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: Flight Radar

Motivation

FIGURE: World air traffic - source: Flight Radar

- 20 000 airplanes — 80 000 flights per day,

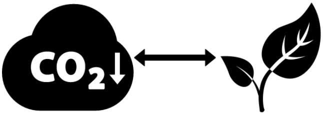
Motivation

Figure: World air traffic - source: Flight Radar

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

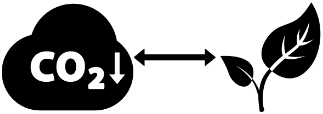
Motivation

- Most polluting means of transportation,



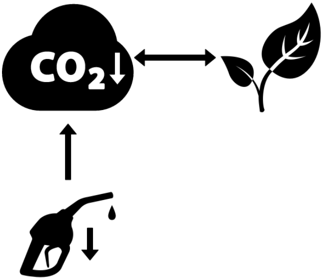
Motivation

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



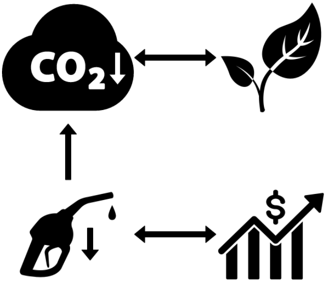
Motivation

- 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 ?

- New hardware ?

Motivation

How to tackle this problem ?

- ~~New hardware ?~~
- Better use of existing fleet,

MOTIVATION

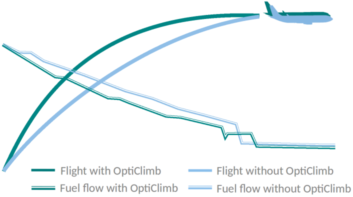
How to tackle this problem ?

- ~~New hardware ?~~
- Better use of existing fleet,
- Climb is most consuming flight phase...

Motivation

How to tackle this problem ?

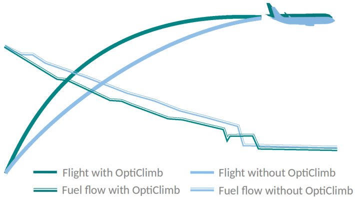
- ~~New hardware ?~~
- Better use of existing fleet,
- Climb is most consuming flight phase...
- Mostly rectilinear trajectories at full thrust,



Motivation

How to tackle this problem ?

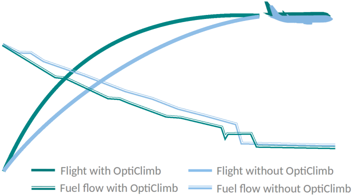
- ~~New hardware ?~~
- Better use of existing fleet,
- Climb is most consuming flight phase...
- Mostly rectilinear trajectories at full thrust,
- Thousands of variables recorded every second,



Motivation

How to tackle this problem ?

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

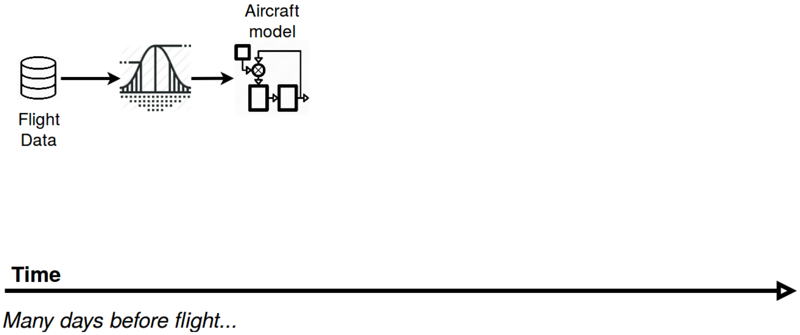


OPTI CLIMB
More savings, less CO_2

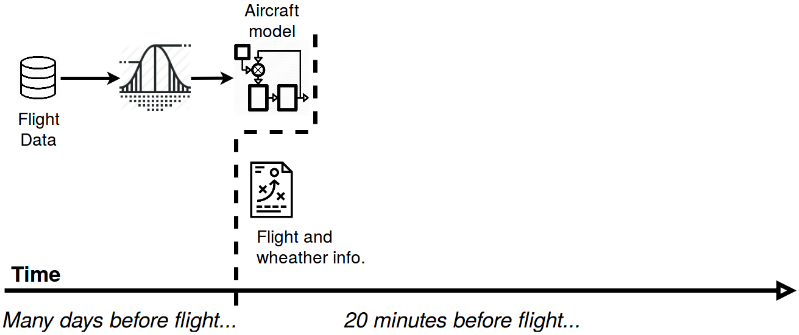
OptiClimb

OptiClimb

OptiClimb

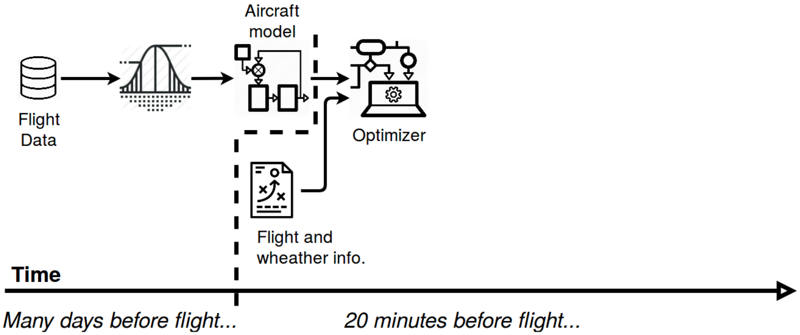


OPTICLIMB



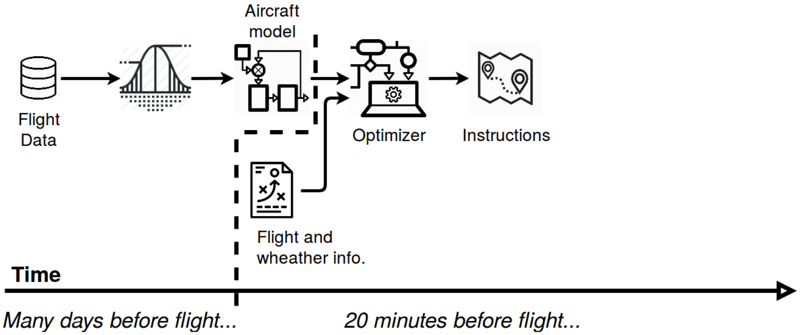
Aircraft
model
Flight
Data
Flight and
wheather info.
Time
Many days before flight...
20 minutes before flight...

OptiClimb



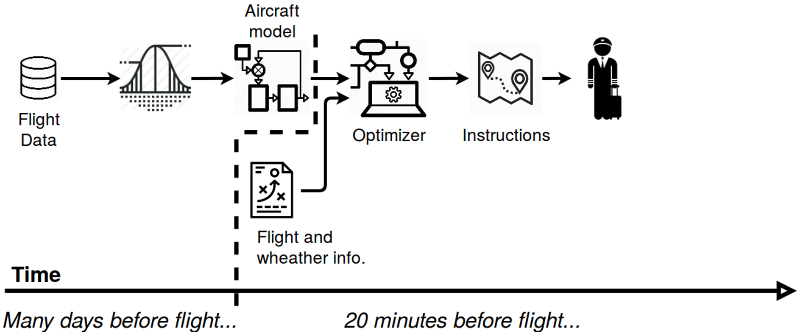
Aircraft
model
Flight
Data
Optimizer
Flight and
wheather info.
Time
Many days before flight...
20 minutes before flight...

OptiClimb

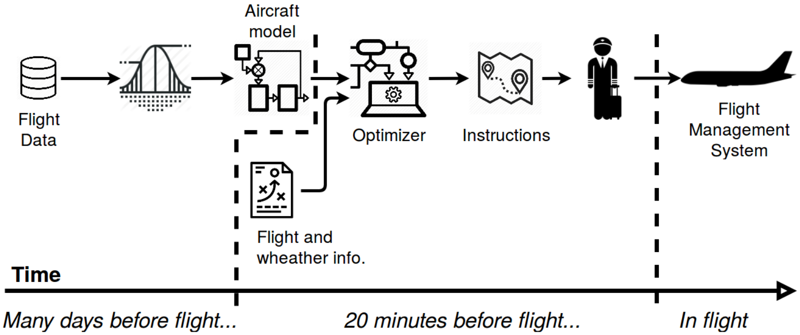


Aircraft
model
Flight
Data
Optimizer
Instructions
Flight and
wheather info.
Time
Many days before flight...
20 minutes before flight...

OptiClimb

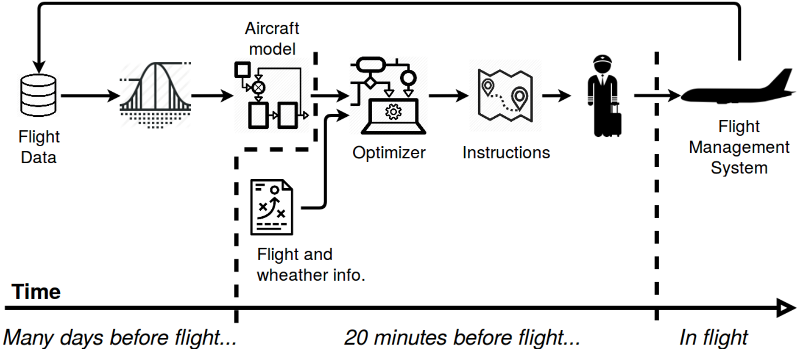

Aircraft
model
Flight
Data
Optimizer
Instructions
Flight and
wheather info.
Time
Many days before flight...
20 minutes before flight...

