - [Jategaonkar, 2006]

- Output-Error Method
- Filter-Error Method

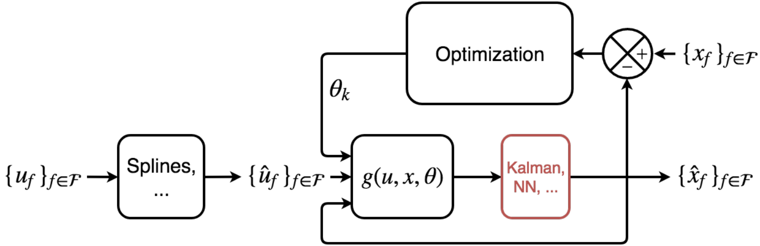
Less scalable to many trajectories



State-of-the-art - [Jategaonkar, 2006]

- Output-Error Method
- Filter-Error Method

} Less scalable to many trajectories



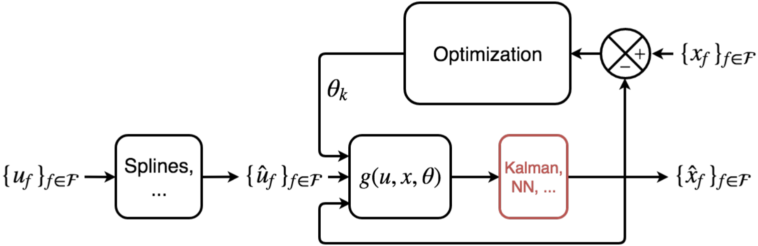
- **Equation-Error Method**

$$\dot{\boldsymbol{x}}(t) = g(\boldsymbol{u}(t), \boldsymbol{x}(t), \boldsymbol{\theta}) + \varepsilon(t), \quad t \in [0, t_f]$$

State-of-the-art - [Jategaonkar, 2006]

- Output-Error Method
- Filter-Error Method

Less scalable to many trajectories



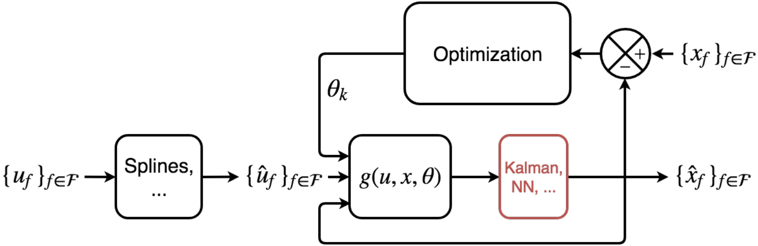
- **Equation-Error Method**

$$\dot{\boldsymbol{x}}_i = g(\boldsymbol{u}_i, \boldsymbol{x}_i, \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, \dots, N$$

State-of-the-art - [Jategaonkar, 2006]

- Output-Error Method
- Filter-Error Method

Less scalable to many trajectories



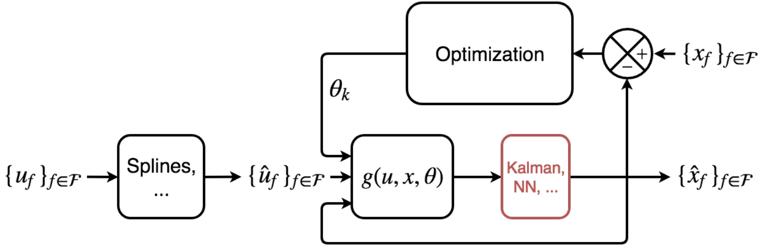
- **Equation-Error Method**

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \ell\left(\dot{\boldsymbol{x}}_i, g\left(\boldsymbol{u}_i, \boldsymbol{x}_i, \boldsymbol{\theta}\right)\right)$$

State-of-the-art - [Jategaonkar, 2006]

- Output-Error Method
- Filter-Error Method

Less scalable to many trajectories



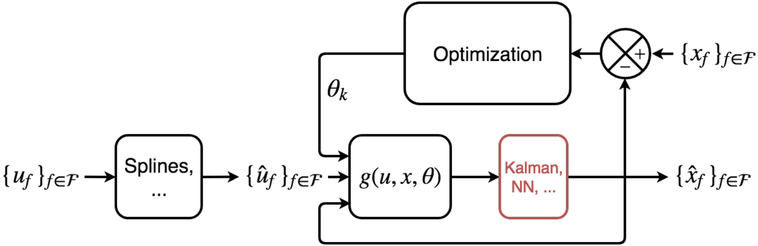
- **Equation-Error Method** *Ex: (Nonlinear) Least-Squares*

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left\| \dot{\boldsymbol{x}}_i - g\left(\boldsymbol{u}_i, \boldsymbol{x}_i, \boldsymbol{\theta}\right) \right\|_2^2$$

State-of-the-art - [Jategaonkar, 2006]

- Output-Error Method
- Filter-Error Method

Less scalable to many trajectories



- **Equation-Error Method** *Ex: (Nonlinear) Least-Squares*

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left\| Y(\boldsymbol{u}_i, \boldsymbol{x}_i, \dot{\boldsymbol{x}}_i) - G(\boldsymbol{u}_i, \boldsymbol{x}_i, \dot{\boldsymbol{x}}_i, \boldsymbol{\theta}) \right\|_2^2$$

Leveraging the dynamics structure

$$
\begin{cases}
\dot{h} = V \sin\gamma \\
\dot{V} = \dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)\cos\alpha - D(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_D) - mg\sin\gamma}{m} \\
\dot{\gamma} = \dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)\sin\alpha + L(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_L) - mg\cos\gamma}{mV} \\
\dot{m} = -\dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)}{I_{sp}(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_{Isp})}
\end{cases}
$$

Leveraging the dynamics structure

$$
\begin{cases}
\dot{h} \; = V \sin\gamma \\[2ex]
\dot{V} = \dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)\cos\alpha - D(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_D) - mg\sin\gamma}{m} \\[2ex]
\dot{\gamma} \; = \dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)\sin\alpha + L(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_L) - mg\cos\gamma}{mV} \\[2ex]
\dot{m} = -\dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)}{I_{sp}(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_{Isp})}
\end{cases}
$$

- Nonlinear in states and controls

Leveraging the dynamics structure

$$
\begin{cases}
\dot{h} = V \sin\gamma \\
\dot{V} = \dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)\cos\alpha - D(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_D) - mg\sin\gamma}{m} \\
\dot{\gamma} = \dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)\sin\alpha + L(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_L) - mg\cos\gamma}{mV} \\
\dot{m} = -\dfrac{T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)}{I_{sp}(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_{Isp})}
\end{cases}
$$

- Nonlinear in states and controls
- Nonlinear in parameters

Leveraging the dynamics structure

$$
\begin{cases}
\dot{h} = V \sin\gamma \\
m\dot{V} + mg\sin\gamma = T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)\cos\alpha - D(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_D) \\
mV\dot{\gamma} + mg\cos\gamma = T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T)\sin\alpha + L(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_L) \\
0 = T(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_T) + \dot{m} I_{sp}(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\theta}_{Isp})
\end{cases}
$$

- Nonlinear in states and controls
- Nonlinear in parameters

Leveraging the dynamics structure

$$\begin{cases} \dot{h} = V \sin\gamma \\ m\dot{V} + mg\sin\gamma = (X_T \cdot \boldsymbol{\theta}_T)\cos\alpha - X_D \cdot \boldsymbol{\theta}_D + \varepsilon_1 \\ mV\dot{\gamma} + mg\cos\gamma = (X_T \cdot \boldsymbol{\theta}_T)\sin\alpha + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\ 0 = X_T \cdot \boldsymbol{\theta}_T + \dot{m}(X_{Isp} \cdot \boldsymbol{\theta}_{Isp}) + \varepsilon_3 \end{cases}$$

$$\underbrace{\qquad\qquad\qquad}_{Y(\boldsymbol{u},\boldsymbol{x},\dot{\boldsymbol{x}})} \qquad \underbrace{\qquad\qquad\qquad\qquad}_{G(\boldsymbol{u},\boldsymbol{x},\dot{\boldsymbol{x}},\boldsymbol{\theta})}$$

- Nonlinear in states and controls
- ~~Nonlinear in parameters~~ $\rightarrow$ Linear in parameters

