



**HAL**  
open science

## Multi-task Bolasso based aircraft dynamics identification

Cédric Rommel, Joseph Frédéric Bonnans, Baptiste Gregorutti, Pierre  
Martinon

► **To cite this version:**

Cédric Rommel, Joseph Frédéric Bonnans, Baptiste Gregorutti, Pierre Martinon. Multi-task Bolasso based aircraft dynamics identification. PGMODays, Nov 2017, Paris, France. hal-01643177

**HAL Id: hal-01643177**

**<https://inria.hal.science/hal-01643177v1>**

Submitted on 21 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# MULTI-TASK BOLASSO BASED AIRCRAFT DYNAMICS IDENTIFICATION

C. Rommel<sup>1,2</sup>, J. F. Bonnans<sup>1</sup>,  
B. Gregorutti<sup>2</sup> and P. Martinon<sup>1</sup>

CMAP Ecole Polytechnique - INRIA<sup>1</sup>  
Safety Line<sup>2</sup>

November 14<sup>th</sup> 2017



# MOTIVATION

- 20 000 airplanes — 80 000 flights per day,

# MOTIVATION

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

# MOTIVATION

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

# MOTIVATION

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

# MOTIVATION

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

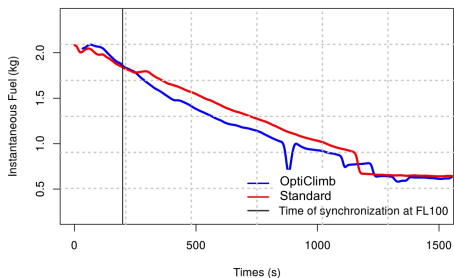
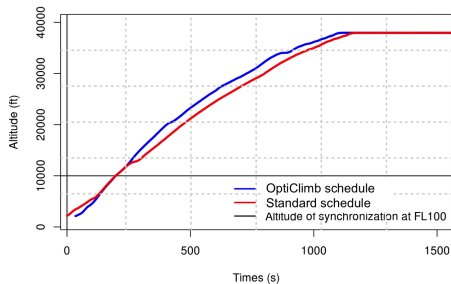
# MOTIVATION

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





# EXAMPLE OF OPTIMIZED TRAJECTORY



<b>Reference</b>	<b>Optimized</b>
1311kg	1141kg

# OPTIMAL CONTROL PROBLEM

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

s.t.  $\left\{ \right.$

# OPTIMAL CONTROL PROBLEM

$$\begin{aligned} & \min_{(\mathbf{x}, \mathbf{u}) \in \mathbb{X} \times \mathbb{U}} \int_0^{t_f} C(t, \mathbf{u}(t), \mathbf{x}(t)) dt, \\ \text{s.t. } & \left\{ \begin{array}{l} \Phi(\mathbf{x}(0), \mathbf{x}(t_f)) \in K_\Phi, \end{array} \right. \end{aligned}$$

# OPTIMAL CONTROL PROBLEM

$$\begin{aligned} & \min_{(\mathbf{x}, \mathbf{u}) \in \mathbb{X} \times \mathbb{U}} \int_0^{t_f} C(t, \mathbf{u}(t), \mathbf{x}(t)) dt, \\ \text{s.t. } & \left\{ \begin{array}{l} \Phi(\mathbf{x}(0), \mathbf{x}(t_f)) \in K_\Phi, \\ \mathbf{u}(t) \in U_{ad}, \quad \text{for a.e. } t \in [0, t_f], \end{array} \right. \end{aligned}$$

# OPTIMAL CONTROL PROBLEM

$$\begin{aligned} & \min_{(\mathbf{x}, \mathbf{u}) \in \mathbb{X} \times \mathbb{U}} \int_0^{t_f} C(t, \mathbf{u}(t), \mathbf{x}(t)) dt, \\ \text{s.t. } & \begin{cases} \Phi(\mathbf{x}(0), \mathbf{x}(t_f)) \in K_\Phi, \\ \mathbf{u}(t) \in U_{ad}, \quad \text{for a.e. } t \in [0, t_f], \\ c_j(\mathbf{x}(t)) \leq 0, \quad j = 1, \dots, n_c, \quad \text{for all } t \in [0, t_f], \end{cases} \end{aligned}$$

# OPTIMAL CONTROL PROBLEM

$$\begin{aligned} & \min_{(\mathbf{x}, \mathbf{u}) \in \mathbb{X} \times \mathbb{U}} \int_0^{t_f} C(t, \mathbf{u}(t), \mathbf{x}(t)) dt, \\ \text{s.t. } & \begin{cases} \Phi(\mathbf{x}(0), \mathbf{x}(t_f)) & \in K_\Phi, \\ \mathbf{u}(t) & \in U_{ad}, \quad \text{for a.e. } t \in [0, t_f], \\ c_j(\mathbf{x}(t)) & \leq 0, \quad j = 1, \dots, n_c, \quad \text{for all } t \in [0, t_f], \\ \dot{\mathbf{x}} & = g(t, \mathbf{u}, \mathbf{x}), \quad \text{for a.e. } t \in [0, t_f]. \end{cases} \end{aligned}$$