OptiClimb



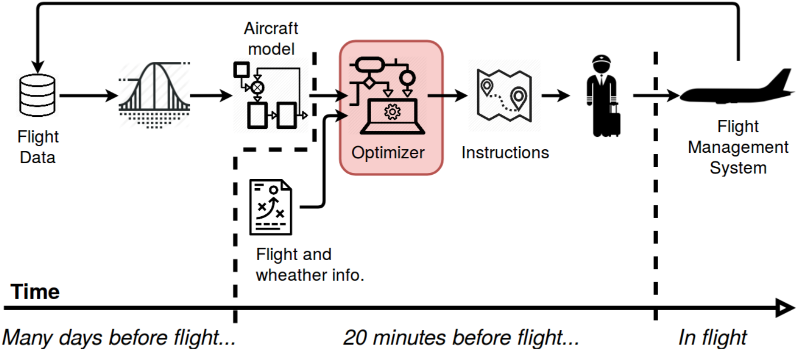
Aircraft model
Flight Data
Optimizer
Instructions
Flight Management System
Flight and wheather info.
Time
Many days before flight...
20 minutes before flight...
In flight

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

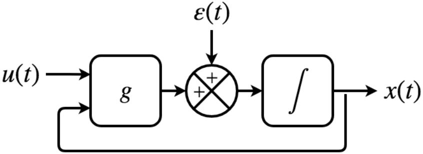
OptiClimb



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

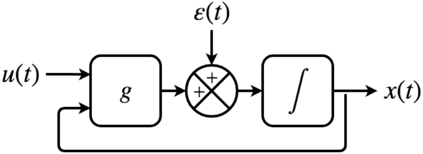
Trajectory Optimization



Trajectory Optimization

Dynamics:

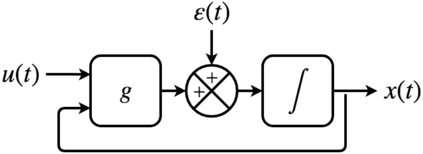
$$\dot{x}(t) = g(\boldsymbol{u}(t), \boldsymbol{x}(t)) + \varepsilon(t)$$



Trajectory Optimization

Dynamics:

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

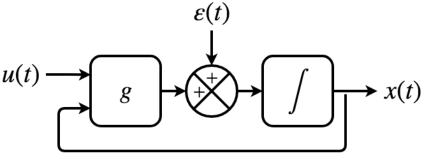


Optimization objective: $C(\boldsymbol{u}(t), \boldsymbol{x}(t))$

Trajectory Optimization

Dynamics:

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

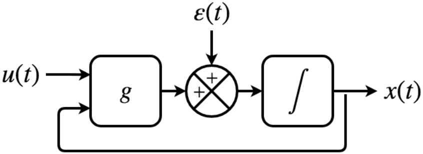


Optimization objective: $C(\boldsymbol{u}(t), \boldsymbol{x}(t)) \Longleftarrow$  ,

Trajectory Optimization

Dynamics:

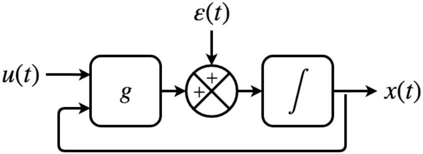
$$\dot{x}(t) = g(\boldsymbol{u}(t), \boldsymbol{x}(t)) + \varepsilon(t)$$



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

Dynamics:

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



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

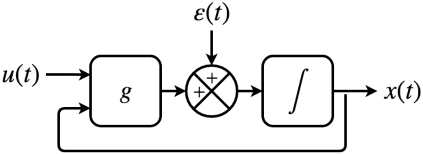
Flight constraints:

$$\left\{ \right.$$

Trajectory Optimization

Dynamics:

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



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

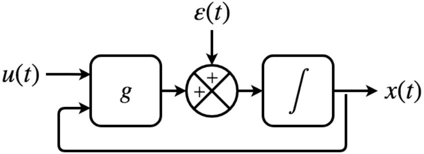
Flight constraints:

$$\begin{cases} \Phi(\boldsymbol{x}(0), \boldsymbol{x}(t_f)) \in K_\Phi & \text{Initial and final conditions} \\ & \end{cases}$$

Trajectory Optimization

Dynamics:

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



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

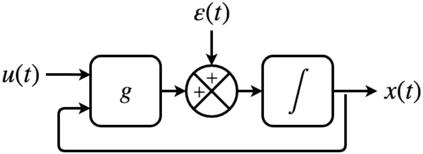
Flight constraints:

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

Trajectory Optimization

Dynamics:

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



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

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

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

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

Trajectory Optimization

Optimal Control Problem

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

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

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

Trajectory Optimization

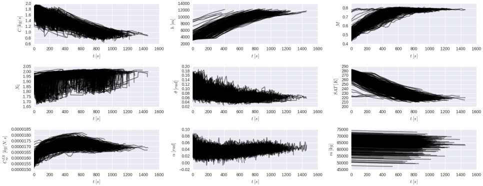
Optimal Control Problem

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

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

System Identification

Black box



QAR data

Trajectory Optimization

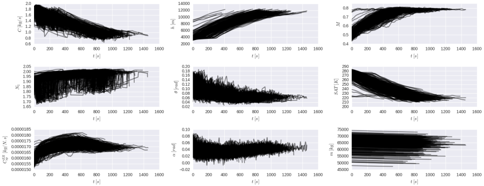
Optimal Control Problem

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

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

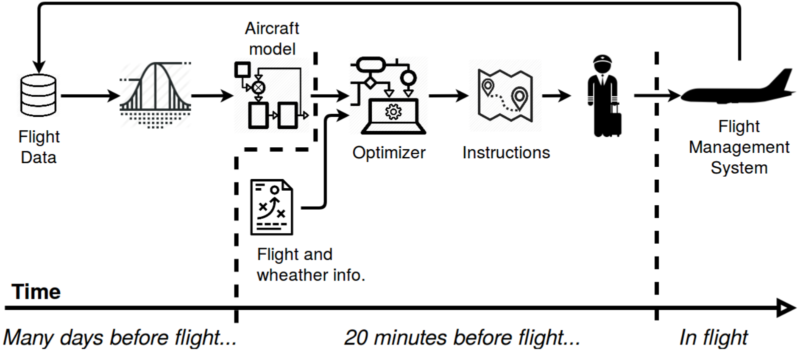
System Identification

Black box

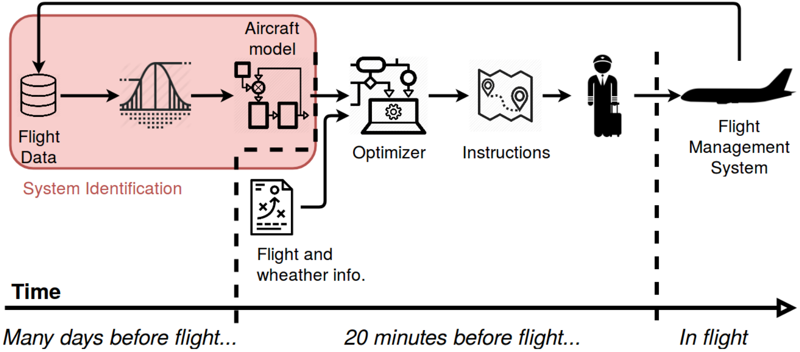


QAR data

System Identification

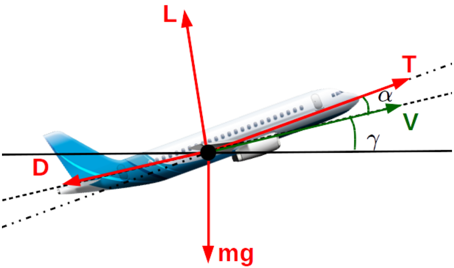


System Identification

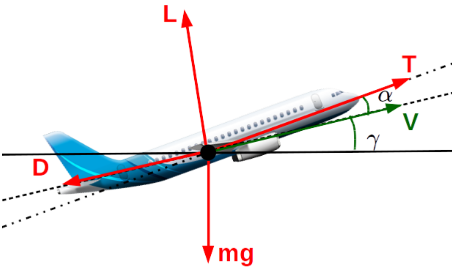


System Identification

Flight dynamics

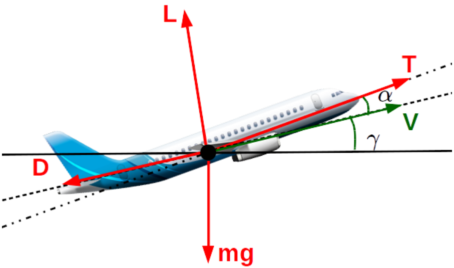


Flight dynamics



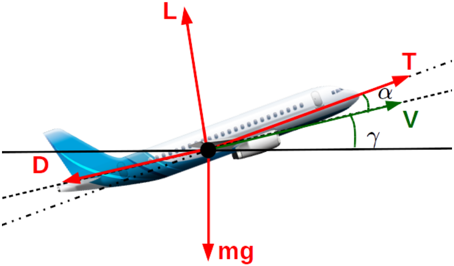
$$
\begin{cases}
\dot{h} = & V \sin\gamma \\
\dot{V} = & \dfrac{T\cos\alpha - D - mg\sin\gamma}{m} \\
\dot{\gamma} = & \dfrac{T\sin\alpha + L - mg\cos\gamma}{mV} \\
\dot{m} = & -\dfrac{T}{I_{sp}}
\end{cases}
$$