Leveraging the dynamics structure

$$
\begin{cases}
\dot{h} = V \sin\gamma \\
m\dot{V} + mg\sin\gamma = (X_T \cdot \boldsymbol{\theta}_T)\cos\alpha - X_D \cdot \boldsymbol{\theta}_D + \varepsilon_1 \\
mV\dot{\gamma} + mg\cos\gamma = (X_T \cdot \boldsymbol{\theta}_T)\sin\alpha + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\
0 = X_T \cdot \boldsymbol{\theta}_T + \dot{m}(X_{Isp} \cdot \boldsymbol{\theta}_{Isp}) + \varepsilon_3
\end{cases}
$$

$$
\underbrace{\hphantom{mV\dot{\gamma} + mg\cos\gamma}}_{Y(\boldsymbol{u},\boldsymbol{x},\dot{\boldsymbol{x}})} \qquad \underbrace{\hphantom{(X_T \cdot \boldsymbol{\theta}_T)\sin\alpha + X_L \cdot \boldsymbol{\theta}_L}}_{G(\boldsymbol{u},\boldsymbol{x},\dot{\boldsymbol{x}},\boldsymbol{\theta})}
$$

- Nonlinear in states and controls
- ~~Nonlinear in parameters~~ $\rightarrow$ Linear in parameters

Leveraging the dynamics structure

$$\begin{cases} \dot{h} = V \sin\gamma \\ Y_1 = X_{T1} \cdot \boldsymbol{\theta}_T - X_D \cdot \boldsymbol{\theta}_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \boldsymbol{\theta}_T + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\ Y_3 = X_T \cdot \boldsymbol{\theta}_T + X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} + \varepsilon_3 \end{cases}$$

- Nonlinear in states and controls
- ~~Nonlinear in parameters~~ $\rightarrow$ Linear in parameters

Leveraging the dynamics structure

$$
\begin{cases}
\dot{h} = V \sin \gamma \\
Y_1 = X_{T1} \cdot \boldsymbol{\theta}_T - X_D \cdot \boldsymbol{\theta}_D + \varepsilon_1 \\
Y_2 = X_{T2} \cdot \boldsymbol{\theta}_T + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\
Y_3 = X_T \cdot \boldsymbol{\theta}_T + X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} + \varepsilon_3
\end{cases}
$$

- Nonlinear in states and controls
- ~~Nonlinear in parameters~~ $\rightarrow$ Linear in parameters
- Structured
- Coupling

Leveraging the dynamics structure

$$\begin{cases} \dot{h} = V \sin\gamma \\ Y_1 = X_{T1} \cdot \boldsymbol{\theta}_T - X_D \cdot \boldsymbol{\theta}_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \boldsymbol{\theta}_T + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\ Y_3 = X_T \cdot \boldsymbol{\theta}_T + X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} + \varepsilon_3 \end{cases}$$

- Nonlinear in states and controls
- ~~Nonlinear in parameters~~ $\rightarrow$ Linear in parameters
- Structured
- Coupling $\rightsquigarrow$ $\boxed{\textbf{Multi-task Learning}}$

Multi-task Regression

Aircraft:

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ Y_3 = X_T \cdot \theta_T + X_{lspm} \cdot \theta_{lsp} + \varepsilon_3 \end{cases}$$

General:

$$\begin{cases} Y_1 = X_{c,1} \cdot \theta_c + X_1 \cdot \theta_1 + \varepsilon_1 \\ Y_2 = X_{c,2} \cdot \theta_c + X_2 \cdot \theta_2 + \varepsilon_2 \\ \vdots \qquad\qquad \vdots \\ Y_K = X_{c,K} \cdot \theta_c + X_K \cdot \theta_K + \varepsilon_K \end{cases}$$

Coupling parameters , **Task specific parameters**

Many other examples:

- *Giant squid neurons* [FitzHugh, 1961, Nagumo et al., 1962],
- *Susceptible-infectious-recovered models* [Anderson and May, 1992],
- *Mechanical systems,...*

Multi-task Regression

Aircraft:

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ Y_3 = X_T \cdot \theta_T + X_{lspm} \cdot \theta_{lsp} + \varepsilon_3 \end{cases}$$

General:

$$\begin{cases} Y_1 = X_{c,1} \cdot \theta_c + X_1 \cdot \theta_1 + \varepsilon_1 \\ Y_2 = X_{c,2} \cdot \theta_c + X_2 \cdot \theta_2 + \varepsilon_2 \\ \vdots \qquad \vdots \\ Y_K = X_{c,K} \cdot \theta_c + X_K \cdot \theta_K + \varepsilon_K \end{cases}$$

Coupling parameters , **Task specific parameters**

Multi-task Linear Least-Squares:

$$\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(Y_{k,i} - X_{c,k,i} \cdot \theta_c - X_{k,i} \cdot \theta_k \right)^2$$

Multi-task Regression

Aircraft:

$$\begin{cases} Y_1 = X_{T1} \cdot \boldsymbol{\theta}_T - X_D \cdot \boldsymbol{\theta}_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \boldsymbol{\theta}_T + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\ Y_3 = X_T \cdot \boldsymbol{\theta}_T + X_{lspm} \cdot \boldsymbol{\theta}_{lsp} + \varepsilon_3 \end{cases}$$

General:

$$\begin{cases} Y_1 = X_{c,1} \cdot \boldsymbol{\theta}_c + X_1 \cdot \boldsymbol{\theta}_1 + \varepsilon_1 \\ Y_2 = X_{c,2} \cdot \boldsymbol{\theta}_c + X_2 \cdot \boldsymbol{\theta}_2 + \varepsilon_2 \\ \vdots \qquad \vdots \\ Y_K = X_{c,K} \cdot \boldsymbol{\theta}_c + X_K \cdot \boldsymbol{\theta}_K + \varepsilon_K \end{cases}$$

Coupling parameters , Task specific parameters

Multi-task Linear Least-Squares: <u>Block-sparse Coupling Structure</u>

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left\| \begin{pmatrix} Y_{1,i} \\ \vdots \\ Y_{K,i} \end{pmatrix} - \begin{pmatrix} X_{c,1,i}^\top & X_{1,i}^\top & 0 & 0 & \dots & 0 \\ X_{c,2,i}^\top & 0 & X_{2,i}^\top & 0 & \dots & 0 \\ \vdots & 0 & 0 & \ddots & 0 & 0 \\ X_{c,K,i}^\top & 0 & 0 & \dots & 0 & X_{K,i}^\top \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_c \\ \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_K \end{pmatrix} \right\|_2^2$$

Multi-task Regression

Aircraft:

$$\begin{cases} Y_1 = X_{T1} \cdot \boldsymbol{\theta}_T - X_D \cdot \boldsymbol{\theta}_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \boldsymbol{\theta}_T + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\ Y_3 = X_T \cdot \boldsymbol{\theta}_T + X_{lspm} \cdot \boldsymbol{\theta}_{lsp} + \varepsilon_3 \end{cases}$$

General:

$$\begin{cases} Y_1 = X_{c,1} \cdot \boldsymbol{\theta}_c + X_1 \cdot \boldsymbol{\theta}_1 + \varepsilon_1 \\ Y_2 = X_{c,2} \cdot \boldsymbol{\theta}_c + X_2 \cdot \boldsymbol{\theta}_2 + \varepsilon_2 \\ \vdots \qquad \vdots \\ Y_K = X_{c,K} \cdot \boldsymbol{\theta}_c + X_K \cdot \boldsymbol{\theta}_K + \varepsilon_K \end{cases}$$

Coupling parameters , Task specific parameters

Multi-task Linear Least-Squares:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \| Y_i - X_i \boldsymbol{\theta} \|_2^2$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_c, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K) \in \mathbb{R}^p$, $p = d_c + \sum_{k=1}^{K} d_k$,
$Y_i \in \mathbb{R}^K$ and $X_i \in \mathbb{R}^{K \times p}$.

Feature Selection

Our model:

$$\begin{aligned}T =&N_1(\boldsymbol{\theta}_{T,1} + \boldsymbol{\theta}_{T,2}\rho + \boldsymbol{\theta}_{T,3}M + \boldsymbol{\theta}_{T,4}\rho^2 + \boldsymbol{\theta}_{T,5}\rho M + \boldsymbol{\theta}_{T,6}M^2 +\\ &\boldsymbol{\theta}_{T,7}\rho^3 + \boldsymbol{\theta}_{T,8}\rho^2 M + \boldsymbol{\theta}_{T,9}\rho M^2 + \boldsymbol{\theta}_{T,10}M^3 + \boldsymbol{\theta}_{T,11}\rho^4 +\\ &\boldsymbol{\theta}_{T,12}\rho^3 M + \boldsymbol{\theta}_{T,13}\rho^2 M^2 + \boldsymbol{\theta}_{T,14}\rho M^3 + \boldsymbol{\theta}_{T,15}M^4).\end{aligned}$$

