Model Based Emissions Control

Matthew Egginton
Jamie Lukins
Jack Skipper
Supervisors: Tim Sullivan, Florian Theil

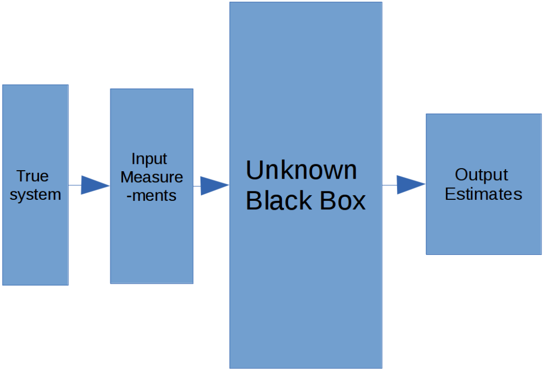
Report Presentation

Mathematics and Statistics Centre for Doctoral Training

THE UNIVERSITY OF WARWICK

The Problem

The aim is to model a black box type scenario, where one has input measurements with noise, and output estimate with error.



Outline

- Formulation of Problem
- Real World Application
- Least Squares Minimisation
- Kalman Filtering
- Further Work

More Mathematical Formulation

- One has a time series $x = (x^i)$ of input measurements where $x^i \in \mathbb{R}^d$ for some d and $x^i \sim \mathcal{N}(\widetilde{x}^i, Q^i)$ for some positive definite symmetric matrices Q^i and $\widetilde{x}^i \in \mathbb{R}^d$.
- We assume that the true system is given by $\widetilde{y} = (\widetilde{y}^i)$.
- One also has output measurements y^i where $y^i \sim \mathcal{N}(\widetilde{y}^i, P^i)$ where P^i are positive definite symmetric matrices.
- No prior knowledge of the physical system is supposed.

The aim is to recreate the function $f : \mathbb{R}^d \to \mathbb{R}^d$ that maps the input x^i to the true system $\widetilde{y}^i$.

The Problem (The practical outlook)

This Problem can be seen to be in two parts:

1. An inverse problem where there is input and output data with noise from a black-box and finding or approximating the black-box is the task.
2. This is to recreate the black-box object and possibly improve so that with new input data the output data can be recreated in real time or faster.

Once the first method is complete then the second problem can be approached.

Why is this Useful?

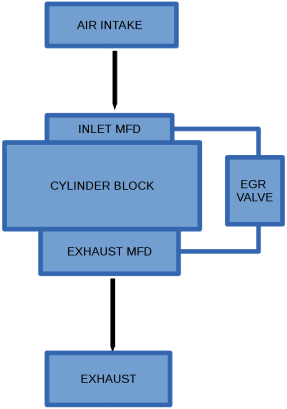
In an industrial application, one has to consider data analysis, in particular:

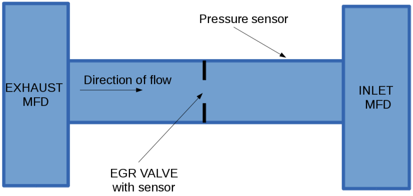
1. Smoothing noisy data (inherent noise associated to measuring)
2. Analysing relationships between variables (for example companies not communicating to keep intellectual property for gains).

Thus companies from just input and output data want to invert the process to gain the same knowledge.

Emissions optimisation-Engine

One technique to reduce emissions is to recirculate exhaust gases.





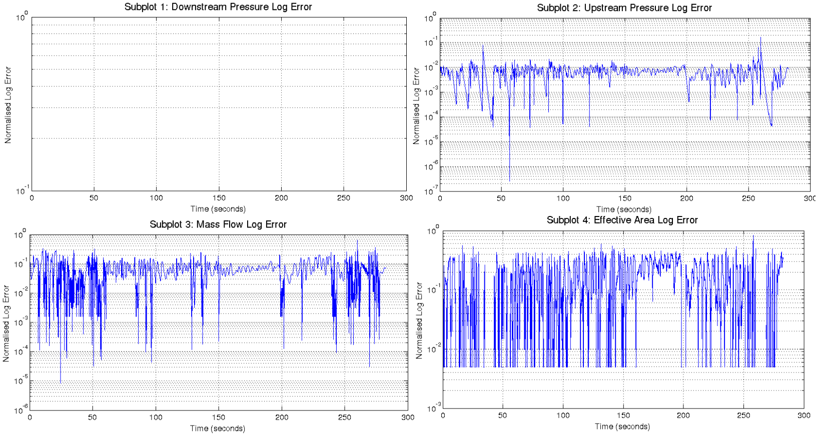
Emissions optimisation-Application

More accurate estimates of the system are desired to improve optimisation of emissions.