FLIGHT DYNAMICS



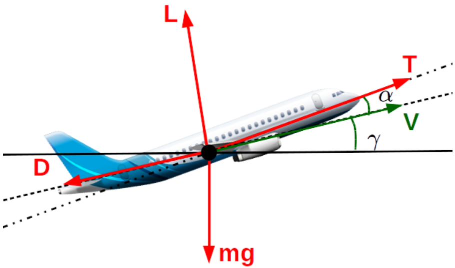
$$\begin{cases} \dot{h} = & V\sin\gamma + \dot{W}_z \\ \dot{V} = & \dfrac{T\cos\alpha - D - mg\sin\gamma - m\dot{W}_{xv}}{m} \\ \dot{\gamma} = & \dfrac{T\sin\alpha + L - mg\cos\gamma - m\dot{W}_{zv}}{mV} \\ \dot{m} = & -\dfrac{T}{I_{sp}} \end{cases}$$

FLIGHT DYNAMICS



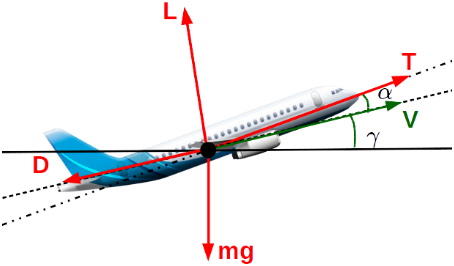
$$
\begin{cases}
\dot{h} = & V \sin\gamma + \dot{W}_z \\
\dot{V} = & \dfrac{T\cos\alpha - D - mg\sin\gamma - m\dot{W}_{xv}}{m} \\
\dot{\gamma} = & \dfrac{(T\sin\alpha + L)\cos\mu - mg\cos\gamma - m\dot{W}_{zv}}{mV} \\
\dot{m} = & -\dfrac{T}{I_{sp}}
\end{cases}
$$

Flight dynamics



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

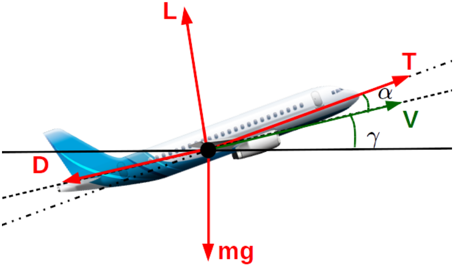
Flight dynamics



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

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

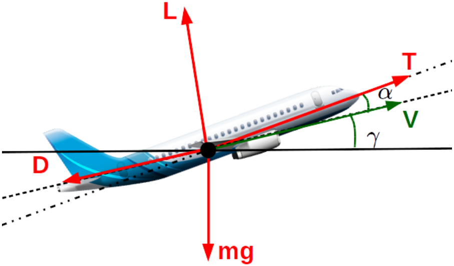
Flight dynamics



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

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

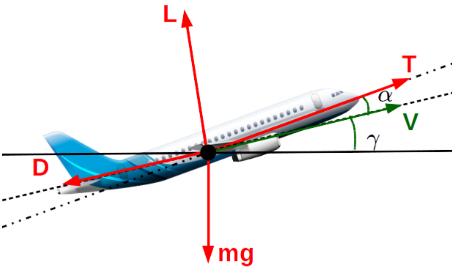
Flight dynamics



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

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

Flight dynamics



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

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

Physical models of nested functions

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

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

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

Physical models of nested functions

$$\begin{cases} T = X_T \cdot \boldsymbol{\theta}_T, \\ D = X_D \cdot \boldsymbol{\theta}_D, \\ L = X_L \cdot \boldsymbol{\theta}_L, \\ I_{sp} = X_{Isp} \cdot \boldsymbol{\theta}_{Isp}, \end{cases}$$

Physical models of nested functions

$$
\begin{cases}
T = X_T \cdot \boldsymbol{\theta}_T, \\
D = X_D \cdot \boldsymbol{\theta}_D, \\
L = X_L \cdot \boldsymbol{\theta}_L, \\
I_{sp} = X_{Isp} \cdot \boldsymbol{\theta}_{Isp},
\end{cases}
$$

$$
\begin{aligned}
X_T &= N_1(1, \rho, M, \rho^2, \rho M, M^2, \ldots) \\
X_D &= q(1, \alpha, M, \alpha^2, \alpha M, M^2, \ldots) \\
X_L &= q(1, \alpha, M, \alpha^2, \alpha M, M^2, \ldots) \\
X_{Isp} &= SAT(1, h, M, h^2, hM, M^2, \ldots)
\end{aligned}
$$

Physical models of nested functions

$$
\begin{cases}
T(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}_T) & = X_T \cdot \boldsymbol{\theta}_T, \\
D(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}_D) & = X_D \cdot \boldsymbol{\theta}_D, \\
L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}_L) & = X_L \cdot \boldsymbol{\theta}_L, \\
I_{sp}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}_{Isp}) & = X_{Isp} \cdot \boldsymbol{\theta}_{Isp},
\end{cases}
$$

$$
\begin{aligned}
X_T &= N_1(1, \rho, M, \rho^2, \rho M, M^2, \ldots) \\
X_D &= q(1, \alpha, M, \alpha^2, \alpha M, M^2, \ldots) \\
X_L &= q(1, \alpha, M, \alpha^2, \alpha M, M^2, \ldots) \\
X_{Isp} &= SAT(1, h, M, h^2, hM, M^2, \ldots)
\end{aligned}
$$

State-of-the-art

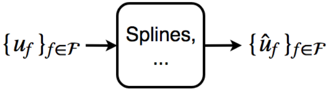
STATE-OF-THE-ART

- Output-Error Method

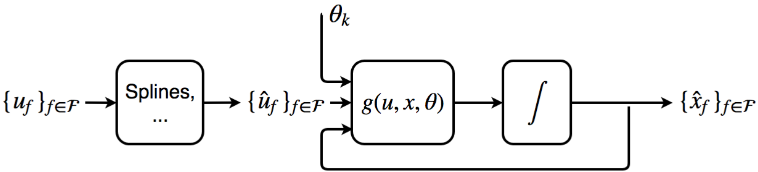
$\{u_f\}_{f \in \mathcal{F}}$

State-of-the-art

- Output-Error Method

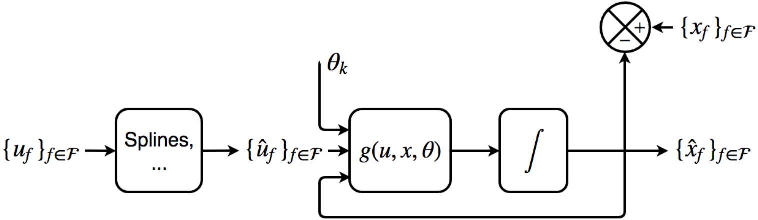


- Output-Error Method



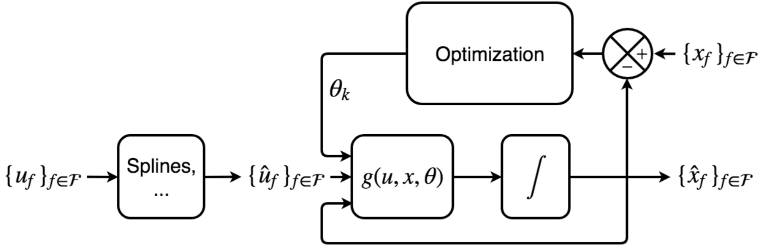
State-of-the-art

- Output-Error Method



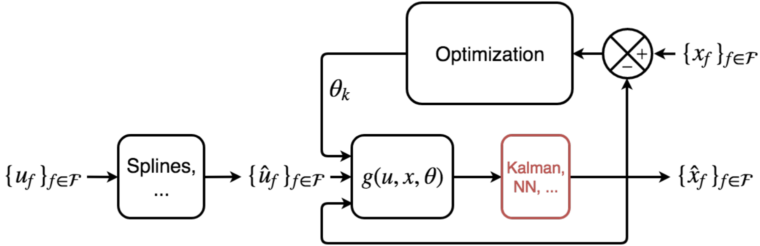
State-of-the-art

- Output-Error Method



State-of-the-art

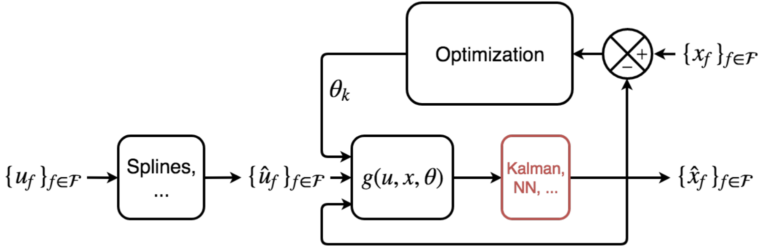
- Output-Error Method
- Filter-Error Method