# OPTIMAL CONTROL PROBLEM

$$\begin{aligned} & \min_{(\mathbf{x}, \mathbf{u}) \in \mathbb{X} \times \mathbb{U}} \int_0^{t_f} C(t, \mathbf{u}(t), \mathbf{x}(t)) dt, \\ \text{s.t. } & \begin{cases} \Phi(\mathbf{x}(0), \mathbf{x}(t_f)) \in K_\Phi, \\ \mathbf{u}(t) \in U_{ad}, \quad \text{for a.e. } t \in [0, t_f], \\ c_j(\mathbf{x}(t)) \leq 0, \quad j = 1, \dots, n_c, \quad \text{for all } t \in [0, t_f], \\ \dot{\mathbf{x}} = g(t, \mathbf{u}, \mathbf{x}), \quad \text{for a.e. } t \in [0, t_f]. \end{cases} \end{aligned}$$

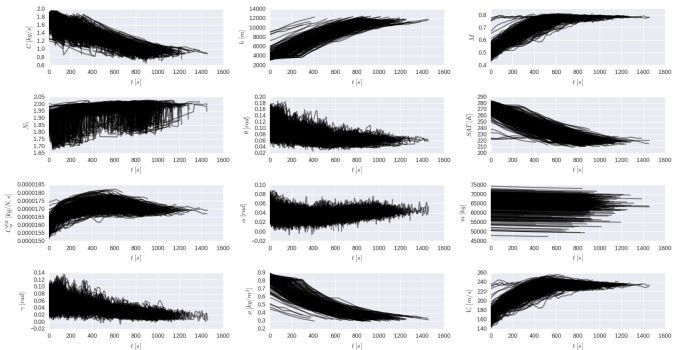
# QAR DATA



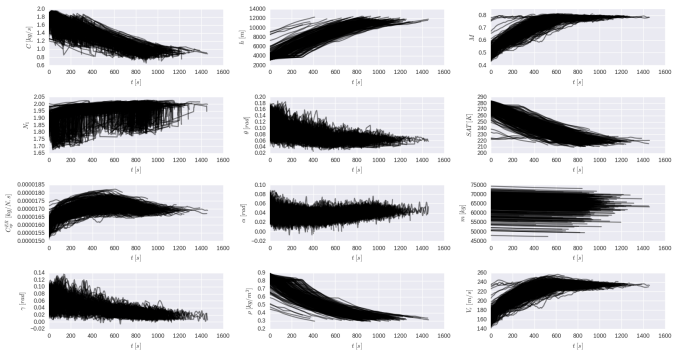
# QAR DATA



# QAR DATA

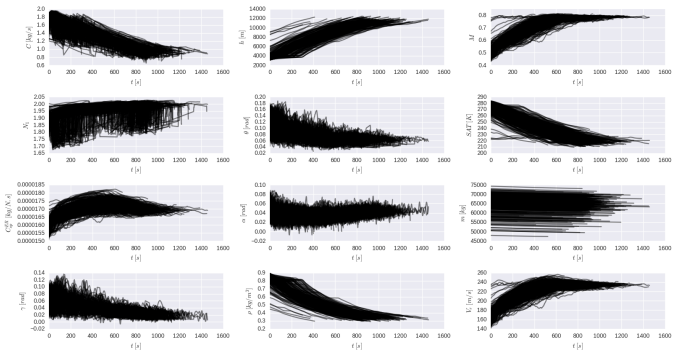


# QAR DATA



Massive ( $> 1000$  variables recorded every second),

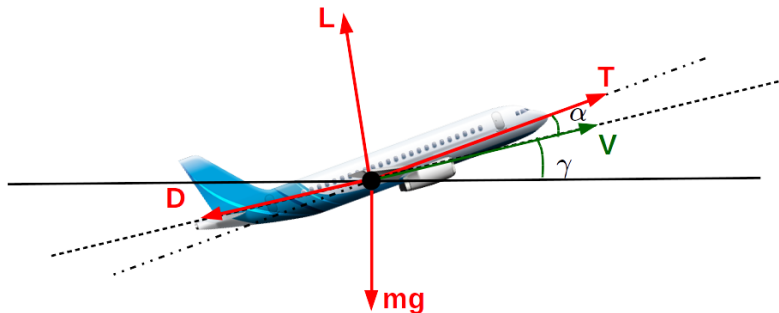
# QAR DATA



Massive ( $> 1000$  variables recorded every second),

$$\dot{\mathbf{x}} = \mathbf{g}(t, \mathbf{u}, \mathbf{x})$$

# FLIGHT MECHANICS AND STATE EQUATION



Classic flight mechanics model

# FLIGHT MECHANICS AND STATE EQUATION

$$\left\{ \begin{array}{l} \dot{h} = V \sin \gamma, \\ \dot{V} = \frac{T \cos \alpha - D - mg \sin \gamma}{m}, \\ \dot{\gamma} = \frac{T \sin \alpha + L - mg \cos \gamma}{mV}, \\ \dot{m} = -\frac{T}{I_{sp}}. \end{array} \right.$$