Mattingly's model [Mattingly et al., 1992]:

$$T = N_1(\boldsymbol{\theta}_{T,1}\rho + \boldsymbol{\theta}_{T,2}\rho M^3).$$

Feature Selection

Our model:

$$\begin{aligned} T = &N_1(\boldsymbol{\theta}_{T,1} + \boldsymbol{\theta}_{T,2}\rho + \boldsymbol{\theta}_{T,3}M + \boldsymbol{\theta}_{T,4}\rho^2 + \boldsymbol{\theta}_{T,5}\rho M + \boldsymbol{\theta}_{T,6}M^2 + \\ &\boldsymbol{\theta}_{T,7}\rho^3 + \boldsymbol{\theta}_{T,8}\rho^2 M + \boldsymbol{\theta}_{T,9}\rho M^2 + \boldsymbol{\theta}_{T,10}M^3 + \boldsymbol{\theta}_{T,11}\rho^4 + \\ &\boldsymbol{\theta}_{T,12}\rho^3 M + \boldsymbol{\theta}_{T,13}\rho^2 M^2 + \boldsymbol{\theta}_{T,14}\rho M^3 + \boldsymbol{\theta}_{T,15}M^4). \end{aligned}$$

Mattingly's model [Mattingly et al., 1992]:

$$T = N_1(\boldsymbol{\theta}_{T,1}\rho + \boldsymbol{\theta}_{T,2}\rho M^3).$$

$\Rightarrow$ <u>High risk of overfitting</u>

Feature Selection

Our (sparse) model:

$$
\begin{aligned}
T =& N_1(\boldsymbol{\theta}_{T,1} + \theta_{T,2}\rho + \boldsymbol{\theta}_{T,3}M + \theta_{T,4}\rho^2 + \theta_{T,5}\rho M + \boldsymbol{\theta}_{T,6}M^2 + \\
& \theta_{T,7}\rho^3 + \theta_{T,8}\rho^2 M + \theta_{T,9}\rho M^2 + \theta_{T,10}M^3 + \boldsymbol{\theta}_{T,11}\rho^4 + \\
& \boldsymbol{\theta}_{T,12}\rho^3 M + \theta_{T,13}\rho^2 M^2 + \theta_{T,14}\rho M^3 + \theta_{T,15}M^4).
\end{aligned}
$$

Mattingly's model [Mattingly et al., 1992]:

$$T = N_1(\boldsymbol{\theta}_{T,1}\rho + \boldsymbol{\theta}_{T,2}\rho M^3).$$

$\Rightarrow$ High risk of overfitting

Feature Selection

Our (sparse) model:

$$T = N_1(\boldsymbol{\theta}_{T,1} + \theta_{T,2}\rho + \boldsymbol{\theta}_{T,3}M + \theta_{T,4}\rho^2 + \theta_{T,5}\rho M + \boldsymbol{\theta}_{T,6}M^2 + \theta_{T,7}\rho^3 + \theta_{T,8}\rho^2 M + \theta_{T,9}\rho M^2 + \theta_{T,10}M^3 + \boldsymbol{\theta}_{T,11}\rho^4 + \boldsymbol{\theta}_{T,12}\rho^3 M + \theta_{T,13}\rho^2 M^2 + \theta_{T,14}\rho M^3 + \theta_{T,15}M^4).$$

Mattingly's model [Mattingly et al., 1992]:

$$T = N_1(\boldsymbol{\theta}_{T,1}\rho + \boldsymbol{\theta}_{T,2}\rho M^3).$$

Sparse models are:

- Less susceptible to overfitting,
- More compliant with physical models,
- More interpretable,
- Lighter/Faster.

Block-sparse Lasso

Lasso [Tibshirani, 1994]: $\{(X_i, Y_i)\}_{i=1}^N \subset \mathbb{R}^{d+1}$ i.i.d sample,

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \lambda \|\boldsymbol{\theta}\|_1 .$$

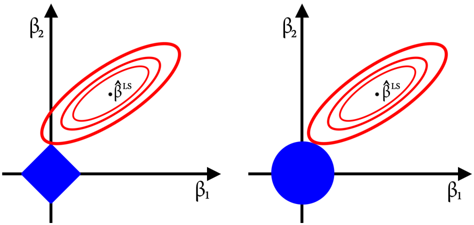


Figure: [1]Sparsity induced by L^1 norm in Lasso.

Source : Wikipedia, Lasso(statistics)

Block-sparse Lasso

$$\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(Y_{k,i} - X_{c,k,i} \cdot \boldsymbol{\theta}_c - X_{k,i} \cdot \boldsymbol{\theta}_k \right)^2 + \lambda_c \|\boldsymbol{\theta}_c\|_1 + \sum_{k=1}^{K} \lambda_k \|\boldsymbol{\theta}_k\|_1$$

Block-sparse Lasso

Block-sparse structure preserved $\rightsquigarrow$ **Equivalent to Lasso problem**

$$\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(Y_{k,i} - X_{c,k,i} \cdot \boldsymbol{\theta}_c - X_{k,i} \cdot \boldsymbol{\theta}_k \right)^2 + \lambda_c \|\boldsymbol{\theta}_c\|_1 + \sum_{k=1}^{K} \lambda_k \|\boldsymbol{\theta}_k\|_1$$

Block-sparse Lasso

Block-sparse structure preserved $\rightsquigarrow$ **Equivalent to Lasso problem**

$$\min_{\beta} \sum_{i=1}^{N} \| Y_i - B_i \beta \|_2^2 + \lambda_c \|\beta\|_1$$

with $\beta = (\boldsymbol{\theta}_c, \frac{\lambda_1}{\lambda_c}\boldsymbol{\theta}_1, \ldots, \frac{\lambda_K}{\lambda_c}\boldsymbol{\theta}_K) \in \mathbb{R}^p$, $p = d_c + \sum_{k=1}^{K} d_k$, $Y_i \in \mathbb{R}^K$ and $B_i \in \mathbb{R}^{K \times p}$.

Block-sparse Lasso

Block-sparse structure preserved $\rightsquigarrow$ **Equivalent to Lasso problem**

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \| Y_i - X_i \boldsymbol{\theta} \|_2^2 + \lambda_c \|\boldsymbol{\theta}\|_1$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_c, \quad \boldsymbol{\theta}_1, \dots, \quad \boldsymbol{\theta}_K) \in \mathbb{R}^p$, $p = d_c + \sum_{k=1}^{K} d_k$,
$Y_i \in \mathbb{R}^K$ and $X_i \in \mathbb{R}^{K \times p}$,
In practice, we choose $\lambda_k = \lambda_c$, for all $k = 1, \dots, 3$ and

$$X_i = \begin{pmatrix} X_{T1,i}^\top & -X_{D,i}^\top & 0 & 0 \\ X_{T2,i}^\top & 0 & X_{L,i}^\top & 0 \\ X_{T,i}^\top & 0 & 0 & X_{lspm,i}^\top \end{pmatrix}, \qquad Y_i = \begin{pmatrix} Y_{1,i} \\ Y_{2,i} \\ Y_{3,i} \end{pmatrix}$$

Bootstrap implementation

High correlations between features...

Bootstrap implementation

High correlations between features...
$\Rightarrow$ **Inconsistent selections via the lasso !**

Bootstrap implementation

High correlations between features...
$\Rightarrow$ **Inconsistent selections via the lasso !**

Bolasso - Bach [2008]

Require: training data $\mathcal{T} = \{(X_i, Y_i)\}_{i=1}^{N} \subset \mathbb{R}^{K\times(K+1)} \times \mathbb{R}^{K}$,
number of bootstrap replicates b,
L^1 penalty parameter λ_c,

```
for k = 1 to b do
    Generate bootstrap sample T_k,
    Compute Block sparse Lasso estimate θ̂^k from T_k,
    Compute support J_k = {j, θ̂_j^k ≠ 0},
end for
Compute intersection J = ∩_{k=1}^{b} J_k,
Compute θ̂_J from selected features using Least-Squares.
```

for $k = 1$ **to** b **do**
  Generate bootstrap sample $\mathcal{T}_k$,
  Compute Block sparse Lasso estimate $\hat{\boldsymbol{\theta}}^k$ from $\mathcal{T}_k$,
  Compute support $J_k = \{j, \hat{\boldsymbol{\theta}}_j^k \neq 0\}$,
end for
Compute intersection $J = \bigcap_{k=1}^{b} J_k$,
Compute $\hat{\boldsymbol{\theta}}_J$ from selected features using Least-Squares.

Bootstrap implementation

High correlations between features...
$\Rightarrow$ Inconsistent selections via the lasso !

Bolasso - Bach [2008]

Require: training data $\mathcal{T} = \{(X_i, Y_i)\}_{i=1}^N \subset \mathbb{R}^{K\times(K+1)} \times \mathbb{R}^K$,
number of bootstrap replicates b,
L^1 penalty parameter λ_c,

1: **for** $k = 1$ **to** b **do**
2: Generate bootstrap sample $\mathcal{T}_k$,
3: Compute Block sparse Lasso estimate $\hat{\boldsymbol{\theta}}^k$ from $\mathcal{T}_k$,
4: Compute support $J_k = \{j, \hat{\boldsymbol{\theta}}_j^k \neq 0\}$,
5: **end for**
6: Compute intersection $J = \bigcap_{k=1}^b J_k$,
7: Compute $\hat{\boldsymbol{\theta}}_J$ from selected features using Least-Squares.