State-of-the-art

- Output-Error Method
- Filter-Error Method

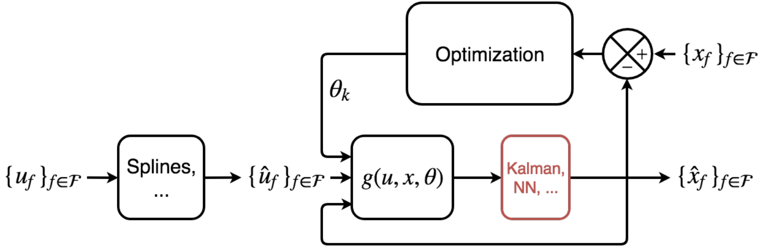
Not scalable to many trajectories



State-of-the-art

- Output-Error Method
- Filter-Error Method

} Not scalable to many trajectories



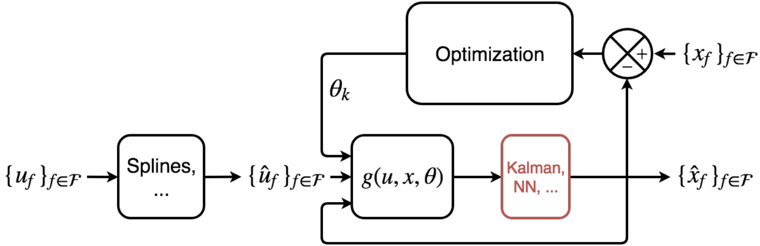
- Equation-Error Method

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

State-of-the-art

- Output-Error Method
- Filter-Error Method

Not scalable to many trajectories



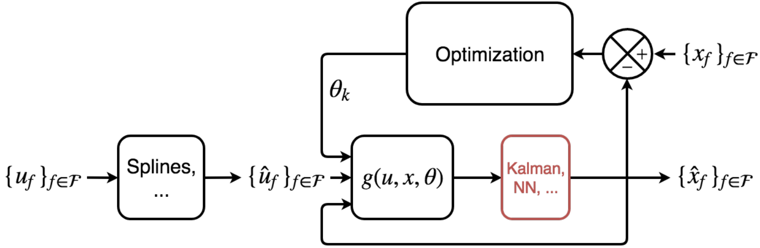
- Equation-Error Method

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

State-of-the-art

- Output-Error Method
- Filter-Error Method

} Not scalable to many trajectories



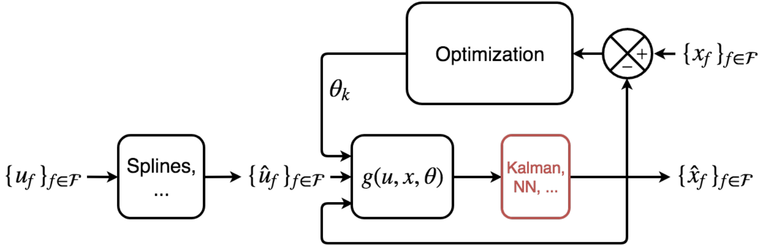
- Equation-Error Method

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

STATE-OF-THE-ART

- Output-Error Method
- Filter-Error Method

Not scalable to many trajectories



- Equation-Error Method

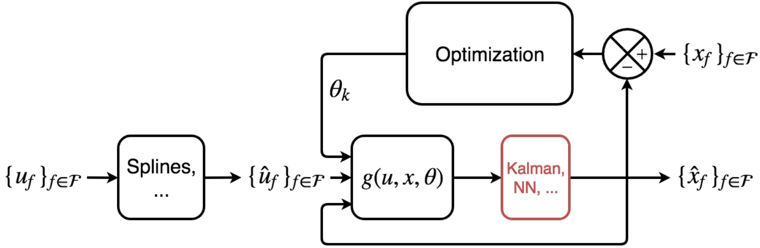
Ex: (Nonlinear) Least-Squares

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

State-of-the-art

- Output-Error Method
- Filter-Error Method

} Not scalable to many trajectories

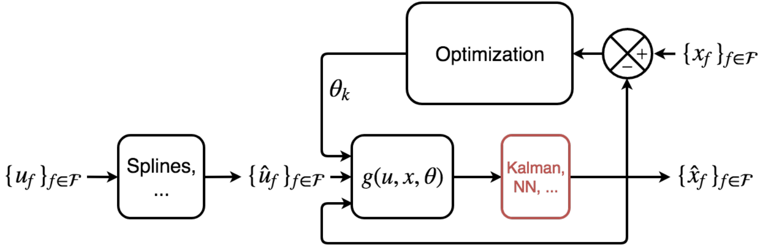


- Equation-Error Method *Ex: (Nonlinear) Least-Squares*

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

- Output-Error Method
- Filter-Error Method

} Not 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

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

- Nonlinear in states and controls

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

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

- Nonlinear in states and controls
- Nonlinear in parameters

Leveraging the dynamics structure

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

- Nonlinear in states and controls
- Nonlinear in parameters

Leveraging the dynamics structure

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

- Nonlinear in states and controls
- Nonlinear in parameters

Leveraging the dynamics structure

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

- 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

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

- Nonlinear in states and controls
- Nonlinear in parameters

Leveraging the dynamics structure

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

- Nonlinear in states and controls
- Nonlinear in parameters

Leveraging the dynamics structure

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

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

- Nonlinear in states and controls
- Nonlinear in parameters

Leveraging the dynamics structure

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

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

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

Leveraging the dynamics structure

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

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

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

LEVERAGING THE DYNAMICS STRUCTURE

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

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

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

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
- Structured

Leveraging the dynamics structure

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

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

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

LEVERAGING THE DYNAMICS STRUCTURE

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

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

- Nonlinear in states and controls
- ~~Nonlinear in parameters~~ → Linear in parameters
- Structured

Leveraging the dynamics structure

$$
\left\{
\begin{aligned}
\dot{h} &= V \sin\gamma \\
\underbrace{m\dot{V} + mg\sin\gamma}_{} &= \underbrace{X_{T1} \cdot \boldsymbol{\theta}_T - X_D \cdot \boldsymbol{\theta}_D}_{} + \varepsilon_1 \\
mV\dot{\gamma} + mg\cos\gamma &= X_{T2} \cdot \boldsymbol{\theta}_T + X_L \cdot \boldsymbol{\theta}_L + \varepsilon_2 \\
0 &= X_T \cdot \boldsymbol{\theta}_T + X_{Ispm} \cdot \boldsymbol{\theta}_{Isp} + \varepsilon_3
\end{aligned}
\right.
$$

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

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

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

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 $\rightsquigarrow$ $\boxed{\textbf{Multi-task Learning}}$

Multi-task Regression

Aircraft:

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

Multi-task Regression

Aircraft:

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

General:

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

Multi-task Regression

Aircraft:

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

General:

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

Coupling parameters , **Task specific parameters**

Multi-task Regression

Aircraft:

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

General:

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

Coupling parameters , **Task specific parameters**

Many other examples:

- *Giant squid neurons* [FitzHugh, 1961, Nagumo et al., 1962],

$$\begin{cases} \dot{V} = \boldsymbol{\theta}_1 \left(V - \dfrac{V^3}{3} + R \right), \\ \dot{R} = -\dfrac{1}{\boldsymbol{\theta}_1} (V - \boldsymbol{\theta}_2 + \boldsymbol{\theta}_3 R), \end{cases}$$

Multi-task Regression

Aircraft:

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

General:

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

Coupling parameters , **Task specific parameters**

Many other examples:

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

$$\begin{cases} \dot{S} = \boldsymbol{\theta}_1 - (\boldsymbol{\theta}_2 I + \boldsymbol{\theta}_3) S, \\ \dot{I} = \boldsymbol{\theta}_2 IS - (\boldsymbol{\theta}_4 + \boldsymbol{\theta}_3) I, \\ \dot{R} = \boldsymbol{\theta}_4 I - \boldsymbol{\theta}_3 R. \end{cases}$$

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:

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

Coupling parameters , Task specific parameters

Many other examples:

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

$$m\left(\frac{d\vec{V}}{dt}\right)_x = (\vec{F}(\boldsymbol{\theta}_F))_x, \quad m\left(\frac{d\vec{V}}{dt}\right)_y = (\vec{F}(\boldsymbol{\theta}_F))_y.$$

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:

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

Multi-task Regression

Aircraft:

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

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

Multi-task Linear Least-Squares: Block-sparse Coupling Structure

$$\min_{\theta} \sum_{i=1}^{N} \left\| \begin{pmatrix} Y_{1,i} \\ \vdots \\ Y_{K,i} \end{pmatrix} - \begin{pmatrix} X_{c,1,i} & X_{1,i} & 0 & 0 & \dots & 0 \\ X_{c,2,i} & 0 & X_{2,i} & 0 & \dots & 0 \\ \vdots & 0 & 0 & \ddots & 0 & 0 \\ X_{c,K,i} & 0 & 0 & \dots & 0 & X_{K,i} \end{pmatrix} \begin{pmatrix} \theta_c \\ \theta_1 \\ \vdots \\ \theta_K \end{pmatrix} \right\|_2^2$$

Multi-task Regression

Aircraft:

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

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**

Multi-task Linear Least-Squares: Block-sparse Coupling Structure

$$\min_{\theta} \sum_{i=1}^{N} \left\| \begin{pmatrix} Y_{1,i} \\ Y_{2,i} \\ Y_{3,i} \end{pmatrix} - \begin{pmatrix} X_{T1,i} & -X_{D,i} & 0 & 0 \\ X_{T2,i} & 0 & X_{L,i} & 0 \\ X_{T,i} & 0 & 0 & X_{Ispm,i} \end{pmatrix} \begin{pmatrix} \theta_T \\ \theta_D \\ \theta_L \\ \theta_{Isp} \end{pmatrix} \right\|_2^2$$

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:

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

Coupling parameters , **Task specific parameters**

Multi-task Linear Least-Squares: Block-sparse Coupling Structure

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

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

Feature Selection

Our model:

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

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]:

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

$\Rightarrow$ High risk of overfitting

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:

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

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

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

Sparse models are:

- Less susceptible to overfitting,

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]:

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

Sparse models are:

- Less susceptible to overfitting,
- More compliant with physical models,

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]:

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

Sparse models are:

- Less susceptible to overfitting,
- More compliant with physical models,
- More interpretable,

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,

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

Block-sparse Lasso

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

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

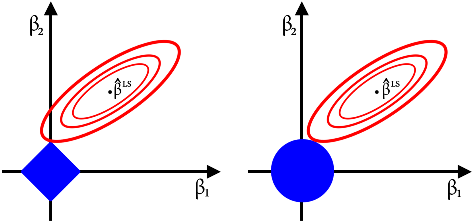


FIGURE: [1]Sparsity induced by L^1 norm in Lasso.

