Distributed Optimization for Aircraft Fleet Monitoring

Eric Chu, Dimitry Gorinevsky, and Stephen Boyd

May 10, 2011

Introduction

- aircraft fleet monitoring problem
  - individualized models for each aircraft and each flight
- distributed computing (optimization)

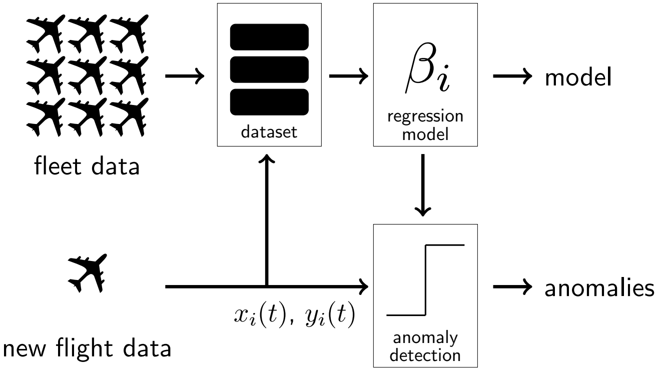
Problem statement

- massive amounts of Flight Operations Quality Assurance (FOQA) data is collected
- goal: use data to monitor performance of individual aircraft
  - reduce maintenance costs
  - detect fuel efficiency problems

Approach

- learn multivariate models from the data
- use model to perform anomaly detection
- features of the approach:
  - handles flight-to-flight and tail-to-tail variability (fixed effects)
  - handles large-scale, distributed dataset
  - scalable through distributed computing

Problem setup



Model identification (regression)

- find model β_i and trend $a_i(t)$ such that

$$y_i(t) = \beta_i x_i(t) + a_i(t) + r_i(t) \quad \text{for all } i, t,$$

  where $r_i(t)$ is model fit residual

- the models β_i, the trends $a_i(t)$, and the residuals $r_i(t)$, $t = 1, \ldots, T_i$, $i = 1, \ldots, N$, are found by solving

$$\text{minimize} \quad J^{\text{res}} + \kappa J^{\text{shift}} + \mu J^{\text{unit}}$$

Model Identification (regression)

- we choose

$$\begin{aligned} J^{\text{res}} &= \sum_{i=1}^{N} \sum_{t=1}^{T_i} \|r_i(t)\|_2^2 \\ J^{\text{shift}} &= \sum_{i=1}^{N} \sum_{t=2}^{T_i} \|a_i(t) - a_i(t-1)\|_2^2 \\ J^{\text{unit}} &= \sum_{i=1}^{N} \|\beta_i - \bar{\beta}\|_F^2 \end{aligned}$$

where $\|\cdot\|_F$ denotes the Frobenius norm ($\|A\|_F = (\sum_{i,j} A_{ij}^2)^{1/2}$)

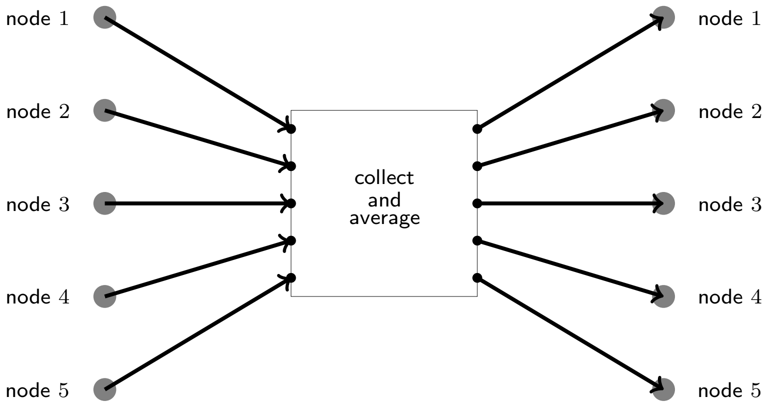
One small hitch...