# FLIGHT MECHANICS AND STATE EQUATION

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

State variables:  $\mathbf{x} = [h, V, \gamma, m]$

# FLIGHT MECHANICS AND STATE EQUATION

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

State variables:  $\mathbf{x} = [h, V, \gamma, m]$

Control variables:  $\mathbf{u} = [\alpha, N_1]$



# FLIGHT MECHANICS AND STATE EQUATION

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

State variables:  $\mathbf{x} = [h, V, \gamma, m]$

Control variables:  $\mathbf{u} = [\alpha, N_1]$

**Unknown functions of the state and control variables**

# FLIGHT MECHANICS AND STATE EQUATION

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

State variables:  $\mathbf{x} = [h, V, \gamma, m]$

Control variables:  $\mathbf{u} = [\alpha, N_1]$

**Unknown functions of the state and control variables**

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T \text{ function of } (M, \rho, N_1), \\ D \text{ function of } (M, \rho, q), \\ L \text{ function of } (M, \rho, q), \\ I_{sp} \text{ function of } (M, h, SAT), \end{array} \right.$$

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T \text{ function of } (M, \rho, N_1) = \varphi_T(\mathbf{x}, \mathbf{u}), \\ D \text{ function of } (M, \rho, q) = \varphi_D(\mathbf{x}, \mathbf{u}), \\ L \text{ function of } (M, \rho, q) = \varphi_L(\mathbf{x}, \mathbf{u}), \\ I_{sp} \text{ function of } (M, h, SAT) = \varphi_{I_{sp}}(\mathbf{x}, \mathbf{u}), \end{array} \right.$$

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T \text{ function of } (M, \rho, N_1) = \varphi_T(\mathbf{x}, \mathbf{u}), \\ D \text{ function of } (M, \rho, q) = \varphi_D(\mathbf{x}, \mathbf{u}), \\ L \text{ function of } (M, \rho, q) = \varphi_L(\mathbf{x}, \mathbf{u}), \\ I_{sp} \text{ function of } (M, h, SAT) = \varphi_{I_{sp}}(\mathbf{x}, \mathbf{u}), \end{array} \right.$$

- Need for smooth models,

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T \text{ function of } (M, \rho, N_1) = \varphi_T(\mathbf{x}, \mathbf{u}), \\ D \text{ function of } (M, \rho, q) = \varphi_D(\mathbf{x}, \mathbf{u}), \\ L \text{ function of } (M, \rho, q) = \varphi_L(\mathbf{x}, \mathbf{u}), \\ I_{sp} \text{ function of } (M, h, SAT) = \varphi_{I_{sp}}(\mathbf{x}, \mathbf{u}), \end{array} \right.$$

- Need for smooth models,
- Need for models which are fast to compute,

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T \text{ function of } (M, \rho, N_1) = \varphi_T(\mathbf{x}, \mathbf{u}), \\ D \text{ function of } (M, \rho, q) = \varphi_D(\mathbf{x}, \mathbf{u}), \\ L \text{ function of } (M, \rho, q) = \varphi_L(\mathbf{x}, \mathbf{u}), \\ I_{sp} \text{ function of } (M, h, SAT) = \varphi_{I_{sp}}(\mathbf{x}, \mathbf{u}), \end{array} \right.$$

- Need for smooth models,
- Need for models which are fast to compute,
- Need for interpretable models for safety,

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T = X_T \cdot \theta_T, \\ D = X_D \cdot \theta_D, \\ L = X_L \cdot \theta_L, \\ I_{sp} = X_{I_{sp}} \cdot \theta_{I_{sp}}. \end{array} \right.$$

- Need for smooth models,
- Need for models which are fast to compute,
- Need for interpretable models for safety,



# MODEL REQUIREMENTS

$$\left\{ \begin{array}{ll} T = X_T \cdot \theta_T, & \text{with } X_T \neq \varphi_T(\mathbf{x}, \mathbf{u}), \\ D = X_D \cdot \theta_D, & \text{with } X_D \neq \varphi_D(\mathbf{x}, \mathbf{u}), \\ L = X_L \cdot \theta_L, & \text{with } X_L \neq \varphi_L(\mathbf{x}, \mathbf{u}), \\ I_{sp} = X_{Isp} \cdot \theta_{Isp}, & \text{with } X_{Isp} \neq \varphi_{Isp}(\mathbf{x}, \mathbf{u}). \end{array} \right.$$

- Need for smooth models,
- Need for models which are fast to compute,
- Need for interpretable models for safety,

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T = X_T \cdot \theta_T, \\ D = X_D \cdot \theta_D, \\ L = X_L \cdot \theta_L, \\ I_{sp} = X_{Isp} \cdot \theta_{Isp}, \end{array} \right. \quad \begin{array}{l} \text{with } X_T \neq \varphi_T(\mathbf{x}, \mathbf{u}), \\ \text{with } X_D \neq \varphi_D(\mathbf{x}, \mathbf{u}), \\ \text{with } X_L \neq \varphi_L(\mathbf{x}, \mathbf{u}), \\ \text{with } X_{Isp} \neq \varphi_{Isp}(\mathbf{x}, \mathbf{u}). \end{array}$$

- Need for smooth models,
- Need for models which are fast to compute,
- Need for interpretable models for safety,
- Need for models which are rich enough.