[1] Source : Wikipedia, Lasso(statistics)

Block-sparse Lasso

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

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

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

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**

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left\| \begin{pmatrix} Y_{1,i} \\ \vdots \\ Y_{K,i} \end{pmatrix} - \begin{pmatrix} B_{c,1,i} & B_{1,i} & 0 & 0 & \dots & 0 \\ B_{c,2,i} & 0 & B_{2,i} & 0 & \dots & 0 \\ \vdots & 0 & 0 & \ddots & 0 & 0 \\ B_{c,K,i} & 0 & 0 & \dots & 0 & B_{K,i} \end{pmatrix} \begin{pmatrix} \beta_c \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} \right\|_2^2$$

$$+ \lambda_c \left\| \begin{pmatrix} \beta_c \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} \right\|_1 ,$$

$$B_{c,k,i} := X_{c,k,i}, \qquad \beta_c := \boldsymbol{\theta}_c,$$

$$B_{k,i} := \frac{\lambda_c}{\lambda_k} X_{k,i}, \qquad \beta_k := \frac{\lambda_k}{\lambda_c} \boldsymbol{\theta}_k .$$

Block-sparse Lasso

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

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left\| \begin{pmatrix} Y_{1,i} \\ Y_{2,i} \\ Y_{3,i} \end{pmatrix} - \begin{pmatrix} X_{T1,i} & -\frac{\lambda_c}{\lambda_1} X_{D,i} & 0 & 0 \\ X_{T2,i} & 0 & \frac{\lambda_c}{\lambda_2} X_{L,i} & 0 \\ X_{T,i} & 0 & 0 & \frac{\lambda_c}{\lambda_3} X_{Ispm,i} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_T \\ \frac{\lambda_1}{\lambda_c} \boldsymbol{\theta}_D \\ \frac{\lambda_2}{\lambda_c} \boldsymbol{\theta}_L \\ \frac{\lambda_3}{\lambda_c} \boldsymbol{\theta}_{Isp} \end{pmatrix} \right\|_2^2 + \left\| \begin{pmatrix} \boldsymbol{\theta}_T \\ \frac{\lambda_1}{\lambda_c} \boldsymbol{\theta}_D \\ \frac{\lambda_2}{\lambda_c} \boldsymbol{\theta}_L \\ \frac{\lambda_3}{\lambda_c} \boldsymbol{\theta}_{Isp} \end{pmatrix} \right\|_1 .$$

Block-sparse Lasso

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

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left\| \begin{pmatrix} Y_{1,i} \\ Y_{2,i} \\ Y_{3,i} \end{pmatrix} - \begin{pmatrix} X_{T1,i} & -\frac{\lambda_c}{\lambda_1} X_{D,i} & 0 & 0 \\ X_{T2,i} & 0 & \frac{\lambda_c}{\lambda_2} X_{L,i} & 0 \\ X_{T,i} & 0 & 0 & \frac{\lambda_c}{\lambda_3} X_{Ispm,i} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_T \\ \frac{\lambda_1}{\lambda_c} \boldsymbol{\theta}_D \\ \frac{\lambda_2}{\lambda_c} \boldsymbol{\theta}_L \\ \frac{\lambda_3}{\lambda_c} \boldsymbol{\theta}_{Isp} \end{pmatrix} \right\|_2^2 + \left\| \begin{pmatrix} \boldsymbol{\theta}_T \\ \frac{\lambda_1}{\lambda_c} \boldsymbol{\theta}_D \\ \frac{\lambda_2}{\lambda_c} \boldsymbol{\theta}_L \\ \frac{\lambda_3}{\lambda_c} \boldsymbol{\theta}_{Isp} \end{pmatrix} \right\|_1 .$$

In practice: $\lambda_k = \lambda_c$, for all $k = 1, \ldots, K$.

Block-sparse Lasso

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

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left\| \begin{pmatrix} Y_{1,i} \\ Y_{2,i} \\ Y_{3,i} \end{pmatrix} - \begin{pmatrix} X_{T1,i} & - & X_{D,i} & 0 & 0 \\ X_{T2,i} & & 0 & X_{L,i} & 0 \\ X_{T,i} & & 0 & 0 & X_{lspm,i} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_T \\ \boldsymbol{\theta}_D \\ \boldsymbol{\theta}_L \\ \boldsymbol{\theta}_{lsp} \end{pmatrix} \right\|_2^2 + \left\| \begin{pmatrix} \boldsymbol{\theta}_T \\ \boldsymbol{\theta}_D \\ \boldsymbol{\theta}_L \\ \boldsymbol{\theta}_{lsp} \end{pmatrix} \right\|_1 .$$

In practice: $\lambda_k = \lambda_c$, for all $k = 1, \ldots, K$.

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

Structured feature selection

state-of-the-art

- Group Lasso [Yuan and Lin, 2005] — $\{(X_i, Y_i)\}_{i=1}^{N} \subset \mathbb{R}^{d+1}$ i.i.d sample, $\boldsymbol{\theta} = (\boldsymbol{\theta}_g)_{g \in \mathcal{G}}$ partition:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g \in \mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2,$$

STRUCTURED FEATURE SELECTION

STATE-OF-THE-ART

- Group Lasso [Yuan and Lin, 2005] — $\{(X_i, Y_i)\}_{i=1}^N \subset \mathbb{R}^{d+1}$ i.i.d sample, $\boldsymbol{\theta} = (\boldsymbol{\theta}_g)_{g \in \mathcal{G}}$ partition:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g \in \mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2,$$

$\Rightarrow$ Sparsity among groups

Structured feature selection

state-of-the-art

- Group Lasso [Yuan and Lin, 2005] — $\{(X_i, Y_i)\}_{i=1}^N \subset \mathbb{R}^{d+1}$ i.i.d sample, $\boldsymbol{\theta} = (\boldsymbol{\theta}_g)_{g \in \mathcal{G}}$ partition:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g \in \mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2,$$

  $\Rightarrow$ Sparsity among groups

- Sparse group Lasso [Friedman et al., 2010]:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g \in \mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2 + \lambda \|\boldsymbol{\theta}\|_1,$$

Structured feature selection

State-of-the-art

- Group Lasso [Yuan and Lin, 2005] — $\{(X_i, Y_i)\}_{i=1}^{N} \subset \mathbb{R}^{d+1}$ i.i.d sample, $\boldsymbol{\theta} = (\boldsymbol{\theta}_g)_{g\in\mathcal{G}}$ partition:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g\in\mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2,$$

  $\Rightarrow$ Sparsity among groups

- Sparse group Lasso [Friedman et al., 2010]:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g\in\mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2 + \lambda \|\boldsymbol{\theta}\|_1,$$

  $\Rightarrow$ Sparsity among **and** inside groups

STRUCTURED FEATURE SELECTION

STATE-OF-THE-ART

- Group Lasso [Yuan and Lin, 2005] — $\{(X_i, Y_i)\}_{i=1}^N \subset \mathbb{R}^{d+1}$ i.i.d sample, $\boldsymbol{\theta} = (\boldsymbol{\theta}_g)_{g\in\mathcal{G}}$ partition:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g\in\mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2,$$

  $\Rightarrow$ Sparsity among groups

- Sparse group Lasso [Friedman et al., 2010]:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g\in\mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2 + \lambda\|\boldsymbol{\theta}\|_1,$$

  $\Rightarrow$ Sparsity among **and** inside groups

- Multi-task Lasso [Obozinski et al., 2006] — K tasks, $X_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}^K$, $\boldsymbol{\theta} \in \mathbb{R}^{d\times K}$, $\boldsymbol{\theta}^k = (\boldsymbol{\theta}_j^k)_{j=1}^d$ parameter of task k:

$$\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \sum_{i=1}^{N} (Y_{k,i} - X_i \cdot \boldsymbol{\theta}^k)^2 + \sum_{j=1}^{d} \lambda_j \|\boldsymbol{\theta}_j\|_2.$$

Structured feature selection

state-of-the-art

- Group Lasso [Yuan and Lin, 2005] — $\{(X_i, Y_i)\}_{i=1}^N \subset \mathbb{R}^{d+1}$ i.i.d sample, $\boldsymbol{\theta} = (\boldsymbol{\theta}_g)_{g\in\mathcal{G}}$ partition:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g\in\mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2,$$

  $\Rightarrow$ Sparsity among groups

- Sparse group Lasso [Friedman et al., 2010]:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (Y_i - X_i \cdot \boldsymbol{\theta})^2 + \sum_{g\in\mathcal{G}} \lambda_g \|\boldsymbol{\theta}_g\|_2 + \lambda\|\boldsymbol{\theta}\|_1,$$

  $\Rightarrow$ Sparsity among **and** inside groups

- Multi-task Lasso [Obozinski et al., 2006] — K tasks, $X_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}^K$, $\boldsymbol{\theta} \in \mathbb{R}^{d\times K}$, $\boldsymbol{\theta}^k = (\boldsymbol{\theta}_j^k)_{j=1}^d$ parameter of task k:

$$\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \sum_{i=1}^{N} (Y_{k,i} - X_i \cdot \boldsymbol{\theta}^k)^2 + \sum_{j=1}^{d} \lambda_j \|\boldsymbol{\theta}_j\|_2.$$

  $\Rightarrow$ Shared sparsity among tasks

Structured feature selection
state-of-the-art

Other methods	Similarity/Difference with Block-sparse Lasso

STRUCTURED FEATURE SELECTION
STATE-OF-THE-ART

Other methods	Similarity/Difference with Block-sparse Lasso
Group Lasso	Parameters divided into groups, but group sparsity is precoded in quadratic loss function, so no $L^{2,1}$-norm,

Structured feature selection
state-of-the-art

Other methods	Similarity/Difference with Block-sparse Lasso
Group Lasso	Parameters divided into groups, but group sparsity is precoded in quadratic loss function, so no $L^{2,1}$-norm,
Sparse Group Lasso	Sparsity induced within group,

STRUCTURED FEATURE SELECTION
STATE-OF-THE-ART

Other methods	Similarity/Difference with Block-sparse Lasso
Group Lasso	Parameters divided into groups, but group sparsity is precoded in quadratic loss function, so no $L^{2,1}$-norm,
Sparse Group Lasso	Sparsity induced within group,
Multi-task Lasso	Multi-task setting, but not same pattern for every task.