This system is characterised by upstream and downstream pressure denoted by P_U and P_D resp., the mass flow through the system M, and the effective area of the valve A.

The black box model is suitable as we were kindly given input measurements and output estimates of these four data streams to model.

Data



Subplot 1: Downstream Pressure Log Error
Normalised Log Error
Time (seconds)
Subplot 2: Upstream Pressure Log Error
Normalised Log Error
Time (seconds)
Subplot 3: Mass Flow Log Error
Normalised Log Error
Time (seconds)
Subplot 4: Effective Area Log Error
Normalised Log Error
Time (seconds)

Least Squares Minimisation

First introduced by Legendre and Gauss [1]. Produces an approximation of the function that minimises the least squares distance between the approximation and the function.

Since the method doesn't try to exactly fit the data values, it is suitable for removing noise in the input or output values.

For a choice of basis, this becomes a linear least squares regression problem which is simpler to computationally solve.

Least Squares Minimisation

Given data $(x^i, y^i) \in \mathbb{R}^{2d}$ for $i = 1, \dots, N$ one aims to find the function $f : \mathbb{R}^d \to \mathbb{R}$ in some space V which minimises

$$S = \sum_{i=1}^{N} \left| f(x^i) - y_k^i \right|^2$$

If $\dim V < \infty$ with basis $\{\phi_j\}$ then one can write the minimising coefficients as

$$\beta = (X^T X)^{-1} X^T y_k$$

where $X_{ij} = \phi_j(x^i)$ using the Gauss-Markov theorem.

Probabilistic framework

Again given data (x^i, y^i) as before, and choosing some parameters β one supposes that there is a relationship of the form

$$y_k^i - f(x^i|\beta) \sim \mathcal{N}(0, C^i)$$

Thus $y_k^i - f(x^i|\beta) \propto e^{-|f(x^i|\beta) - y_k^i|_2^2}$. Finding the Maximum Likelihood estimator then corresponds to maximising

$$\mathcal{L}(x^1, \ldots, x^N) = \prod_{i=1}^{N} f(x^i|\beta) \propto e^{-\sum_{i=1}^{N} |f(x^i|\beta) - y_k^i|_2^2}$$

and the maximum of this coincides with the minimum least squares distance of $f(x^i|\beta)$ with y_k^i for all i.

Choosing a basis

We see that if we choose a basis of the function space $\{\phi_j\}$ then the least square regression above becomes a simpler linear regression problem with the matrix $\mathbf{C}$ made of elements of the the evaluations of $\phi_j(x^i)$

$$\sum_{i=1}^{N} |f(x^i) - y_k^i|^2 = \sum_{i=1}^{N} \left| \sum_{j=1}^{K} \beta_j \phi_j(x^i) - y_k^i \right|^2 = \|\mathbf{C}\beta - \mathbf{y}_k\|_2^2$$

Choice of basis

Considering this result on existence and uniqueness of a minimiser is independent of the basis, one is free to choose the space over which one minimises.

Local Support

1. Gives the matrix X as sparse, so computation is less involved.
2. A poor approximation if data is found only in a small region of space.

Global Support

1. More involved computation as to evaluate at a point, all basis functions must be used.
2. Useful if data is clustered, and doesn't fill the whole space.
3. Can predict outside of the data domain.

Choice of basis

If it is the case that the data is clustered, and prediction outside these ranges is highly desirable, the a global basis consisting of a partial Fourier basis could be used. This would take the form of

$$\cos\left(\frac{x_1 - s_1}{2n_1\pi} i\right) \cos\left(\frac{x_2 - s_2}{2n_2\pi} j\right) \cos\left(\frac{x_3 - s_3}{2n_3\pi} k\right) \cos\left(\frac{x_4 - s_4}{2n_4\pi} l\right)$$