- Consistency even under high correlations proved in Bach [2008],

Bootstrap implementation

High correlations between features...
$\Rightarrow$ Inconsistent selections via the lasso !

Bolasso - Bach [2008]

Require: training data $\mathcal{T} = \{(X_i, Y_i)\}_{i=1}^N \subset \mathbb{R}^{K\times(K+1)} \times \mathbb{R}^K$,
number of bootstrap replicates b,
L^1 penalty parameter λ_c,

for $k = 1$ **to** b **do**
    Generate bootstrap sample $\mathcal{T}_k$,
    Compute Block sparse Lasso estimate $\hat{\boldsymbol{\theta}}^k$ from $\mathcal{T}_k$,
    Compute support $J_k = \{j, \hat{\boldsymbol{\theta}}_j^k \neq 0\}$,
end for
Compute intersection $J = \bigcap_{k=1}^b J_k$,
Compute $\hat{\boldsymbol{\theta}}_J$ from selected features using Least-Squares.

- Consistency even under high correlations proved in Bach [2008],
- Efficient implementations exist: LARS [Efron et al., 2004].

Problem with intra-group correlations

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \|Y_i - X_i\boldsymbol{\theta}\|_2^2 + \lambda_c \|\boldsymbol{\theta}\|_1 \Rightarrow \hat{\boldsymbol{\theta}}_{\boldsymbol{T}} = \hat{\boldsymbol{\theta}}_{\boldsymbol{Isp}} = \mathbf{0}!$$

Problem with intra-group correlations

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \| Y_i - X_i \boldsymbol{\theta} \|_2^2 + \lambda_c \| \boldsymbol{\theta} \|_1 \Rightarrow \hat{\boldsymbol{\theta}}_{\boldsymbol{T}} = \hat{\boldsymbol{\theta}}_{\boldsymbol{Isp}} = \mathbf{0}!$$

$$\begin{cases} Y_1 = X_{T1} \cdot \boldsymbol{\theta}_T - \quad X_D \cdot \boldsymbol{\theta}_D \quad + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \boldsymbol{\theta}_T + \quad X_L \cdot \boldsymbol{\theta}_L \quad + \varepsilon_2 \\ \color{red}{0 \; = \; X_T \cdot \boldsymbol{\theta}_T \; + X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} + \varepsilon_3} \end{cases}$$

Problem with intra-group correlations

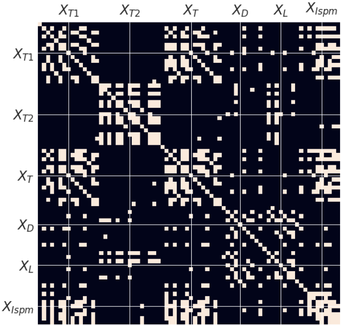


Figure: Features correlations higher than 0.9 in absolute value in white.

PROBLEM WITH INTRA-GROUP CORRELATIONS

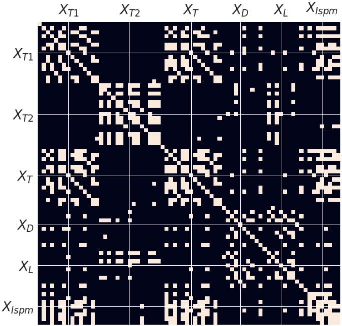


FIGURE: Features correlations higher than 0.9 in absolute value in white.

$\Rightarrow \boldsymbol{\theta} \mapsto \sum_{i=1}^{N} \| Y_i - X_i \boldsymbol{\theta} \|_2^2$ not injective...

Ill-posed problem !

Problem with intra-group correlations

$$\begin{cases} Y_1 &= X_{T1} \cdot \boldsymbol{\theta}_T &- X_D \cdot \boldsymbol{\theta}_D &+\varepsilon_1 \\ Y_2 &= X_{T2} \cdot \boldsymbol{\theta}_T &+ X_L \cdot \boldsymbol{\theta}_L &+\varepsilon_2 \\ 0 &= X_T \cdot \boldsymbol{\theta}_T &+ X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} &+\varepsilon_3 \\ & & & \end{cases}$$

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \| Y_i - X_i \boldsymbol{\theta} \|_2^2 + \lambda_c \|\boldsymbol{\theta}\|_1$$

Prior model $\tilde{I}_{sp}$ from Roux [2005] $\rightsquigarrow$ $\tilde{I}_{sp,i} = \tilde{I}_{sp}(\boldsymbol{u}_i, \boldsymbol{x}_i)$, $i = 1, \ldots, N$.

Problem with intra-group correlations

$$\begin{cases} Y_1 = X_{T1} \cdot \boldsymbol{\theta}_T - X_D \cdot \boldsymbol{\theta}_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \boldsymbol{\theta}_T + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\ 0 = X_T \cdot \boldsymbol{\theta}_T + X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} + \varepsilon_3 \end{cases}$$

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left(\| Y_i - X_i \boldsymbol{\theta} \|_2^2 + \lambda_t \| \tilde{I}_{sp,i} - X_{Isp,i} \cdot \boldsymbol{\theta}_{Isp} \|_2^2 \right) + \lambda_c \|\boldsymbol{\theta}\|_1$$

Prior model $\tilde{I}_{sp}$ from Roux [2005] $\rightsquigarrow \tilde{I}_{sp,i} = \tilde{I}_{sp}(\boldsymbol{u}_i, \boldsymbol{x}_i),\ i = 1, \ldots, N.$

Problem with intra-group correlations

$$\begin{cases} Y_1 &= X_{T1} \cdot \boldsymbol{\theta}_T &- & X_D \cdot \boldsymbol{\theta}_D & +\varepsilon_1 \\ Y_2 &= X_{T2} \cdot \boldsymbol{\theta}_T &+ & X_L \cdot \boldsymbol{\theta}_L & +\varepsilon_2 \\ 0 &= X_T \cdot \boldsymbol{\theta}_T &+ & X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} & +\varepsilon_3 \\ \sqrt{\lambda_t}\tilde{I}_{sp} &= \sqrt{\lambda_t} X_{Isp} \cdot \boldsymbol{\theta}_{Isp} & & & +\varepsilon_4 \end{cases}$$

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left(\| Y_i - X_i \boldsymbol{\theta} \|_2^2 + \lambda_t \| \tilde{I}_{sp,i} - X_{Isp,i} \cdot \boldsymbol{\theta}_{Isp} \|_2^2 \right) + \lambda_c \| \boldsymbol{\theta} \|_1$$

Prior model $\tilde{I}_{sp}$ from Roux [2005] $\rightsquigarrow$ $\tilde{I}_{sp,i} = \tilde{I}_{sp}(\boldsymbol{u}_i, \boldsymbol{x}_i)$, $i = 1, \dots, N$.

Problem with intra-group correlations

$$\begin{cases} Y_1 & = & X_{T1} \cdot \boldsymbol{\theta}_T & - & X_D \cdot \boldsymbol{\theta}_D & +\varepsilon_1 \\ Y_2 & = & X_{T2} \cdot \boldsymbol{\theta}_T & + & X_L \cdot \boldsymbol{\theta}_L & +\varepsilon_2 \\ 0 & = & X_T \cdot \boldsymbol{\theta}_T & + & X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} & +\varepsilon_3 \\ \sqrt{\lambda_t}\tilde{I}_{sp} & = & \sqrt{\lambda_t} X_{Isp} \cdot \boldsymbol{\theta}_{Isp} & & & +\varepsilon_4 \end{cases}$$

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \| \tilde{Y}_i - \tilde{X}_i \boldsymbol{\theta} \|_2^2 + \lambda_c \|\boldsymbol{\theta}\|_1$$

$$\tilde{Y}_i = \begin{pmatrix} Y_{1,i} \\ Y_{2,i} \\ 0 \\ \sqrt{\lambda_t}\tilde{I}_{sp,i} \end{pmatrix}, \quad \tilde{X}_i = \begin{pmatrix} X_{T1,i}^\top & -X_{D,i}^\top & 0 & 0 \\ X_{T2,i}^\top & 0 & X_{L,i}^\top & 0 \\ X_{T,i}^\top & 0 & 0 & X_{Ispm,i}^\top \\ 0 & 0 & 0 & \sqrt{\lambda_t} X_{Isp,i}^\top \end{pmatrix},$$

Prior model $\tilde{I}_{sp}$ from Roux [2005] $\rightsquigarrow \tilde{I}_{sp,i} = \tilde{I}_{sp}(\boldsymbol{u}_i, \boldsymbol{x}_i)$, $i = 1, \ldots, N$.