Bootstrap implementation

High correlations among between polynomial features...

Bootstrap implementation

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

Bootstrap implementation

High correlations among between polynomial 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.

Bootstrap implementation

High correlations among between polynomial 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 under high correlations proved in Bach [2008],

Bootstrap implementation

High correlations among between polynomial 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 under high correlations proved in Bach [2008],
- Efficient implementations exists: LARS [Efron et al., 2004].

PROBLEM WITH INTRA-GROUP CORRELATIONS

Problem with intra-group correlations

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

Problem with intra-group correlations

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

Problem with intra-group correlations

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

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

Problem with intra-group correlations

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

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

Problem with intra-group correlations

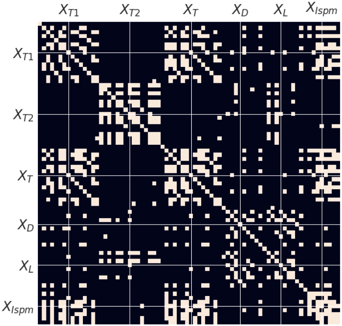


Figure: Features correlations higher than 0.9 in absolute value.

Problem with intra-group correlations

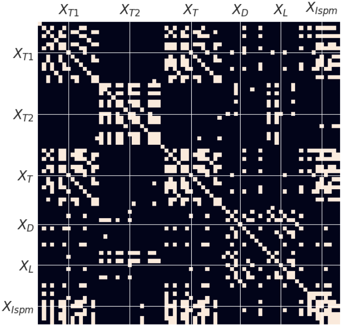


FIGURE: Features correlations higher than 0.9 in absolute value.

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

Problem with intra-group correlations

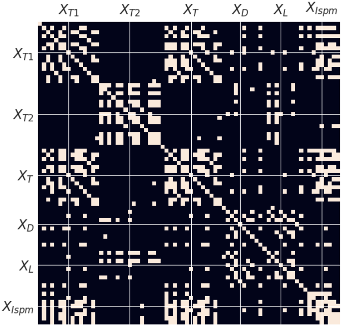


Figure: Features correlations higher than 0.9 in absolute value.

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

Ill-posed problem !

Problem with intra-group correlations

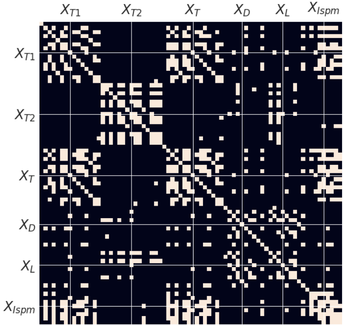


FIGURE: Features correlations higher than 0.9 in absolute value.

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

Ill-posed problem !

Tikhonov penalty [Tikhonov, 1943] known to improve problem conditioning $\|\boldsymbol{\theta}\|_2^2$,

Problem with intra-group correlations

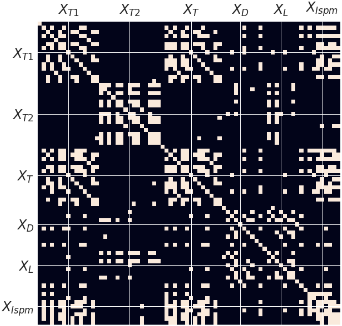


FIGURE: Features correlations higher than 0.9 in absolute value.

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

Ill-posed problem !

Tikhonov penalty [Tikhonov, 1943] known to improve problem conditioning $\|\boldsymbol{\theta}\|_2^2$, but shrinks parameters towards 0...

Problem with intra-group correlations

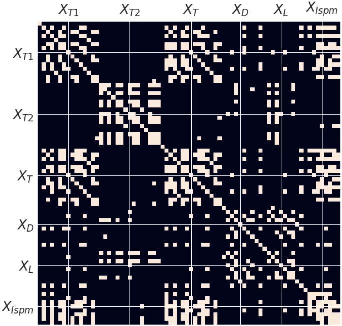


FIGURE: Features correlations higher than 0.9 in absolute value.

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

Ill-posed problem !

Tikhonov penalty [Tikhonov, 1943] known to improve problem conditioning $\|\boldsymbol{\theta}\|_2^2$, but shrinks parameters towards 0...

$\Rightarrow$ Generalized Tikhonov

$$\boxed{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^\top Q (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) = \|\Gamma(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\|_2^2,}$$

where $\tilde{\boldsymbol{\theta}}$ is a prior and $Q \in \mathbb{R}^{p \times p}$.

Problem with intra-group correlations

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

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

Problem with intra-group correlations

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

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

Use prior $\tilde{I}_{sp} \rightsquigarrow \{\tilde{I}_{sp,i} = \tilde{I}_{sp}(\boldsymbol{u}_i, \boldsymbol{x}_i)\}$

Problem with intra-group correlations

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

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

Use prior $\tilde{I}_{sp} \rightsquigarrow \{ \tilde{I}_{sp,i} = \tilde{I}_{sp}(\boldsymbol{u}_i, \boldsymbol{x}_i) \}$

Problem with intra-group correlations

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

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

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

Problem with intra-group correlations

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

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

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

Problem with intra-group correlations

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

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

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

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

Feature selection results

Feature selection results

- 25 different B737-800,

Feature selection results

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

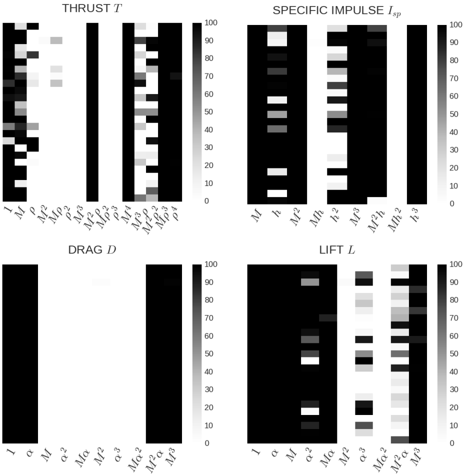
Feature selection results

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

FEATURE SELECTION RESULTS

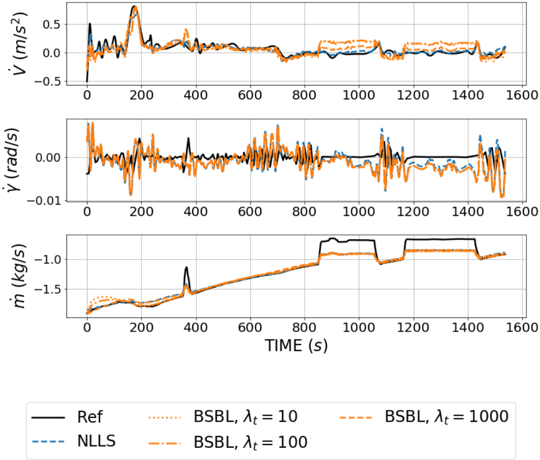
- 25 different B737-800,
- 10 471 flights = 8 261 619 observations,
- Block sparse Bolasso used for T, D, L and I_{sp},
- We expect similar model structures,

Feature selection results



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

ACCURACY OF DYNAMICS PREDICTIONS



Accuracy of dynamics predictions

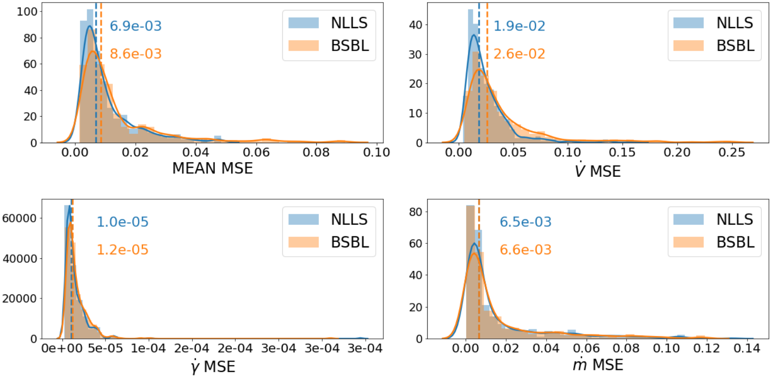
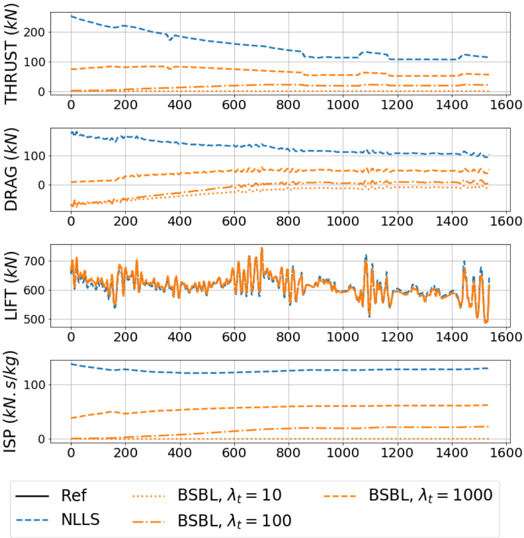


FIGURE: Leave-one-out off-sample errors distributions for nonlinear least-squares NLLS and block-sparse bolasso BSBL. Median errors are annotated and marked by dashed vertical lines.