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T = X_T \cdot \theta_T, \\ D = X_D \cdot \theta_D, \\ L = X_L \cdot \theta_L, \\ I_{sp} = X_{Isp} \cdot \theta_{Isp}, \end{array} \right. \quad \begin{array}{l} \text{with } X_T = \Phi_d \circ \varphi_T(\mathbf{x}, \mathbf{u}), \\ \text{with } X_D = \Phi_d \circ \varphi_D(\mathbf{x}, \mathbf{u}), \\ \text{with } X_L = \Phi_d \circ \varphi_L(\mathbf{x}, \mathbf{u}), \\ \text{with } X_{Isp} = \Phi_d \circ \varphi_{Isp}(\mathbf{x}, \mathbf{u}). \end{array}$$

- Need for smooth models,
- Need for models which are fast to compute,
- Need for interpretable models for safety,
- Need for models which are rich enough.

# MODEL REQUIREMENTS

$$\left\{ \begin{array}{l} T = X_T \cdot \theta_T, \\ D = X_D \cdot \theta_D, \\ L = X_L \cdot \theta_L, \\ I_{sp} = X_{I_{sp}} \cdot \theta_{I_{sp}}, \end{array} \right. \quad \begin{array}{l} \text{with } X_T = N_1(1, \rho, M, \rho^2, \rho M, M^2, \dots), \\ \text{with } X_D = q(1, \alpha, M, \alpha^2, \alpha M, M^2, \dots), \\ \text{with } X_L = q(1, \alpha, M, \alpha^2, \alpha M, M^2, \dots), \\ \text{with } X_{I_{sp}} = (1, h, M, h^2, hM, M^2, \dots). \end{array}$$

- Need for smooth models,
- Need for models which are fast to compute,
- Need for interpretable models for safety,
- Need for models which are rich enough.

# REGRESSION PROBLEMS

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

# REGRESSION PROBLEMS

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

# REGRESSION PROBLEMS

$$\left\{ \begin{array}{l} \dot{h} = V \sin \gamma \\ m\dot{V}_r + mg \sin \gamma = T \cos \alpha - D \\ mV_r\dot{\gamma} + mg \cos \gamma = T \sin \alpha + L \\ \dot{m} = -\frac{T}{I_{sp}}. \end{array} \right.$$

# REGRESSION PROBLEMS

$$\left\{ \begin{array}{l} \dot{h} = V \sin \gamma \\ m\dot{V}_r + mg \sin \gamma = T \cos \alpha - D \\ mV_r\dot{\gamma} + mg \cos \gamma = T \sin \alpha + L \\ 0 = T + \dot{m}l_{sp}. \end{array} \right.$$



# REGRESSION PROBLEMS

$$\left\{ \begin{array}{l} \dot{h} = V \sin \gamma \\ m\dot{V}_r + mg \sin \gamma = T \cos \alpha - D \\ mV_r \dot{\gamma} + mg \cos \gamma = T \sin \alpha + L \\ 0 = T + \dot{m}l_{sp} \end{array} \right.$$

⇓

$$\left\{ \begin{array}{l} Y_1 = X_T \cos \alpha \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_T \sin \alpha \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + \dot{m}X_{lsp} \cdot \theta_{lsp} + \varepsilon_3 \end{array} \right.$$

**Targets to fit**

**Unknown**

**Random error**

# MULTI-TASK REGRESSION FRAMEWORK

$$\begin{cases} Y_1 = X_T \cos \alpha \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_T \sin \alpha \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + \dot{m} X_{Isp} \cdot \theta_{Isp} + \varepsilon_3 \end{cases}$$

# MULTI-TASK REGRESSION FRAMEWORK

$$\begin{cases} Y_1 = X_T \cos \alpha \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_T \sin \alpha \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + \dot{m} X_{Isp} \cdot \theta_{Isp} + \varepsilon_3 \end{cases}$$

# MULTI-TASK REGRESSION FRAMEWORK

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Isp} + \varepsilon_3 \end{cases}$$

# MULTI-TASK REGRESSION FRAMEWORK

$$Y = X\theta + \varepsilon,$$

# MULTI-TASK REGRESSION FRAMEWORK

$$Y = X\theta + \varepsilon,$$

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ 0 \end{pmatrix}}_Y = \underbrace{\begin{pmatrix} X_{T1}^\top & -X_D^\top & 0 & 0 \\ X_{T2}^\top & 0 & X_L^\top & 0 \\ X_T^\top & 0 & 0 & X_{lspm}^\top \end{pmatrix}}_X \underbrace{\begin{pmatrix} \theta_T \\ \theta_D \\ \theta_L \\ \theta_{lspm} \end{pmatrix}}_\theta + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}}_\varepsilon.$$

# MULTI-TASK REGRESSION FRAMEWORK

$$Y = X\theta + \varepsilon,$$

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ 0 \end{pmatrix}}_Y = \underbrace{\begin{pmatrix} X_{T1}^\top & -X_D^\top & 0 & 0 \\ X_{T2}^\top & 0 & X_L^\top & 0 \\ X_T^\top & 0 & 0 & X_{lspm}^\top \end{pmatrix}}_X \underbrace{\begin{pmatrix} \theta_T \\ \theta_D \\ \theta_L \\ \theta_{lspm} \end{pmatrix}}_\theta + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}}_\varepsilon.$$