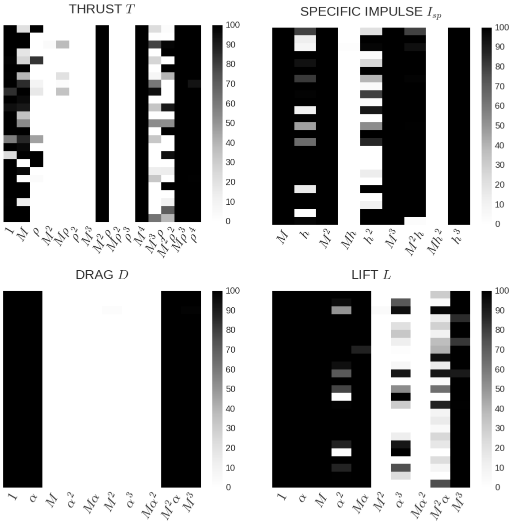
Feature selection results

- 25 different B737-800,
- 10 471 flights = 8 261 619 observations,

Feature selection results

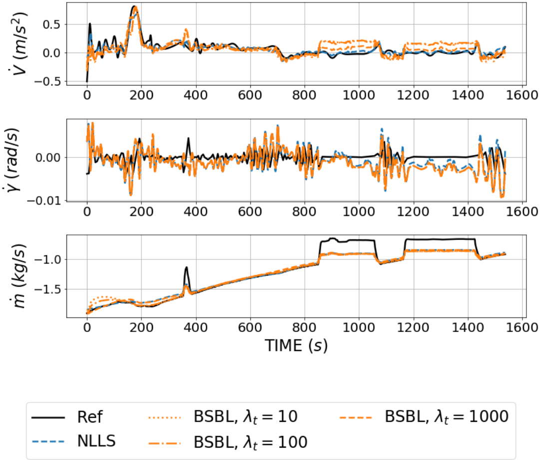
- 25 different B737-800,
- 10 471 flights = 8 261 619 observations,
- Block sparse Bolasso used for T, D, L and I_{sp},
- We expect similar model structures,

FEATURE SELECTION RESULTS

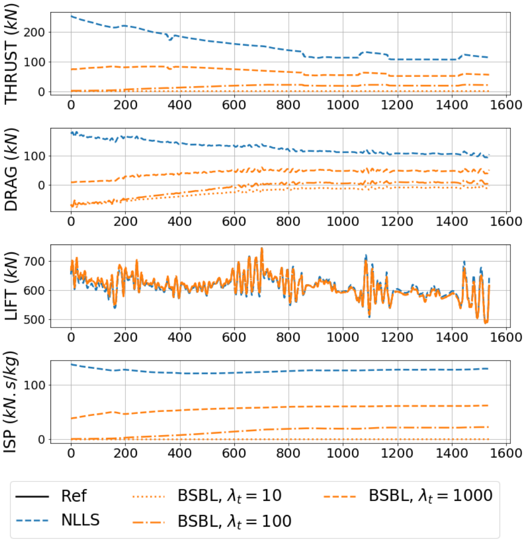


Feature selection results for the thrust, drag, lift and specific impulse models.

Accuracy of dynamics predictions



Realism of hidden elements



Flight resimulation

Flight resimulation

- Last assessment criterion $=$ static;

Flight resimulation

- Last assessment criterion = static;
- Does not incorporate the fact that the observations are time dependent;

Flight resimulation

- Last assessment criterion = static;
- Does not incorporate the fact that the observations are time dependent;
- Does not take into account the goal of optimally controlling the aircraft system.

Flight resimulation

- Last assessment criterion = static;
- Does not incorporate the fact that the observations are time dependent;
- Does not take into account the goal of optimally controlling the aircraft system.

Another possible dynamic criterion:

$$\min_{(\boldsymbol{x},\boldsymbol{u})} \int_{t_0}^{t_n} \left(\|\boldsymbol{u}(t) - \boldsymbol{u}_{test}(t)\|_{\boldsymbol{u}}^2 + \|\boldsymbol{x}(t) - \boldsymbol{x}_{test}(t)\|_{\boldsymbol{x}}^2 \right) dt$$

$$\text{s.t.} \quad \dot{\boldsymbol{x}}(t) = g(\boldsymbol{x}(t), \boldsymbol{u}(t), \hat{\boldsymbol{\theta}}),$$

where $\|\cdot\|_{\boldsymbol{u}}$, $\|\cdot\|_{\boldsymbol{x}}$ denote scaling norms.

Flight resimulation

- Last assessment criterion = static;
- Does not incorporate the fact that the observations are time dependent;
- Does not take into account the goal of optimally controlling the aircraft system.

Another possible dynamic criterion:

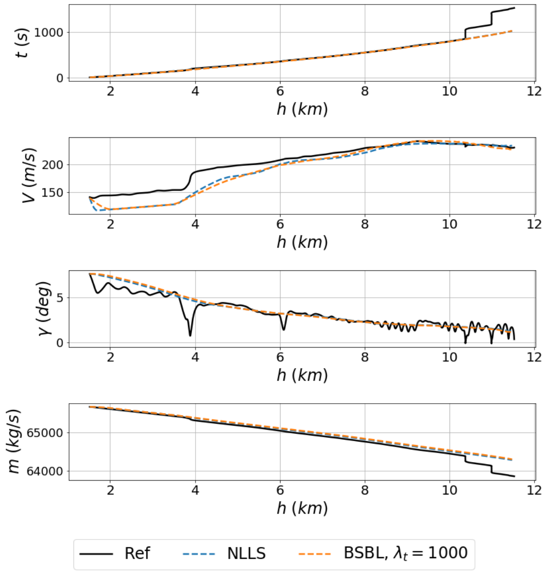
$$\min_{(\boldsymbol{x},\boldsymbol{u})} \int_{t_0}^{t_n} \left(\|\boldsymbol{u}(t) - \boldsymbol{u}_{test}(t)\|_{\boldsymbol{u}}^2 + \|\boldsymbol{x}(t) - \boldsymbol{x}_{test}(t)\|_{\boldsymbol{x}}^2 \right) dt$$

$$\text{s.t.} \quad \dot{\boldsymbol{x}}(t) = g(\boldsymbol{x}(t), \boldsymbol{u}(t), \hat{\boldsymbol{\theta}}),$$

where $\|\cdot\|_{\boldsymbol{u}}$, $\|\cdot\|_{\boldsymbol{x}}$ denote scaling norms.

For practical applications: $t \leftrightarrow h$

Flight resimulation



Flight resimulation

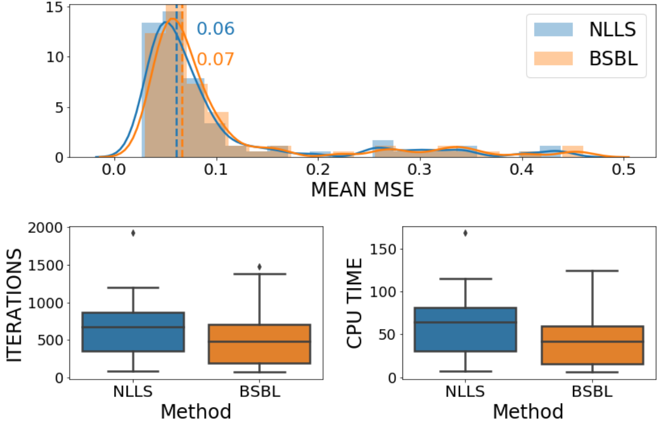


Figure: Distribution of the off-sample simulation error and boxplot of the optimization number of iterations and CPU time.

System Identification conclusion

1. Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),

System Identification conclusion

1. Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
2. Applicable to large amounts of data,

System Identification conclusion

1. Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
2. Applicable to large amounts of data,
3. Block-sparse estimators are proved to lead to consistent structured feature selection,

System Identification conclusion

1. Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
2. Applicable to large amounts of data,
3. Block-sparse estimators are proved to lead to consistent structured feature selection,
4. Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,

System Identification conclusion

1. Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
2. Applicable to large amounts of data,
3. Block-sparse estimators are proved to lead to consistent structured feature selection,
4. Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
5. Compared to regular Nonlinear Least-Squares:
   - Similar performances in accuracy and training time,

System Identification conclusion

1. Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
2. Applicable to large amounts of data,
3. Block-sparse estimators are proved to lead to consistent structured feature selection,
4. Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
5. Compared to regular Nonlinear Least-Squares:
   - Similar performances in accuracy and training time,
   - No initialization required,

System Identification conclusion

1. Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
2. Applicable to large amounts of data,
3. Block-sparse estimators are proved to lead to consistent structured feature selection,
4. Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
5. Compared to regular Nonlinear Least-Squares:
   - Similar performances in accuracy and training time,
   - No initialization required,
   - Light, interpretable and compact data-dependent models (more than 50% compression),

System Identification conclusion

1. Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
2. Applicable to large amounts of data,
3. Block-sparse estimators are proved to lead to consistent structured feature selection,
4. Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
5. Compared to regular Nonlinear Least-Squares:
   - Similar performances in accuracy and training time,
   - No initialization required,
   - Light, interpretable and compact data-dependent models (more than 50% compression),
   - Faster convergence when applied to control problems.