- **example**: FOQA data with regressors $x_i(t)$ of dimension $n = 20$ and performance vector $y_i(t)$ of dimension $m = 10$, $N = 100$ aircraft in the fleet, 500 flights per year over 5 years
- total number of measurements $M^{\rm tot}$ and variables $N^{\rm tot}$:

$$M^{\rm tot} = 2{,}500{,}000, \qquad N^{\rm tot} = 2{,}520{,}000$$

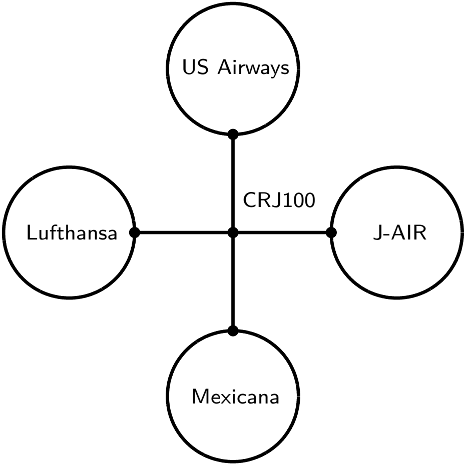
solve time: approximately 500 years using standard matrix arithmetic

Distributed computing



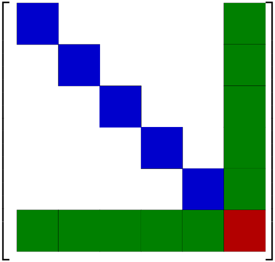
Distributed computing

additional benefit: data privacy



Distributed algorithm

- normal equations have *block arrow* structure:

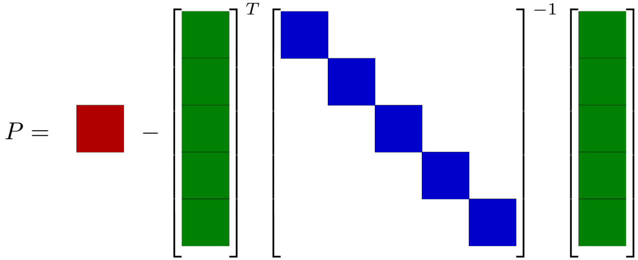


$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix}$$

- blue blocks are easily invertible
- suggests we should solve for x_6 first, then $x_1, \ldots, x_5$ can be found in parallel

Distributed algorithm

- to solve for x_6 we must first compute the Schur complement matrix



- for second term, invert blocks of block diagonal matrix in parallel
- details can be found in *Scalable Statistical Monitoring of Fleet Data*, to appear in *Proceedings of IFAC World Congress*, 2011

Anomaly detection

- the trained model yields $\beta_i^\star$, $\bar{\beta}^\star$, $a_i^\star(t)$, and $r_i^\star(t)$
  - $\beta_i^\star$ are individual models for each aircraft
  - $\bar{\beta}^\star$ is the average model for the fleet
  - $a_i^\star(t)$ are the trend in the offsets
  - $r_i^\star(t)$ are the model prediction residuals
- wish to detect three types of anomalies:

  A1 performance anomaly: $r_i^\star(t)$ becomes 'large'

  A2 performance shift: $a_i^\star(t)$ becomes 'large'

  A3 anomalous unit: $\beta_i^\star - \bar{\beta}^\star$ becomes 'large'

Anomaly detection

- measure 'size' of monitored signal by using
  - A1 $T^{\mathrm{res}} = (r_i^\star(t) - \bar{r}^\star)^T (\Sigma^{\mathrm{res}})^{-1} (r_i^\star(t) - \bar{r}^\star)$
  - A2 $T^{\mathrm{shift}} = (a_i^\star(t) - \bar{a}^\star)^T (\Sigma^{\mathrm{shift}})^{-1} (a_i^\star(t) - \bar{a}^\star)$
  - A3 $T^{\mathrm{unit}} = (\beta_i^\star - \bar{\beta}^\star)^T (\Sigma^{\mathrm{unit}})^{-1} (\beta_i^\star - \bar{\beta}^\star)$
- declare anomaly if any of these exceed threshold on Hotelling T^2 chart