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

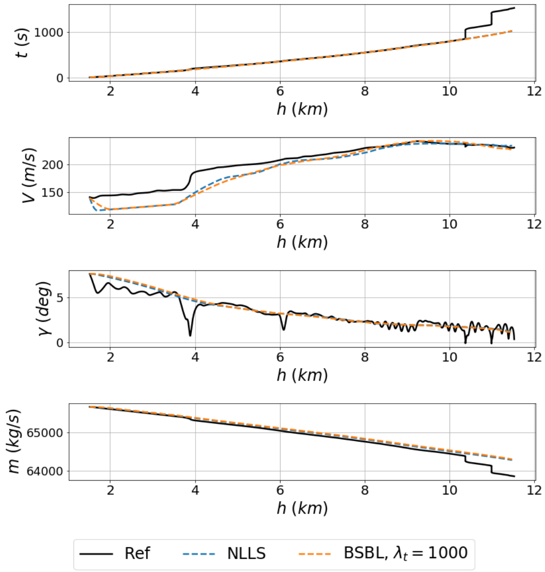
- Last assessment criterion = static;
- Does not incorporate the fact that the observations are time dependent;
- Does not take into account the goal of optimally controlling the aircraft system.

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

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

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

Flight resimulation



Flight resimulation

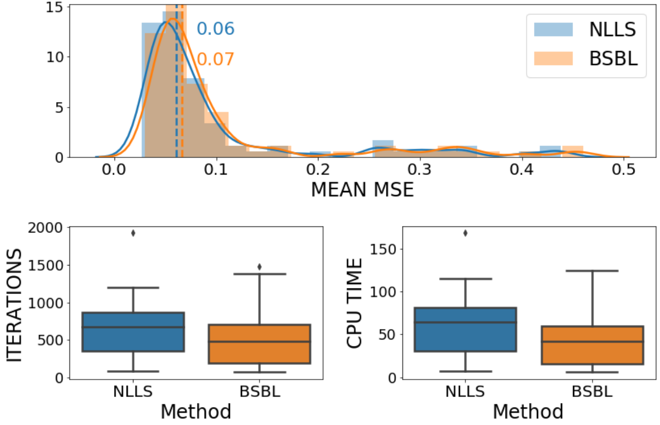


Figure: Distribution of the off-sample simulation error and boxplot of the optimization number of iterations and CPU time.

System Identification conclusion

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

System Identification conclusion

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

System Identification conclusion

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

System Identification conclusion

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

System Identification conclusion

- Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
- Applicable to large amounts of data, unlike traditional Output-Error Methods,
- Block-sparse estimators are proved to lead to consistent structured feature selection,
- Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
- No improvement in accuracy or training time compared to NLLS,

System Identification conclusion

- Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
- Applicable to large amounts of data, unlike traditional Output-Error Methods,
- Block-sparse estimators are proved to lead to consistent structured feature selection,
- Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
- No improvement in accuracy or training time compared to NLLS,
- Light, interpretable and compact data-dependent models (more than 50% compression),

System Identification conclusion

- Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
- Applicable to large amounts of data, unlike traditional Output-Error Methods,
- Block-sparse estimators are proved to lead to consistent structured feature selection,
- Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
- No improvement in accuracy or training time compared to NLLS,
- Light, interpretable and compact data-dependent models (more than 50% compression),
- Faster convergence of control problems,

System Identification conclusion

- Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
- Applicable to large amounts of data, unlike traditional Output-Error Methods,
- Block-sparse estimators are proved to lead to consistent structured feature selection,
- Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
- No improvement in accuracy or training time compared to NLLS,
- Light, interpretable and compact data-dependent models (more than 50% compression),
- Faster convergence of control problems,
- Consistency with physical knowledge,

System Identification conclusion

- Proposed Equation-Error Method approaches which extend to the System Identification framework well-known supervised learning techniques (Lasso, Ridge, bootstrap,...),
- Applicable to large amounts of data, unlike traditional Output-Error Methods,
- Block-sparse estimators are proved to lead to consistent structured feature selection,
- Can be efficiently trained using LARS algorithm as they are equivalent to successive Lasso problems,
- No improvement in accuracy or training time compared to NLLS,
- Light, interpretable and compact data-dependent models (more than 50% compression),
- Faster convergence of control problems,
- Consistency with physical knowledge,
- No initialization required.

Trajectory acceptability

Trajectory acceptability

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

Trajectory acceptability

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

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

Trajectory acceptability

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

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

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

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

Trajectory acceptability

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

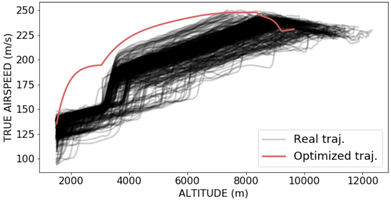
$$\text{s.t.} \begin{cases} \dot{\boldsymbol{x}}(t) = \hat{g}(\boldsymbol{u}(t),\boldsymbol{x}(t)), & \text{a.e. } t\in[0,t_f], \\ \text{Other constraints...} \end{cases} \tag{AOCP}$$

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

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

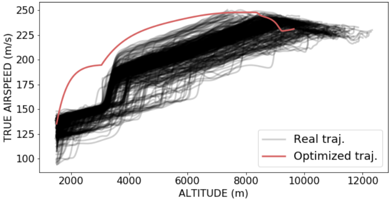
Does it look like a real trajectory ?

Trajectory acceptability



[2]NATS UK air traffic control

Trajectory acceptability

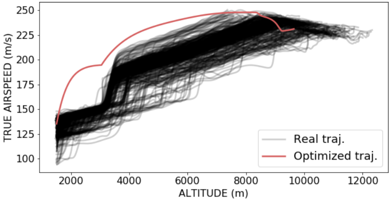


Pilots acceptance

---

[2] NATS UK air traffic control

Trajectory acceptability

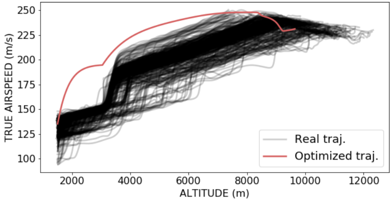


Pilots acceptance

Air Traffic Control[2]

[2]NATS UK air traffic control

Trajectory acceptability



Pilots acceptance Air Traffic Control[2]

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

[2]NATS UK air traffic control

Likelihood

- $X \sim f_{\theta^*}$,
- x: observation of X,
- Likelihood function of θ, given x:

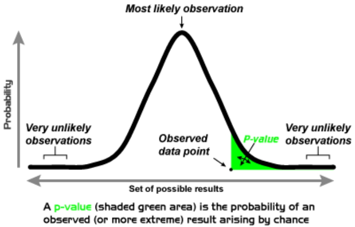
$$\mathcal{L}(\theta|x) = f_\theta(x).$$

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

Likelihood

- $X \sim f_{\theta^*}$,
- x: observation of X,
- Likelihood function of θ, given x:

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

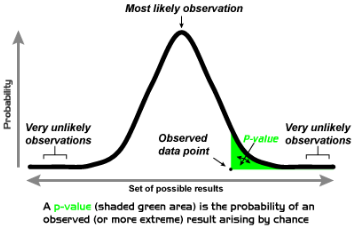


[0] Picture source: wikipedia, P-Value, author: Repapetilto CC.

Likelihood

- $X \sim f_{\theta^*}$,
- x: observation of X,
- Likelihood function of θ, given x:

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



In our case:

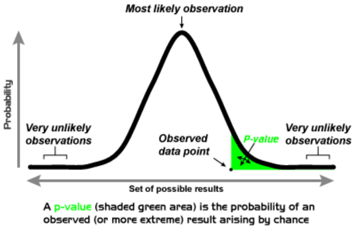
- the optimized trajectory plays the role of θ,

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

LIKELIHOOD

- $X \sim f_{\theta^*}$,
- x: observation of X,
- Likelihood function of θ, given x:

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



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 $\mathbb{T} = [0, t_f]$.

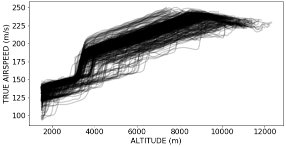
How to apply this to functional data?

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...

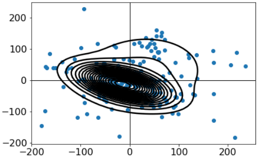
How to apply this to functional data?