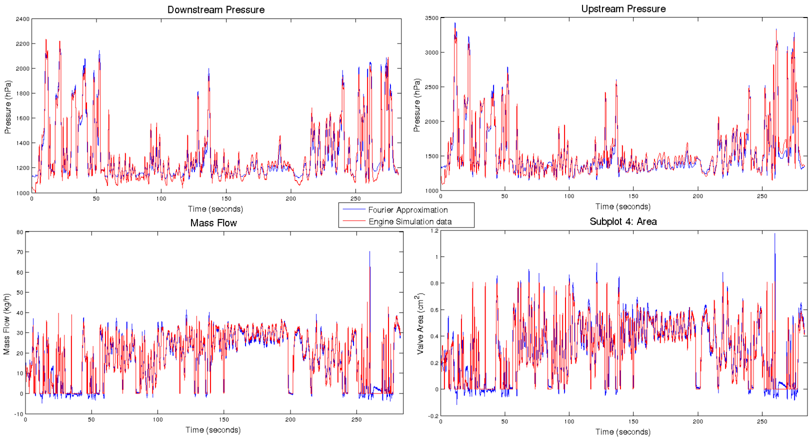
where i, j, k, l vary from $1, \ldots, K$.

Conditioning of the Problem

One also needs to ensure that the data is not being over-fitted, namely one chooses a dimension of the minimisation space that is very much smaller than the number of data points used in the minimisation.

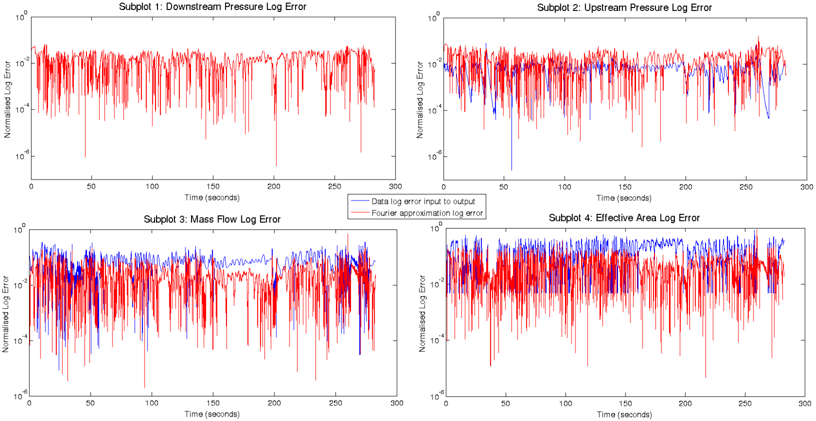
There are methods of functional principal component analysis (P.C.A.) that give values of the largest dimension of space that is suitable for consideration. [2]

Results



Downstream Pressure
Pressure (hPa)
Time (seconds)
Upstream Pressure
Pressure (hPa)
Time (seconds)
Fourier Approximation
Engine Simulation data
Mass Flow
Mass Flow (kg/h)
Time (seconds)
Subplot 4: Area
Valve Area (cm^2)
Time (seconds)

Results



Subplot 1: Downstream Pressure Log Error
Subplot 2: Upstream Pressure Log Error
Subplot 3: Mass Flow Log Error
Subplot 4: Effective Area Log Error
Normalised Log Error
Time (seconds)
Data log error input to output
Fourier approximation log error

Kalman Filtering

Problem: black-box object recreation

Kalman filtering is an algorithm that tries to reconcile outputs from a mathematical model of a physical system and observations of the same system.

Filtering combines the approximation function for the black-box with the observations in such a way as to smooth out noise coming from inaccurate observations in an effort to more accurately estimate the true state of the system at the present time.

Kalman Filtering cont.

Consider the process $x \in \mathbb{R}^d$ given by the stochastic difference equation, for $A \in \mathbb{R}^{d \times d}$

$$x^i = Ax^{i-1} + w^{i-1} \tag{1}$$

and consider also measurements, for $H \in \mathbb{R}^{l \times d}$,

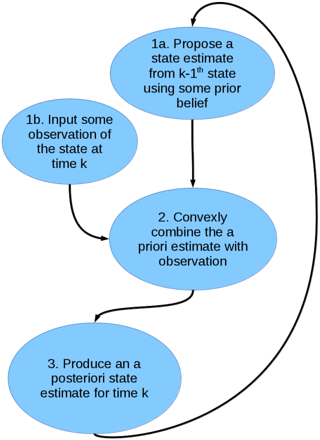
$$z^i = Hx^i + v^i \tag{2}$$

where w and v represent noise terms, and are assumed to be independent multivariate normals with distributions

$$w \sim \text{MVN}(0, Q) \qquad v \sim \text{MVN}(0, R). \tag{3}$$

It should be recognised that this is somewhat of a special case of the usual Kalman filter, [3], because one may well expect the noise to vary over time, and so then the above would be indexed by i.