- Ensures all components of  $\hat{g}$  to share the same thrust  $\hat{T}$ ,

# MULTI-TASK REGRESSION FRAMEWORK

$$Y = X\theta + \varepsilon,$$

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ 0 \end{pmatrix}}_Y = \underbrace{\begin{pmatrix} X_{T1}^\top & -X_D^\top & 0 & 0 \\ X_{T2}^\top & 0 & X_L^\top & 0 \\ X_T^\top & 0 & 0 & X_{lspm}^\top \end{pmatrix}}_X \underbrace{\begin{pmatrix} \theta_T \\ \theta_D \\ \theta_L \\ \theta_{lspm} \end{pmatrix}}_\theta + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}}_\varepsilon.$$

- Ensures all components of  $\hat{g}$  to share the same thrust  $\hat{T}$ ,
- Better predictive accuracy from tight coupling,



# MULTI-TASK REGRESSION FRAMEWORK

$$Y = X\theta + \varepsilon,$$

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ 0 \end{pmatrix}}_Y = \underbrace{\begin{pmatrix} X_{T1}^\top & -X_D^\top & 0 & 0 \\ X_{T2}^\top & 0 & X_L^\top & 0 \\ X_T^\top & 0 & 0 & X_{lspm}^\top \end{pmatrix}}_X \underbrace{\begin{pmatrix} \theta_T \\ \theta_D \\ \theta_L \\ \theta_{lspm} \end{pmatrix}}_\theta + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}}_\varepsilon.$$

- Ensures all components of  $\hat{g}$  to share the same thrust  $\hat{T}$ ,
- Better predictive accuracy from tight coupling,
- Helps in high correlations setting.

# FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,

# FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

# FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

## EMPIRICAL RISK MINIMIZATION

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(Y^i, X^i \theta),$$

# FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

## LEAST SQUARES REGRESSION

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2,$$

## FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

## LEAST SQUARES REGRESSION

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2,$$

## FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

$\rightsquigarrow$  **Overfitting...**

## LEAST SQUARES REGRESSION

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2,$$

# FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

$\rightsquigarrow$  **Overfitting...**

## $L^1$ PENALIZATION

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda \|\theta\|_1,$$



# FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

$\rightsquigarrow$  **Overfitting...**

## $L^1$ PENALIZATION

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda \|\theta\|_1,$$

$\simeq$  Lasso [Tibshirani, 1994]

## FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

$\rightsquigarrow$  **Overfitting...**

## BLOCK SPARSE LASSO

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda \|\theta\|_1,$$

$\simeq$  Lasso [Tibshirani, 1994]

## FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

$\rightsquigarrow$  **Overfitting...**

## BLOCK SPARSE LASSO

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda \|\theta\|_1,$$

$\simeq$  Lasso [Tibshirani, 1994]

Polynomial regression

## FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

$\rightsquigarrow$  **Overfitting...**

## BLOCK SPARSE LASSO

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda \|\theta\|_1,$$

$\simeq$  Lasso [Tibshirani, 1994]

Polynomial regression  $\rightsquigarrow$  high correlations between elements of  $X^i$

## FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

$\rightsquigarrow$  **Overfitting...**

## BLOCK SPARSE LASSO

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda \|\theta\|_1,$$

$\simeq$  Lasso [Tibshirani, 1994]

Polynomial regression  $\rightsquigarrow$  high correlations between elements of  $X^i$

$\rightsquigarrow$  **Unstable selections...**

## FEATURE SELECTION

Let  $\{(\mathbf{x}^i, \mathbf{u}^i, \dot{\mathbf{x}}^i)\}_{i=1}^N$  set of  $N$  observations,  $\rightsquigarrow \{X^i, Y^i\}_{i=1}^N$ ,

Maybe not all monomials are relevant for  $T, D, L$  and/or  $I_{sp}$  model...

$\rightsquigarrow$  **Overfitting...**

## BLOCK SPARSE LASSO

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda \|\theta\|_1,$$

$\simeq$  Lasso [Tibshirani, 1994]

Polynomial regression  $\rightsquigarrow$  high correlations between elements of  $X^i$

$\rightsquigarrow$  **Unstable selections...**

$\Rightarrow$  **Bolasso** [Bach, 2008]

# BOOTSTRAP IMPLEMENTATION

---

## Block sparse Bolasso

---

training data  $\mathcal{T} = \{(X^i, Y^i)\}_{i=1}^N$ ,

**Require:** number of bootstrap replicates  $m$ ,  
 $L^1$  penalization parameter  $\lambda$ ,

- 1: **for**  $k = 1$  **to**  $m$  **do**
  - 2:   Generate bootstrap sample  $\mathcal{T}^k$ ,
  - 3:   Compute Block sparse Lasso estimate  $\hat{\theta}^k$  from  $\mathcal{T}^k$ ,
  - 4:   Compute support  $J_k = \{j, \hat{\theta}_j^k \neq 0\}$ ,
  - 5: **end for**
  - 6: Compute  $J = \bigcap_{k=1}^m J_k$ ,
  - 7: Compute  $\hat{\theta}_J$  from  $\mathcal{T}_J = \{(X_J^i, Y^i)\}_{i=1}^N$  using Least-Squares.
-