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



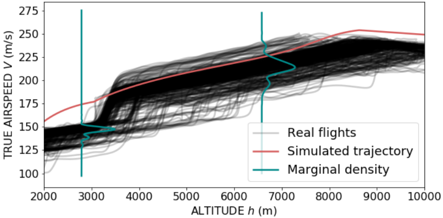
$\Longrightarrow$



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**



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

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

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

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

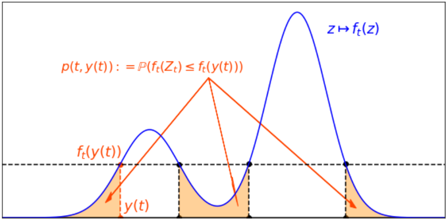
How do we aggregate the marginal likelihoods?

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 $\hat{f}^m_{\tilde{t}_j}$ of the marginal densities $f_{\tilde{t}_j}$:

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

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],

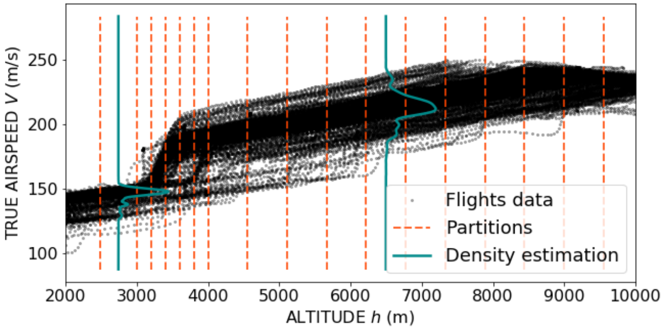
How can we estimate marginal densities?

- Suppose that sampling times $\{t_j^r : j = 1, \dots, n; r = 1, \dots, m\}$ 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 $(X, Y) = (T, Z_T)$.

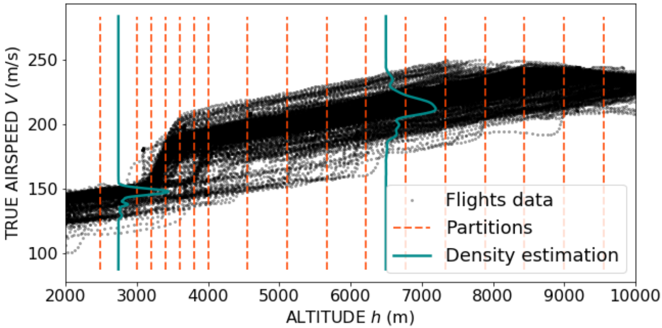
$\Rightarrow$ We could apply SOA conditional density estimation techniques, such as LS-CDE [Sugiyama et al., 2010],

$\Rightarrow$ **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

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

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.

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

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:

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

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:

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

Note that:

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

Consistency

Theorem 1

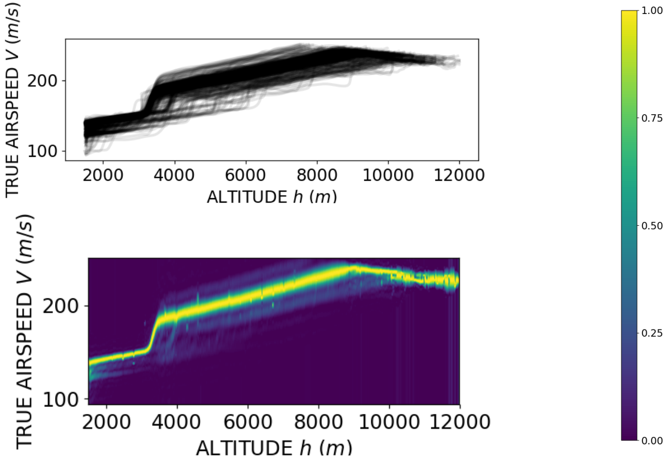
Under assumptions 1 to 4, for any $z \in E$ and $t \in \mathbb{T}$, $\hat{f}^m_{\ell^m(t)}(z)$ consistently approximates the marginal density $f_t(z)$ as the number of curves m grows:

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

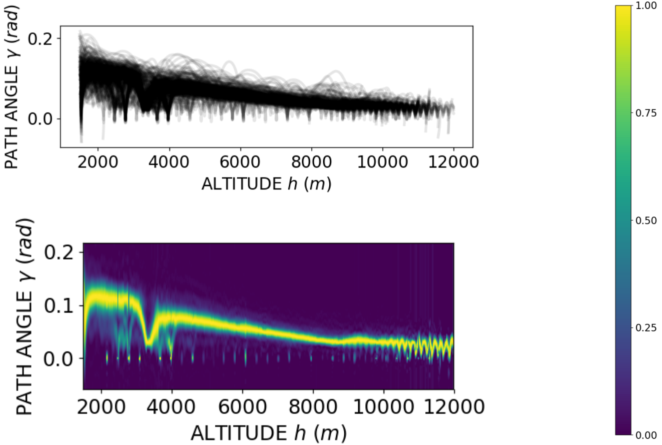
Note that:

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

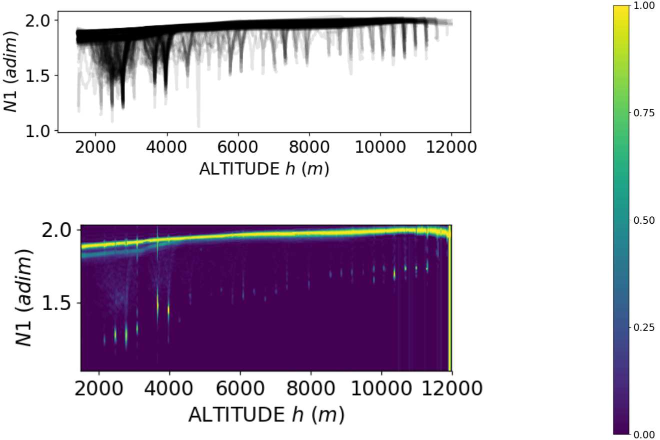
Marginal density estimation results



Marginal density estimation results



Marginal density estimation results



How good is it compared to other methods?

How good is it compared to other methods?

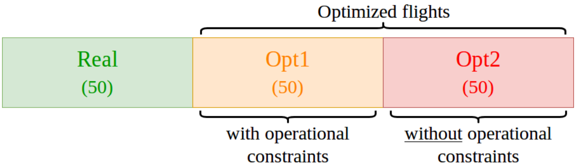
- Training set of $m = 424$ flights $\simeq$ 334 531 point observations,

How good is it compared to other methods?

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

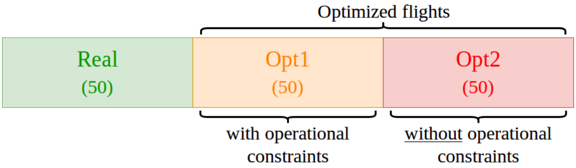
How good is it compared to other methods?

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



HOW GOOD IS IT COMPARED TO OTHER METHODS?

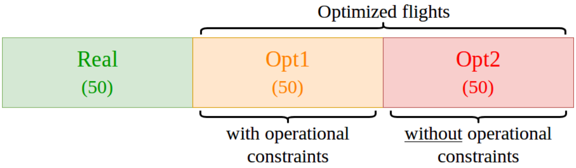
- Training set of $m = 424$ flights $\simeq 334\,531$ point observations,
- Test set of 150 flights



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

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

MML PENALTY

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

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

MML penalty

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

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

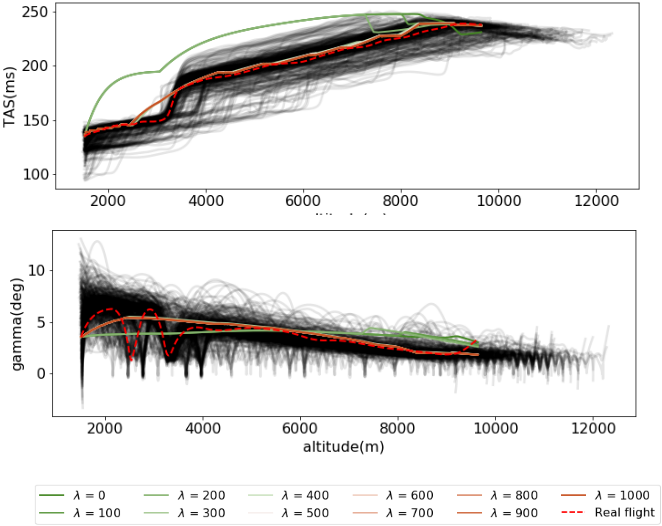
MML PENALTY

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

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

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

Penalty effect



Consumption x Acceptability trade-off

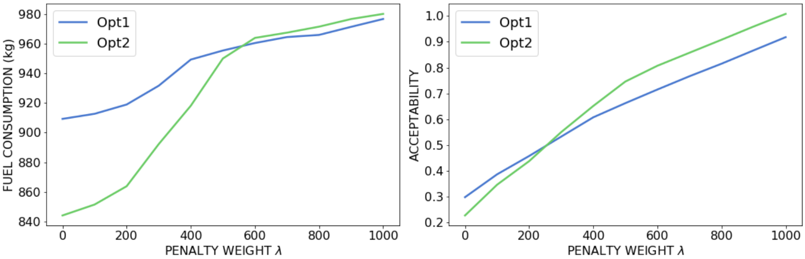


Figure: Average over 20 flights of the fuel consumption and MML score (called acceptability here) of optimized trajectories with varying MML-penalty weight λ.

Trajectory acceptability conclusion

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

Trajectory acceptability 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,

TRAJECTORY ACCEPTABILITY 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,

Trajectory acceptability 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,

Trajectory acceptability 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 Adaptive Kernel and Gaussian mixture implementation,

Trajectory acceptability 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 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 Proceedings of the 25th international conference on Machine learning, pages 33–40.

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.

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