Trajectory acceptability

Trajectory acceptability

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

$$\text{s.t.} \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), & \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases}$$

Trajectory acceptability

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$
$$\text{s.t. } \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), \quad \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases} \tag{AOCP}$$

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

Trajectory acceptability

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(\boldsymbol{u}(t),\boldsymbol{x}(t))dt,$$
$$\text{s.t. } \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), \quad \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases} \tag{AOCP}$$

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

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

Trajectory acceptability

$$\begin{aligned}
&\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(\boldsymbol{u}(t),\boldsymbol{x}(t))dt, \\
&\text{s.t.} \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), & \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases}
\end{aligned} \tag{AOCP}$$

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

- Is $\hat{\boldsymbol{z}}$ inside the validity region of the dynamics model $\hat{g}$?
- Does it look like a real trajectory ?

Trajectory acceptability

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

$$\text{s.t.} \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), & \text{a.e. } t \in [0, t_f], \\ \text{Other constraints...} \end{cases}$$

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

- Is $\hat{\boldsymbol{z}}$ inside the validity region of the dynamics model $\hat{g}$?
- Does it look like a real trajectory ?

Pilots acceptance

Air Traffic Control[2]

Trajectory acceptability

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

$$\text{s.t.} \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), \quad \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases}$$

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

- Is $\hat{\boldsymbol{z}}$ inside the validity region of the dynamics model $\hat{g}$?
- Does it look like a real trajectory ?

Pilots acceptance

Air Traffic Control[2]

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

Optimized trajectory likelihood

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 likely is it to draw the optimized trajectory from the law of Z ?

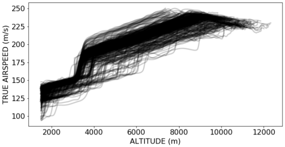
How to apply this to functional data?

Problem: Computation of probability densities in infinite dimensional space.

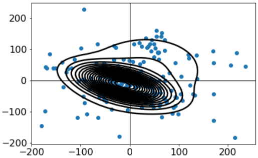
How to apply this to functional data?

Problem: Computation of probability densities in infinite dimensional space.

- Standard approach in Functional Data Analysis: use Functional Principal Component Analysis to decompose the data in a small number of coefficients



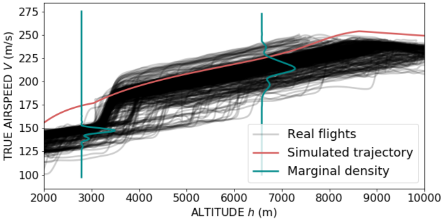
$\Rightarrow$



How to apply this to functional data?

Problem: Computation of probability densities in infinite dimensional space.

- Standard approach in Functional Data Analysis: use Functional Principal Component Analysis to decompose the data in a small number of coefficients
- **Or: we can use the marginal densities**



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 $Z_t = \boldsymbol{y}(t)$.

Mean marginal likelihood

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

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?

- 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

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

where $\psi : L^1(E, \mathbb{R}_+) \times \mathbb{R} \to [0; 1]$ is a continuous scaling map, because marginal densities may have really different shapes.

How do we aggregate the marginal likelihoods?

Possible scalings are the normalized density

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

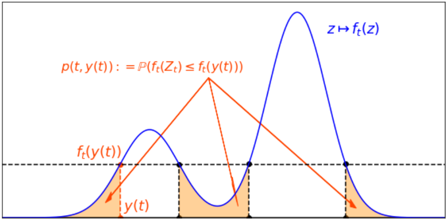
How do we aggregate the marginal likelihoods?

Possible scalings are the normalized density

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

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?

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

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

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

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?

How can we estimate marginal densities?

- In practice, the altitude plays the role of time, so we can't assume the same sampling for each trajectory;

How can we estimate marginal densities?

- In practice, the altitude plays the role of time, so we can't assume the same sampling for each trajectory;
- Assume sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ to be i.i.d. observations of a r.v. T, indep. Z;

How can we estimate marginal densities?

- In practice, the altitude plays the role of time, so we can't assume the same sampling for each trajectory;
- Assume sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ to be i.i.d. observations of a r.v. T, indep. Z;
- Our problem can be seen as a conditional probability density learning problem with $(X, Y) = (T, Z_T)$, where f_t is the density of $Z_t = (Z_T | T = t) = (Y|X)$.

How can we estimate marginal densities?

- In practice, the altitude plays the role of time, so we can't assume the same sampling for each trajectory;
- Assume sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ to be i.i.d. observations of a r.v. T, indep. Z;
- Our problem can be seen as a conditional probability density learning problem with $(X, Y) = (T, Z_T)$, where f_t is the density of $Z_t = (Z_T | T = t) = (Y|X)$.

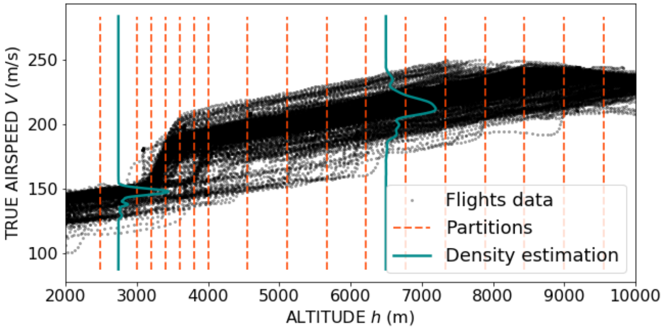
1. We can apply SOA conditional density estimation techniques, such as LS-CDE [Sugiyama et al., 2010],

How can we estimate marginal densities?

- In practice, the altitude plays the role of time, so we can't assume the same sampling for each trajectory;
- Assume sampling times $\{t_j^r : j = 1, \ldots, n; r = 1, \ldots, m\}$ to be i.i.d. observations of a r.v. T, indep. Z;
- Our problem can be seen as a conditional probability density learning problem with $(X, Y) = (T, Z_T)$, where f_t is the density of $Z_t = (Z_T | T = t) = (Y|X)$.

1. We can apply SOA conditional density estimation techniques, such as LS-CDE [Sugiyama et al., 2010],
2. **We can use a fine partitioning of the time domain.**

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

We denote by:

- $\Theta : \mathcal{S} \to L^1(E, \mathbb{R}_+)$ multivariate density estimation statistic,
- $\mathcal{S} = \{(z_k)_{k=1}^N \in E^N : N \in \mathbb{N}^*\}$ set of finite sequences,

Consistency

We denote by:

- $\Theta : \mathcal{S} \to L^1(E, \mathbb{R}_+)$ multivariate density estimation statistic,
- $\mathcal{S} = \{(z_k)_{k=1}^N \in E^N : N \in \mathbb{N}^*\}$ set of finite sequences,
- m the number of random curves;
- $\mathcal{T}_t^m$ subset of data points whose sampling times fall in the bin containing t;

Consistency

We denote by:

- $\Theta : \mathcal{S} \to L^1(E, \mathbb{R}_+)$ multivariate density estimation statistic,
- $\mathcal{S} = \{(z_k)_{k=1}^N \in E^N : N \in \mathbb{N}^*\}$ set of finite sequences,
- m the number of random curves;
- $\mathcal{T}_t^m$ subset of data points whose sampling times fall in the bin containing t;
- $\hat{f}_t^m := \Theta[\mathcal{T}_t^m]$ estimator trained using $\mathcal{T}_t^m$.

Consistency

Assumption 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

Assumption 1 - Positive time density
$\nu \in L^\infty(E, \mathbb{R}_+)$ density function of T, s.t.