# BOOTSTRAP IMPLEMENTATION

---

## Block sparse Bolasso

---

training data  $\mathcal{T} = \{(X^i, Y^i)\}_{i=1}^N$ ,

**Require:** number of bootstrap replicates  $m$ ,  
 $L^1$  penalization parameter  $\lambda$ ,

- 1: **for**  $k = 1$  **to**  $m$  **do**
  - 2:   Generate bootstrap sample  $\mathcal{T}^k$ ,
  - 3:   Compute Block sparse Lasso estimate  $\hat{\theta}^k$  from  $\mathcal{T}^k$ ,
  - 4:   Compute support  $J_k = \{j, \hat{\theta}_j^k \neq 0\}$ ,
  - 5: **end for**
  - 6: Compute  $J = \bigcap_{k=1}^m J_k$ ,
  - 7: Compute  $\hat{\theta}_J$  from  $\mathcal{T}_J = \{(X_J^i, Y^i)\}_{i=1}^N$  using Least-Squares.
- 

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



# BOOTSTRAP IMPLEMENTATION

---

## Block sparse Bolasso

---

training data  $\mathcal{T} = \{(X^i, Y^i)\}_{i=1}^N$ ,

**Require:** number of bootstrap replicates  $m$ ,  
 $L^1$  penalization parameter  $\lambda$ ,

- 1: **for**  $k = 1$  **to**  $m$  **do**
  - 2:   Generate bootstrap sample  $\mathcal{T}^k$ ,
  - 3:   Compute Block sparse Lasso estimate  $\hat{\theta}^k$  from  $\mathcal{T}^k$ ,
  - 4:   Compute support  $J_k = \{j, \hat{\theta}_j^k \neq 0\}$ ,
  - 5: **end for**
  - 6: Compute  $J = \bigcap_{k=1}^m J_k$ ,
  - 7: Compute  $\hat{\theta}_J$  from  $\mathcal{T}_J = \{(X_J^i, Y^i)\}_{i=1}^N$  using Least-Squares.
- 

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

# IDENTIFIABILITY ISSUES

# IDENTIFIABILITY ISSUES

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Isp} + \varepsilon_3 \end{cases}$$

# IDENTIFIABILITY ISSUES

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Isp} + \varepsilon_3 \end{cases}$$

# IDENTIFIABILITY ISSUES

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Isp} + \varepsilon_3 \end{cases}$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda_1 \|\theta\|_1,$$

# IDENTIFIABILITY ISSUES

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Isp} + \varepsilon_3 \end{cases}$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda_1 \|\theta\|_1,$$

$$\Rightarrow \hat{\theta}_T = \hat{\theta}_{Isp} = \mathbf{0} !$$

# IDENTIFIABILITY ISSUES

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Ispm} + \varepsilon_3 \end{cases}$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|Y^i - X^i \theta\|_2^2 + \lambda_1 \|\theta\|_1,$$

$$\Rightarrow \hat{\theta}_T = \hat{\theta}_{Ispm} = \mathbf{0} !$$

Use prior  $\tilde{l}_{sp} \rightsquigarrow \{\tilde{l}_{sp}^i = \tilde{l}_{sp}(\mathbf{x}^i, \mathbf{u}^i)\}$ .

# IDENTIFIABILITY ISSUES

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Ispm} + \varepsilon_3 \end{cases}$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \left( \|Y^i - X^i \theta\|_2^2 + \lambda_2 \|\tilde{l}_{sp}^i - X_{Ispm}^i \cdot \theta_{Ispm}\|_2^2 \right) + \lambda_1 \|\theta\|_1,$$

$$\Rightarrow \hat{\theta}_T = \hat{\theta}_{Ispm} = \mathbf{0} !$$

Use prior  $\tilde{l}_{sp} \rightsquigarrow \{\tilde{l}_{sp}^i = \tilde{l}_{sp}(\mathbf{x}^i, \mathbf{u}^i)\}$ .



# IDENTIFIABILITY ISSUES

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Isp} + \varepsilon_3 \\ \lambda_2 \tilde{l}_{sp} = \lambda_2 X_{Isp} \cdot \theta_{Isp} + \varepsilon_4 \end{cases}$$
$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \left( \|Y^i - X^i \theta\|_2^2 + \lambda_2 \|\tilde{l}_{sp}^i - X_{Isp}^i \cdot \theta_{Isp}\|_2^2 \right) + \lambda_1 \|\theta\|_1,$$
$$\Rightarrow \hat{\theta}_T = \hat{\theta}_{Isp} = \mathbf{0} !$$

Use prior  $\tilde{l}_{sp} \rightsquigarrow \{\tilde{l}_{sp}^i = \tilde{l}_{sp}(\mathbf{x}^i, \mathbf{u}^i)\}$ .

# IDENTIFIABILITY ISSUES

$$\left\{ \begin{array}{l} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ \mathbf{0} = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Isp} + \varepsilon_3 \\ \lambda_2 \tilde{l}_{sp} = \lambda_2 X_{Isp} \cdot \theta_{Isp} + \varepsilon_4 \end{array} \right.$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|\tilde{Y}^i - \tilde{X}^i \theta\|_2^2 + \lambda_1 \|\theta\|_1,$$

$$\Rightarrow \hat{\theta}_T = \hat{\theta}_{Isp} = \mathbf{0} !$$

Use prior  $\tilde{l}_{sp} \rightsquigarrow \{\tilde{l}_{sp}^i = \tilde{l}_{sp}(\mathbf{x}^i, \mathbf{u}^i)\}$ .