Example

- $m = 1$ performance measurement for an angle-of-attack channel, $n = 4$ input measurements, $N = 200$ aircraft, and $T_i = 5000$ for all $i = 1, \ldots, N$

- generate observations $y_i(t)$ via

$$y_i(t) = \beta_i x_i(t) + a_i(t) + r_i(t)$$

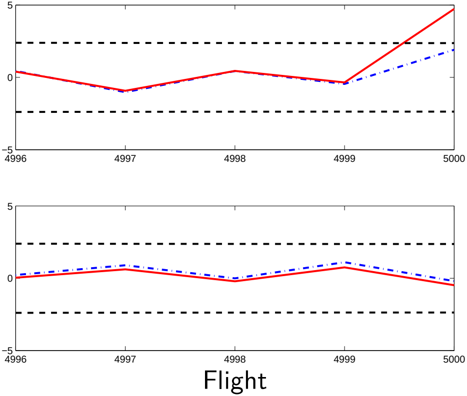
  where β_i, $r_i(t)$, and $x_i(t)$ drawn from Gaussian distributions, $a_i(t) = 0$ for all flights without faults

Example anomalies

- inject faults:
  - A1 a large value of $r_i(t)$
  - A2 ramp up $a_i(t)$ over successive flights of aircraft i
  - A3 a large value of $\beta_i - \bar{\beta}$
- out of 200 aircraft, seed 2 faults of each kind

Example results

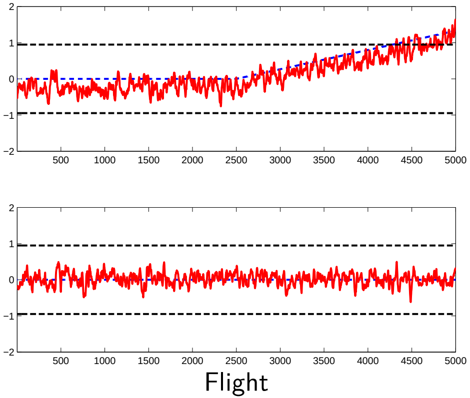
plot of $r_i(t)$ for last 5 flights



blue is ground truth; red is fitted curve

Example results

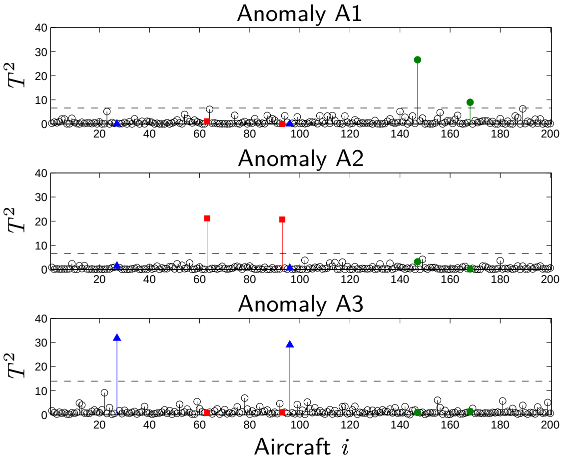
plot of $a_i(t)$ for all flights



blue is ground truth; red is fitted curve

Example results

T^2 values for last flight ($t = 5000$) of all 200 aircraft



Conclusions

- fleet monitoring applied to FOQA data allows us to find:
  - abnormal flights
  - unusual performance trends
  - aircraft performing differently from the rest of the fleet
- demonstrated implementation for 1 million flights in simulation
  - sequential implementation on laptop takes minutes for 1 channel
- scalable computational performance achieved by distributed optimization

Next steps

- demonstrate for real FOQA data
- technology transition: new NASA SBIR project with Mitek Analytics
- more in Dimitry's talk on the architecture tomorrow

Questions

Piecewise constant offsets

- piecewise constant offsets corresponds to onset of a persistent anomaly
- we might use a different choice of $J^{\rm shift}$

$$J^{\rm shift} = \sum_{i=1}^{N} \sum_{t=2}^{T_i} \|a_i(t) - a_i(t-1)\|_2$$

- norm is *not* squared
- promotes sparsity in change of offset (*i.e.*, faults are rare)
- with this $J^{\rm shift}$, no analytical solution