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

Assumption 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

Assumption 1 - Positive time density

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

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

Assumption 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

- $\mathcal{G}$ 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

Theorem 1

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:

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

Consistency

Theorem 1

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:

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

Note that:

- $m \to \infty \neq N \to \infty$,

Consistency

Theorem 1

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:

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

Note that:

- $m \to \infty \neq N \to \infty$,
- Number of samples = random,

Consistency

Theorem 1

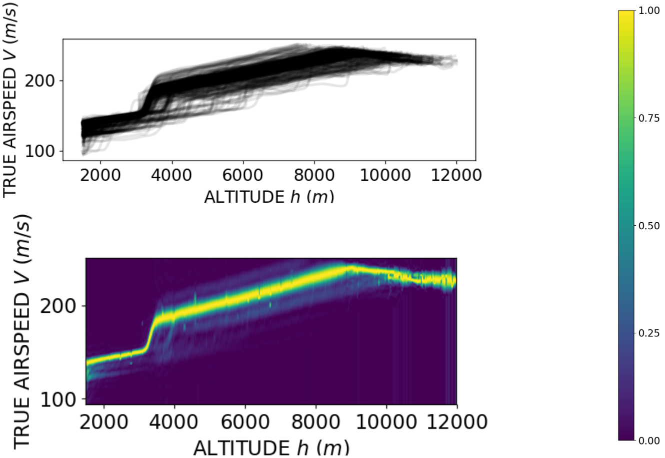
Under assumptions 1 to 4, for any $z \in E$ and $t \in \mathbb{T}$, $\hat{f}^m_{\ell^m(t)}(z)$ consistently approximates the marginal density $f_t(z)$ as the number of curves m grows:

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

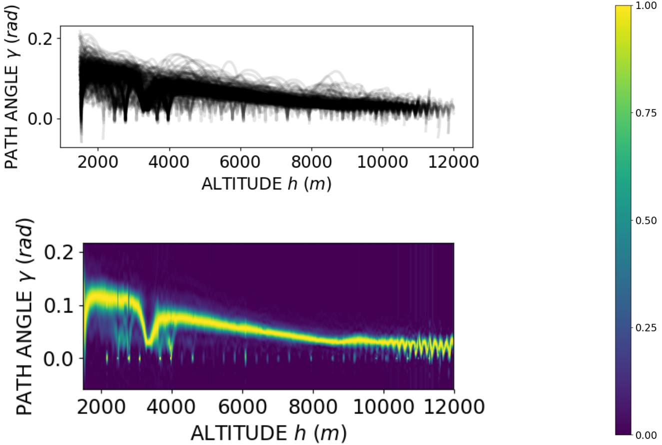
Note that:

- $m \to \infty \neq N \to \infty$,
- Number of samples = random,
- Training data not i.i.d.

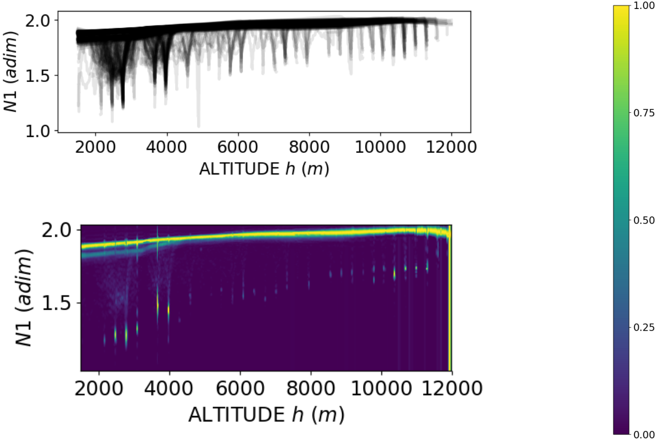
Marginal density estimation results



MARGINAL DENSITY ESTIMATION RESULTS



Marginal density estimation results



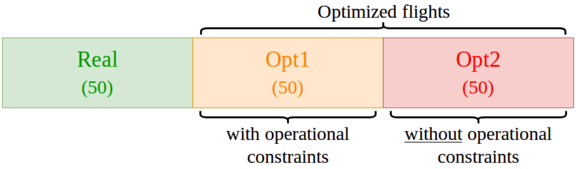
How good is it compared to other methods?

How good is it compared to other methods?

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

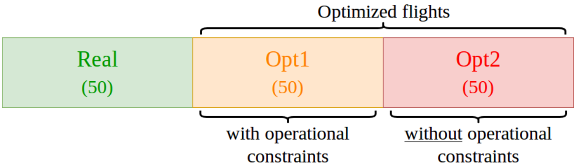
HOW GOOD IS IT COMPARED TO OTHER METHODS?

- 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

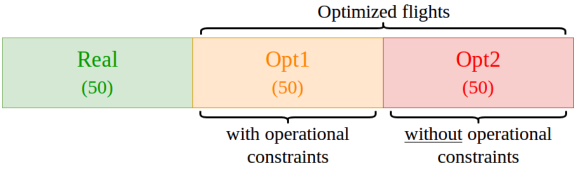


- Discrimination power comparison with (gmm-)FPCA and (integrated) LS-CDE:

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

How good is it compared to other methods?

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



- Discrimination power comparison with (gmm-)FPCA and (integrated) LS-CDE:

Var.	Estimated Likelihoods			Tr. Time
	Real	Opt1	Opt2	
MML	$\mathbf{0.63 \pm 0.07}$	$\mathbf{0.43 \pm 0.08}$	$\mathbf{0.13 \pm 0.02}$	5s
FPCA	0.16 ± 0.12	6.4E-03 $\pm$ 3.8E-03	3.6E-03 $\pm$ 5.4E-03	20s
LS-CDE	0.77 ± 0.05	0.68 ± 0.04	0.49 ± 0.06	14h

MML penalty

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

$$
\begin{aligned}
&\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(\boldsymbol{u}(t),\boldsymbol{x}(t))dt \\
&\text{s.t.} \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), \quad \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases}
\end{aligned}
\qquad (\qquad \text{AOCP})
$$

MML penalty

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

$$\begin{aligned} &\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(\boldsymbol{u}(t),\boldsymbol{x}(t))dt - \lambda\,\mathrm{MML}(Z,\boldsymbol{x}), \\ &\text{s.t.} \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), & \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases} \end{aligned} \qquad \text{(MML-AOCP)}$$

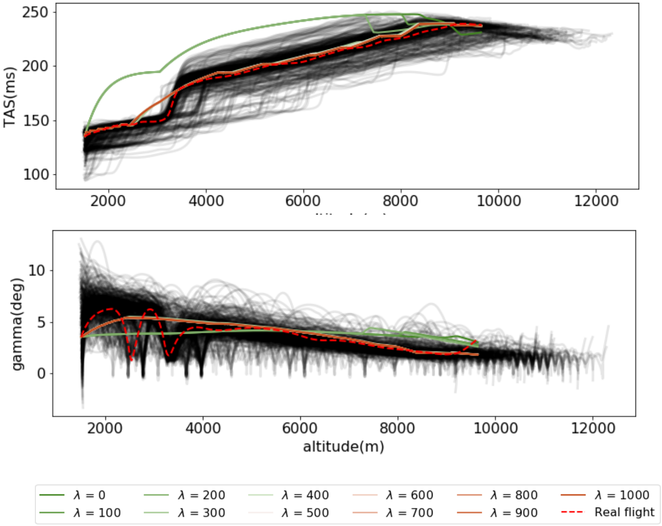
MML PENALTY

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

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathbb{X}\times\mathbb{U}} \int_0^{t_f} C(\boldsymbol{u}(t),\boldsymbol{x}(t))dt - \lambda\,\mathrm{MML}(Z,\boldsymbol{x}),$$
$$\text{s.t.} \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), & \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases} \tag{MML-AOCP}$$

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

PENALTY EFFECT



Trajectory acceptability conclusion

1. General probabilistic criterion using marginal densities to quantify the closeness between a curve and a set of random trajectories,

TRAJECTORY ACCEPTABILITY CONCLUSION

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

Trajectory acceptability conclusion

1. General probabilistic criterion using marginal densities to quantify the closeness between a curve and a set of 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,

Trajectory acceptability conclusion

1. General probabilistic criterion using marginal densities to quantify the closeness between a curve and a set of 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,

Trajectory acceptability conclusion

1. General probabilistic criterion using marginal densities to quantify the closeness between a curve and a set of 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 Adaptive Kernel and Gaussian mixture implementation,

Trajectory acceptability conclusion

1. General probabilistic criterion using marginal densities to quantify the closeness between a curve and a set of 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 Adaptive Kernel and Gaussian mixture implementation,
   - Showed that it can be used in optimal control problems to obtain solutions close to optimal, and still realistic.

THANK YOU FOR YOUR ATTENTION

REFERENCES

Anderson, R. M. and May, R. M. (1992). Infectious Diseases of Humans: Dynamics and Control. Oxford university press.

Bach, F. (2008). Bolasso: model consistent Lasso estimation through the bootstrap. In ICML, pages 33–40.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. JRSS-B, pages 1–38.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. The Annals of Statistics, 32:407–499.

FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. Biophysical journal, 1(6):445–466.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). A note on the Group Lasso and a Sparse group Lasso. arXiv:1001.0736.