# IDENTIFIABILITY ISSUES

$$\begin{cases} Y_1 = X_{T1} \cdot \theta_T - X_D \cdot \theta_D + \varepsilon_1 \\ Y_2 = X_{T2} \cdot \theta_T + X_L \cdot \theta_L + \varepsilon_2 \\ 0 = X_T \cdot \theta_T + X_{Ispm} \cdot \theta_{Isp} + \varepsilon_3 \\ \lambda_2 \tilde{l}_{sp} = \lambda_2 X_{Isp} \cdot \theta_{Isp} + \varepsilon_4 \end{cases}$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|\tilde{Y}^i - \tilde{X}^i \theta\|_2^2 + \lambda_1 \|\theta\|_1,$$

$$\tilde{Y}^i = \begin{pmatrix} Y_1^i \\ Y_2^i \\ 0 \\ \lambda_2 \tilde{l}_{sp}^i \end{pmatrix}, \quad \tilde{X}^i = \begin{pmatrix} (X_{T1}^i)^\top & -(X_D^i)^\top & 0 & 0 \\ (X_{T2}^i)^\top & 0 & (X_L^i)^\top & 0 \\ (X_T^i)^\top & 0 & 0 & (X_{Ispm}^i)^\top \\ 0 & 0 & 0 & \lambda_2 (X_{Isp}^i)^\top \end{pmatrix},$$

Use prior  $\tilde{l}_{sp} \rightsquigarrow \{\tilde{l}_{sp}^i = \tilde{l}_{sp}(\mathbf{x}^i, \mathbf{u}^i)\}$ .

# 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

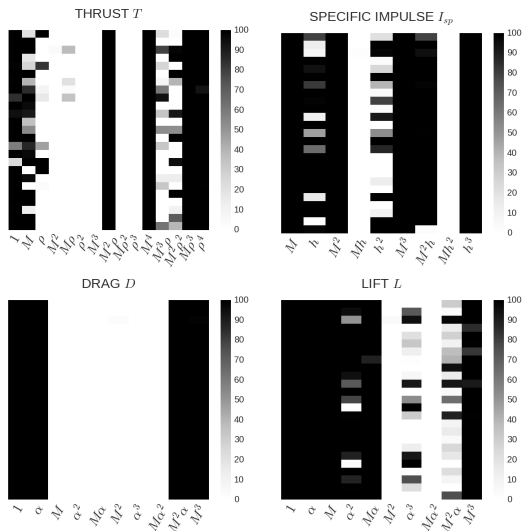
- 25 different B737-800,
- 10 471 flights = 8 261 619 observations,
- Block sparse Bolasso used for  $T$ ,  $D$ ,  $L$  and  $I_{sp}$ ,

# 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,

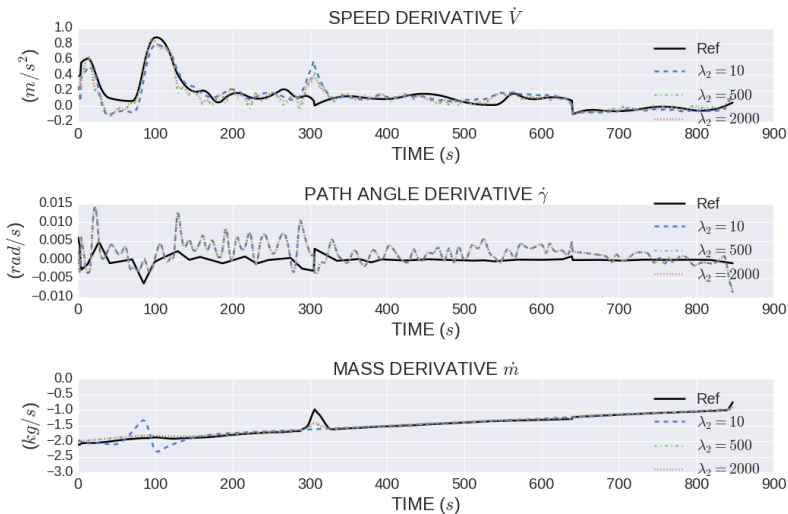


# FEATURE SELECTION RESULTS

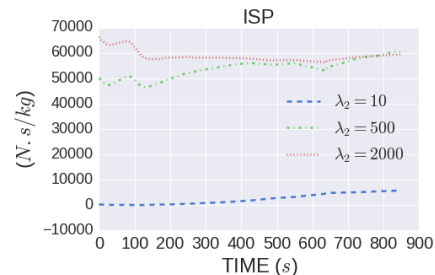
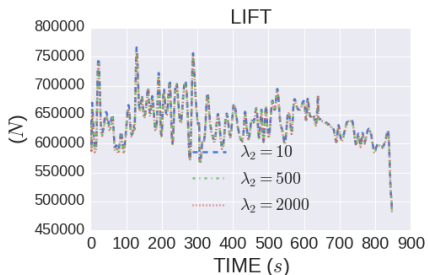
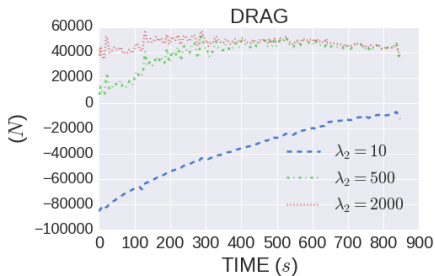
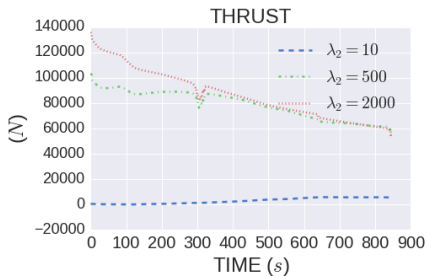


Feature selection results for the thrust, drag, lift and specific impulse models.