Distributed optimization

- basic idea from quadratic case suggests an iterative method with three steps:

Step 1 solve a local problem

Step 2 collect solutions and 'take average', then broadcast

Step 3 update local solution

- the alternating direction method of multipliers (ADMM) provides an algorithm for distributing computation across data following above pattern[1]

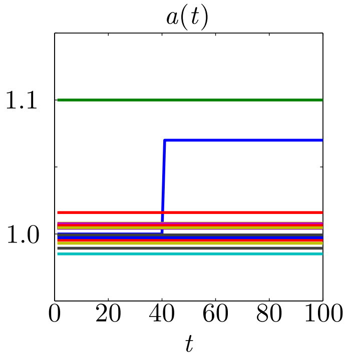
- additionally provides data privacy

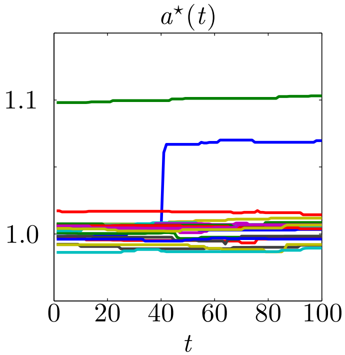
[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. To appear in *Foundations and Trends in Machine Learning*, 2011.

Aircraft example with piecewise constant offsets

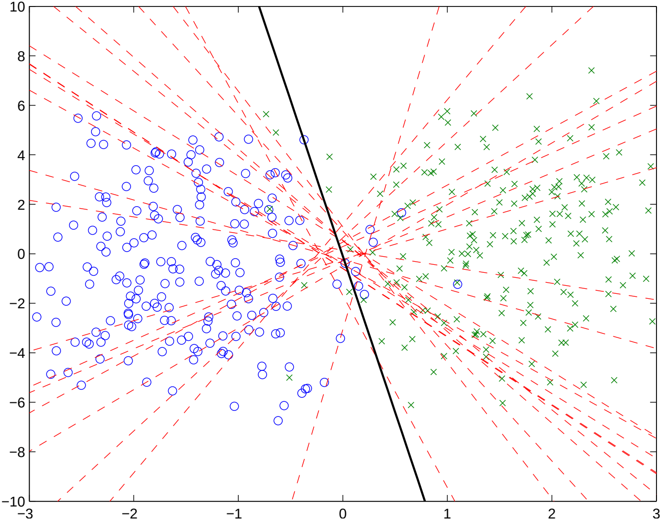
- same optimization problem, with different J^{shift}
- $m = n = 2$, $T = 100$, $N = 20$
- seeded one anomalous aircraft (anomaly A3) and one offset (anomaly A2)
- solved problem using ADMM

Aircraft example

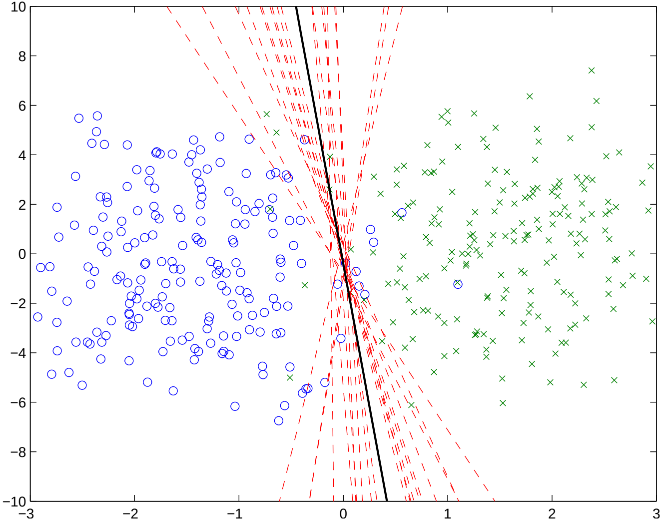




SVM example



SVM example



SVM example