Kalman Filtering cont.



The algorithm is as follows

1. At time n, given the previous a posteriori estimates $x^{n-1}, \ldots, x^{n-l}$ of the system, a prediction $\hat{x}^{n|n-1}$ is made based upon the prior belief or physical dynamics of the system at time n.
2. The system is observed at time n and this observation z^n is used to correct the a priori estimate $\hat{x}^{n|n-1}$ and produce an updated estimate $\hat{x}^{n|n}$
3. Repeat for time $n+1$.

MAS DOC

Kalman Filter Update Equations

In the case, as above, where one has Gaussian noise, one can explicitly write the estimation formulas as follows:

$$
\begin{aligned}
\hat{x}^{k|k-1} &= A\hat{x}^{k-1|k-1} \\
P^{k|k-1} &= AP^{k-1|k-1}A^T + Q \\
K^k &= P^{k|k-1}H^T\left(HP^{k|k-1}H^T + R\right)^{-1} \\
\hat{x}^{k|k} &= \hat{x}^{k|k-1} + K^k\left(z^i - H\hat{x}^{k|k-1}\right) \\
P^{k|k} &= (I - K^kH)P^{k|k-1}
\end{aligned}
$$

Extended Kalman Filtering

A modification of the Kalman filter where one has a non-linear update equation of the form

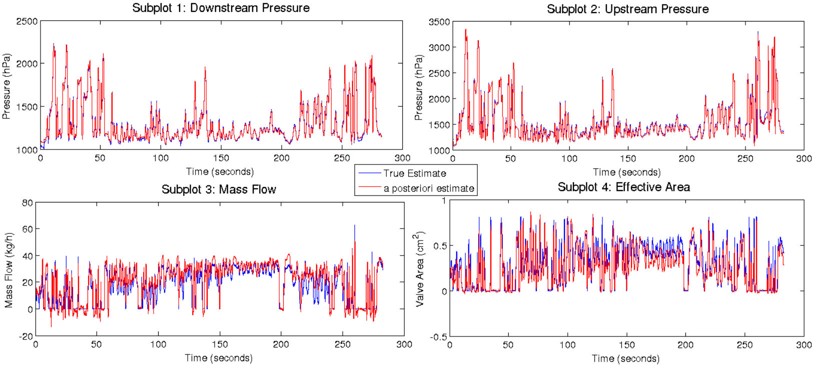
$$x^{i+1} = f(x^i) + w^i$$

where w is again the noise term, and we have measurements

$$z^i = h(x^i) + v^i.$$

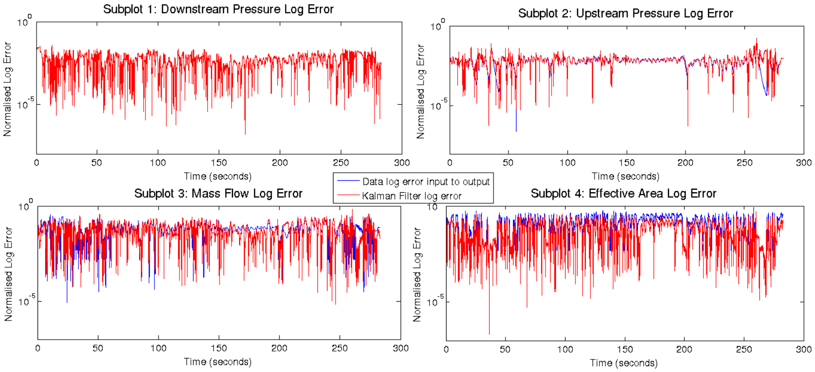
One linearises the f and the h with the derivative and then uses the Kalman filter as above with this linearisation.