Jategaonkar, R. V. (2006). Flight Vehicle System Identification: A Time Domain Methdology. AIAA.

Mattingly, J. D., Heiser, W. H., and Daley, D. H. (1992). Aircraft Engine Design. University Press.

Nagumo, J., Arimoto, S., and Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. Proceedings of the IRE, 50(10):2061–2070.

Obozinski, G., Taskar, B., and Jordan, M. I. (2006). Multi-task feature selection. In ICML-06 Workshop on Structural Knowledge Transfer for Machine Learning.

Roux, E. (2005). Pour une approche analytique de la dynamique du vol. PhD thesis, Supaero.

Sugiyama, M., Takeuchi, I., Suzuki, T., Kanamori, T., Hachiya, H., and Okanohara, D. (2010). Conditional density estimation via least-squares density ratio estimation. In AISTAT, pages 781–788.

Tibshirani, R. (1994). Regression shrinkage and selection via the Lasso. JRSS-B, 58:267–288.

Tikhonov, A. N. (1943). On the stability of inverse problems. In Doklady Akademii Nauk SSSR, volume 39, pages 195–198.

Yuan, M. and Lin, Y. (2005). Model selection and estimation in regression with grouped variables. JRSS-B, 68:49–67.

ACCURACY OF DYNAMICS PREDICTIONS

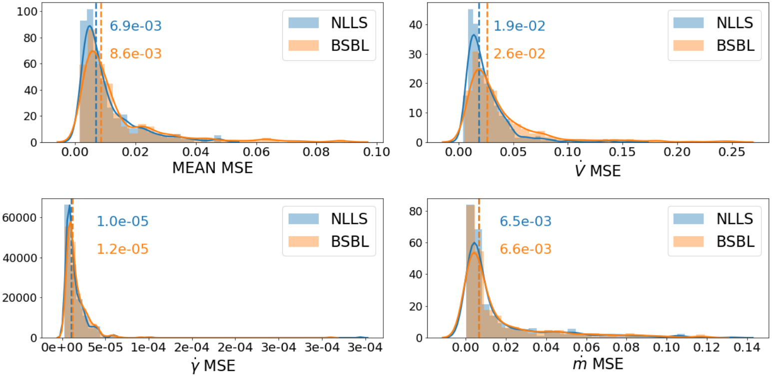


FIGURE: Leave-one-out off-sample errors distributions for nonlinear least-squares NLLS and block-sparse bolasso BSBL. Median errors are annotated and marked by dashed vertical lines.

STRUCTURED FEATURE SELECTION
STATE-OF-THE-ART

Other methods	Difference with Block-sparse Lasso
Group Lasso [Yuan and Lin, 2005]	Groups sparsity is fixed *a priori*,
Sparse Group Lasso [Friedman et al., 2010]	Sparsity induced <u>only</u> within group,
Multi-task Lasso [Obozinski et al., 2006]	Not same pattern for every task.

THEOREM (BOLASSO CONSISTENCY - BACH [2008])

For $\lambda = \lambda_0 N^{-\frac{1}{2}}$ *and* $\lambda_0 > 0$*, assume that*

(H1) *the cumulant generating functions* $\mathbb{E}\left[\exp(s\|X\|_2^2)\right]$ *and* $\mathbb{E}\left[\exp(s\|Y\|_2^2)\right]$ *are finite for some* $s > 0$*.*

(H2) *the joint matrix of second order moments* $Q = \mathbb{E}\left[XX^\top\right] \in \mathbb{R}^{p\times p}$ *is invertible.*

(H3) $\mathbb{E}\left[Y|X\right] = X \cdot \boldsymbol{\theta}$ *and* $\mathrm{Var}\left[Y|X\right] = \sigma^2$ *a.s. for some* $\boldsymbol{\theta} \in \mathbb{R}^p$ *and* $\sigma \in \mathbb{R}_+^*$*.*

Then, for any $b > 0$*, the probability that algorithm 1 does not exactly select the correct model has the following upper bound:*

$$\mathbb{P}\left[J \neq J^*\right] \leq bA_1 e^{-A_2 N} + A_3 \frac{\log N}{N^{1/2}} + A_4 \frac{\log b}{b},$$

where $A_1, A_2, A_3, A_4 > 0$*.*

Generalized Tikhonov Regularization of Isp

Equivalent to $\|\Gamma(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\|_2^2$ with $\Gamma_i = (\underbrace{0, \ldots, 0,}_{d_T + d_D + d_L} X_{\boldsymbol{Isp}}^\top)$ and

$$\Gamma_i \tilde{\boldsymbol{\theta}} = \tilde{I}_{sp,i}.$$

MML consistency for standard kernel estimator

Assumption 5
The function $(t, z) \in \mathbb{T} \times E \mapsto f_t(z)$ is $\mathcal{C}^4(E)$ in z and $\mathcal{C}^1(\mathbb{T})$ in t ; the Lipschitz constant of the function

$$t \mapsto \frac{d^2 f_t}{dz^2}(z) := f_t''(z)$$

is denoted by $L'' > 0$: for any $z \in E$ and $t_1, t_2 \in \mathbb{T}$,

$$|f_{t_1}''(z) - f_{t_2}''(z)| \leq L''|t_1 - t_2|.$$

MML consistency for standard kernel estimator

$$\sigma^2_{K_\sigma} = \int w^2 K_\sigma(w) dw = \sigma^2 \int w^2 K(w) dw = \sigma^2 \sigma^2_K,$$

$$\sigma^2_{K^2_\sigma} = \int w^2 K_\sigma(w)^2 dw = \sigma \int w^2 K(w)^2 dw = \sigma \sigma^2_{K^2},$$

$$R(K_\sigma) = \int K_\sigma(w)^2 dw = \frac{1}{\sigma} \int K(w)^2 dw = \frac{1}{\sigma} R(K).$$

Theorem 2

Under assumptions 1, 3 and 5, if $\hat{f}^m_{\ell^m(t)}$ is a KDE where the kernel K and the bandwidth $\sigma := \sigma_m$ are deterministic, such that $\sigma_K < \infty$, $\sigma_{K^2} < \infty$, $R(K) < \infty$ and if

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

then

$$\lim_{m\to\infty} \mathbb{E}\left[(\hat{f}^m_{\ell^m(t)}(z) - f_t(z))^2 \right] = 0.$$

Theorem 1 proof sketch

$$\lim_{m\to\infty} |f_t(z) - f^m_{\ell^m(t)}(z)| = 0.$$

$$\lim_{m\to\infty} \mathbb{P}(N^m_{r,\ell^m(t)} \leq 1) = 1, \quad r = 1, \ldots, m,$$

$$\forall M > 0, \quad \lim_{m\to\infty} \mathbb{P}\left(N^m_{\ell^m(t)} > M\right) = 1.$$

$$C_M := \{N^m_{\ell^m(t)} > M\} \bigcap_{r=1}^{m} \{N^m_{r,\ell^m(t)} \leq 1\}.$$

$$\forall M > 0, \quad \lim_{m\to\infty} \mathbb{P}\left(C_M\right) = 1.$$

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

Flight mechanics models

$$\rho = \frac{P}{R_s SAT}$$

$$SAT(h) = T_0 + \alpha_T h, \qquad SAT(TAT, M) = \frac{TAT}{1 + \frac{\lambda - 1}{2} M^2}$$

$$M = \frac{V}{V_{sound}} = \frac{V}{(\lambda R_s SAT)^{\frac{1}{2}}}$$

Consumption x Acceptability trade-off

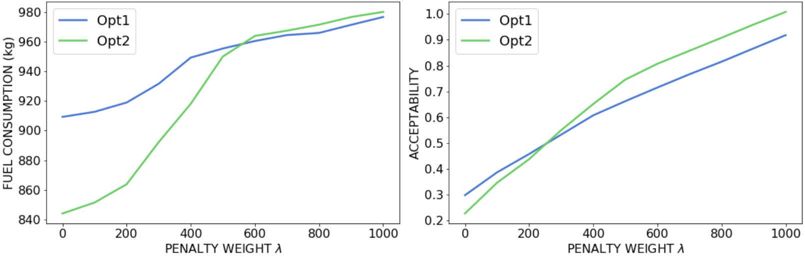
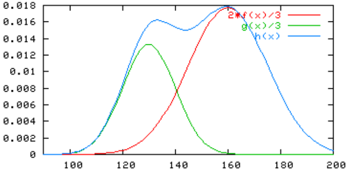


Figure: Average over 20 flights of the fuel consumption and MML score (called acceptability here) of optimized trajectories with varying MML-penalty weight λ.

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:

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

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

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

EM Algorithm

- 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 $i = 1, \ldots, N$,

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

Maximization:

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