# EFFECT OF $\lambda_2$ ON PREDICTED STATES DERIVATIVES



# EFFECT OF $\lambda_2$ ON HIDDEN ELEMENTS



# IDENTIFICATION RESULTS ASSESSMENT

$$\dot{\mathbf{x}} - \hat{\mathbf{g}}(\mathbf{x}, \mathbf{u})$$

# IDENTIFICATION RESULTS ASSESSMENT

$$\mathbf{x} - \hat{\mathbf{x}}$$

$$\mathbf{u} - \hat{\mathbf{u}}$$

# IDENTIFICATION RESULTS ASSESSMENT

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{u} \in \mathbb{X} \times \mathbb{U}} \int_0^{t_f} (\|\mathbf{u}(t) - \mathbf{u}_m(t)\|_u^2 + \|\mathbf{x}(t) - \mathbf{x}_m(t)\|_x^2) dt \\ \text{s.t. } \dot{\mathbf{x}} = \hat{\mathbf{g}}(\mathbf{x}, \mathbf{u}) \end{aligned}$$

# IDENTIFICATION RESULTS ASSESSMENT

$$\min_{\mathbf{x}, \mathbf{u} \in \mathbb{X} \times \mathbb{U}} \int_0^{t_f} (\|\mathbf{u}(t) - \mathbf{u}_m(t)\|_u^2 + \|\mathbf{x}(t) - \mathbf{x}_m(t)\|_x^2) dt$$

s.t.  $\dot{\mathbf{x}} = \hat{\mathbf{g}}(\mathbf{x}, \mathbf{u})$

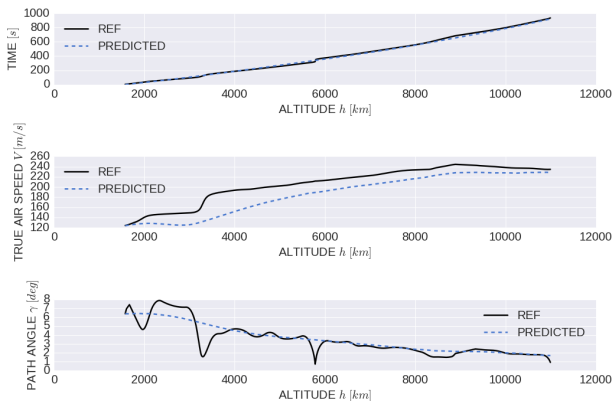


FIGURE : Obtained using BOCOP [Bonnans et al., 2017]

# IDENTIFICATION RESULTS ASSESSMENT

$$\min_{\mathbf{x}, \mathbf{u} \in \mathbb{X} \times \mathbb{U}} \int_0^{t_f} (\|\mathbf{u}(t) - \mathbf{u}_m(t)\|_u^2 + \|\mathbf{x}(t) - \mathbf{x}_m(t)\|_x^2) dt$$

s.t.  $\dot{\mathbf{x}} = \hat{\mathbf{g}}(\mathbf{x}, \mathbf{u})$

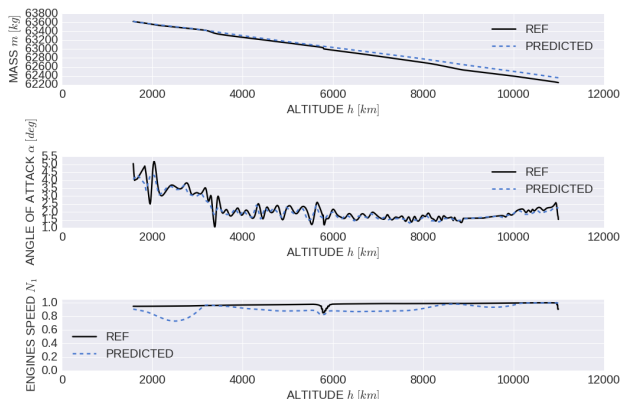






FIGURE : Obtained using BOCOP [Bonnans et al., 2017]



**THANK YOU FOR YOUR ATTENTION !**

# REFERENCES

-  Bach, F. (2008).  
Bolasso: model consistent Lasso estimation through the bootstrap.  
pages 33–40. Proceedings of the 25th International Conference on  
Machine Learning (ICML).
-  Bonnans, J. F., Giorgi, D., Grelard, V., Heymann, B., Maindrault, S.,  
Martinon, P., Tissot, O., and Liu, J. (2017).  
Bocop – A collection of examples.  
Technical report, INRIA.
-  Efron, B. et al. (2004).  
Least angle regression.  
Annals of Statistics, 32:407–499.
-  Tibshirani, R. (1994).  
Regression shrinkage and selection via the Lasso.  
Journal of the Royal Statistical Society, 58:267–288.