Results



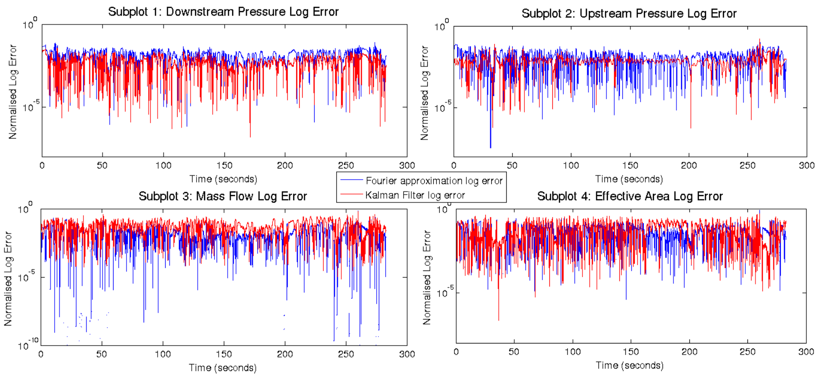
Subplot 1: Downstream Pressure
Pressure (hPa)
2500
2000
1500
1000
0
50
100
150
200
250
300
Time (seconds)
Subplot 2: Upstream Pressure
Pressure (hPa)
3500
3000
2500
2000
1500
1000
0
50
100
150
200
250
300
Time (seconds)
True Estimate
a posteriori estimate
Subplot 3: Mass Flow
Mass Flow (kg/h)
80
60
40
20
0
-20
0
50
100
150
200
250
300
Time (seconds)
Subplot 4: Effective Area
Valve Area (cm^2)
1
0.5
0
-0.5
0
50
100
150
200
250
300
Time (seconds)

Results



Subplot 1: Downstream Pressure Log Error
Subplot 2: Upstream Pressure Log Error
Subplot 3: Mass Flow Log Error
Subplot 4: Effective Area Log Error
Normalised Log Error
Time (seconds)
Data log error input to output
Kalman Filter log error

Results



Subplot 1: Downstream Pressure Log Error
Subplot 2: Upstream Pressure Log Error
Subplot 3: Mass Flow Log Error
Subplot 4: Effective Area Log Error
Normalised Log Error
Time (seconds)
Fourier approximation log error
Kalman Filter log error

MCMC

The object is to find an unknown function f in a Banach space V, possibly the space of continuous functions.

- A Gaussian prior distribution $\mathcal{N}(m, C)$ where $C : V \to V$ is the covariance operator is a natural choice to model an unknown function
- Karhunen-Loève is used to sample from this distribution
- Implement a Markov Chain Monte Carlo (MCMC) algorithm to sample from the posterior.

MCMC example

Click above to play video

Other Function Spaces

The physics of the problem may well suggest a more natural space to consider minimisation over.
Other global basis functions that could be better especially if the function was in higher dimensions are radial functions, defined as follows:

Given a smooth function $f : \mathbb{R}^d \to \mathbb{R}^d$ and points Ξ where one knows the evaluations of this function $f(\xi_i)$ for these values, then a radial basis function is the composition of a continuous function $\phi\colon \mathbb{R}_+ \to \mathbb{R}$ with the Euclidean norm. In other words one can write a function ν as

$$\nu(x) = \sum_{\xi \in \Xi} \lambda_\xi \phi\left(\|x - \xi\|_2\right)$$

for some λ_ξ.

Constrained Minimisation

One could instead consider minimisation of the form

$$\text{Minimise } \|C\alpha - d\|_2^2 \text{ in } V \text{ subject to } A\alpha \leq \gamma$$

where the $A\alpha \leq \gamma$ corresponds to some estimate on the norm of the gradient of the function. This is equivalent to the regularisation problem of Tychonov.

Theorem (Tychonov)

Let $A : \mathcal{H} \to \mathcal{K}$ be a linear operator between Hilbert Spaces such that $R(A)$ is a closed subspace of $\mathcal{K}$. Let $Q : \mathcal{H} \to \mathcal{H}$ be self adjoint and positive definite, and $b \in \mathcal{K}$ and $x_0 \in \mathcal{H}$ be given as well. Then

$$\hat{x} \in \mathrm{argmin}_{x \in \mathcal{H}} \left(\|Ax - b\|^2 + \|x - x_0\|_Q^2 \right) \iff (A^\star A + Q)\hat{x} = A^\star b + Qx_0$$

References

Carl-Friedrich Gauss.
Theoria combinationis observationum erroribus minimis obnoxiae.-Gottingae, Henricus Dieterich 1823.
Henricus Dieterich, 1823.

Peter Hall, Hans-Georg Müller, and Jane-Ling Wang.
Properties of principal component methods for functional and longitudinal data analysis.
The Annals of Statistics, 34(3):1493–1517, 06 2006.

R. E. Kalman.
A new approach to linear filtering and prediction problems.
Journal of basic Engineering, 82(1):35–45, 1960.

Thank you