# Optimisation perspective on state space models

- recursive least squares
- block least squares
- Kalman filter as an optimisation problem
- relation to additive models
- extensions to non-Gaussian state and observation noise

D. Semenovich, University of New South Wales

# Least squares

standard least squares problem:

$$\underset{\mathbf{b}}{\text{minimise}} \quad \|X\mathbf{b} - \mathbf{y}\|_2^2 = \textstyle\sum_{i=1}^m (\mathbf{x}_i^T \mathbf{b} - y_i)^2$$

where $\mathbf{x}_i \in \mathbb{R}^n,\ i = 1, .., m$ are the rows of $X \in \mathbb{R}^{m \times n}$

- $\mathbf{b} \in \mathbb{R}^m$ is the parameter vector to be estimated
- each pair $(y_i, \mathbf{x}_i)$ corresponds to an observation
- solution is given by:

$$\mathbf{b}^* = (X^T X)^{-1} X^T \mathbf{y} = \left( \sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=1}^m y_i \mathbf{x}_i$$

# Incremental least squares

- assume pairs $(y_i, \mathbf{x}_i)$ become available over time, i.e. $m$ increases, then we can express the least squares solution $\mathbf{b}^*$ as a function of $m$

- this corresponds to the following state estimation problem:

$$\mathbf{b}_{i+1} = \mathbf{b}_i \qquad\qquad \begin{aligned} y_i &= \mathbf{x}_i^T \mathbf{b}_i + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0,1) \end{aligned}$$

- one approach is to directly compute:

$$\mathbf{b}^*(m) = \left( \sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=1}^{m} y_i \mathbf{x}_i$$

# Recursive incremental least squares

- if speed is important (hard real time constraints, Monte Carlo simulation etc) we can compute $\mathbf{x}^*(m)$ recursively:

- initialise $\Sigma(0) = 0 \in \mathbb{R}^{n \times n}, \mathbf{r}(0) = 0 \in \mathbb{R}^n$

- then we have:

$$\Sigma(m+1) = \Sigma(m) + \mathbf{x}_{m+1}\mathbf{x}_{m+1}^T \quad \mathbf{r}(m+1) = \mathbf{r}(m) + y_{m+1}\mathbf{x}_{m+1}$$

- if $\Sigma(m)$ is invertible we can calculate $\mathbf{b}^*(m) = \Sigma^{-1}(m)\mathbf{r}(m)$

# Rank one update

- can further speed up calculation by applying **rank one update** to $\Sigma^{-1}$:

$$(\Sigma + \mathbf{x}\mathbf{x}^T)^{-1} = \Sigma^{-1} - \frac{1}{1 + \mathbf{x}^T \Sigma^{-1} \mathbf{x}} (\Sigma^{-1}\mathbf{x})(\Sigma^{-1}\mathbf{x})^T$$

- reduces computational cost of computing $\Sigma^{-1}(m+1)$ from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$

- this is a computational trick and is conceptually largely irrelevant, yet the tratidional presentation of state estimation (Kalman filter etc) makes it appear central.

# Block least squares

partition design matrix $X$ and response vector $\mathbf{y}$ into $m$ row blocks:

$$X = \begin{bmatrix} X_1 \\ \cdots \\ X_m \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \cdots \\ \mathbf{y}_m \end{bmatrix}$$

then we can transform the least squares problem:

$$\underset{\mathbf{b}}{\text{minimise}} \quad \|X\mathbf{b} - \mathbf{y}\|_2^2$$

to the following equivalent form with $m$ copies of the parameter vector:

$$\begin{array}{ll} \underset{\mathbf{b}_1,\ldots,\mathbf{b}_m,\mathbf{z}}{\text{minimise}} & \sum_{i=1}^m \|X_i\mathbf{b}_i - \mathbf{y}_i\|_2^2 \\ \text{subject to} & \mathbf{b}_i = \mathbf{b}_{i+1}, \quad i = 1, \ldots, m-1 \end{array}$$

## State space estimation as an optimisation problem

- The above block least squares problem corresponds to this state space model (same as recursive least squares earlier but with blocks of observations in each time period):

$$\mathbf{b}_{i+1} = \mathbf{b}_i \qquad \begin{aligned} \mathbf{y}_i &= X_i \mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{\epsilon}_i &\sim \mathcal{N}(0, I) \end{aligned}$$

- now add state transition noise (or equivalently, relax the equality constraints on the parameter vectors):

$$\begin{aligned} \mathbf{b}_{i+1} &= \mathbf{b}_i + \boldsymbol{\nu}_i \\ \boldsymbol{\nu}_i &\sim \mathcal{N}(0, I) \end{aligned} \qquad \begin{aligned} \mathbf{y}_i &= X_i \mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{\epsilon}_i &\sim \mathcal{N}(0, I) \end{aligned}$$

- we can now write down the resulting estimation problem as follows:

$$\underset{\mathbf{b}_1,\ldots,\mathbf{b}_m}{\text{minimise}} \quad \textstyle\sum_{i=1}^{m} \|X_i \mathbf{b}_i - \mathbf{y}_i\|_2^2 + \sum_{i=1}^{m-1} \|\mathbf{b}_{i+1} - \mathbf{b}_i\|_2^2$$

# Least squares formulation of state estimation

- state estimation can also be expressed as a standard least squares problem:

$$\underset{\mathbf{b}_1,\ldots,\mathbf{b}_m}{\text{minimise}} \left\| \begin{bmatrix} X_1 & & & & \\ -I & I & & & \\ & & \ddots & & \\ & & & -I & I \\ & & & & X_m \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_{m-1} \\ \mathbf{b}_m \end{bmatrix} - \begin{bmatrix} \mathbf{y}_1 \\ 0 \\ \vdots \\ 0 \\ \mathbf{y}_m \end{bmatrix} \right\|_2^2$$

- the above formulation performs both "filtering" and "smoothing" conditional on all the observations up to time $m$.

- if new information becomes available, augment the optimisation problem and solve again to obtain new estimates $\mathbf{b}^* = (\mathbf{b}_1^*, \ldots, \mathbf{b}_m^*, \mathbf{b}_{m+1}^*)$

# State estimation and additive models

- the simplified state estimation model is closely related to additive models.

- we could fit an additive model with a a single smooth effect in the time dimension by solving the following optimisation problem:

$$\underset{b_1,\ldots,b_m}{\text{minimise}} \quad \sum_{i=1}^{m} \|\mathbf{1}b_i - \mathbf{y}_i\|_2^2 + \sum_{i=1}^{m-1} \|b_{i+1} - b_i\|_2^2$$

- which is the same as state space estimation applied to a constant term.

# General linear Gaussian setting

- state equations for the general linear Gaussian state space model are as follows:

$$\begin{array}{ll} \mathbf{b}_{i+1} = F\mathbf{b}_i + \boldsymbol{\nu}_i & \qquad \mathbf{y}_i = X_i^T\mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{\nu}_i \sim \mathcal{N}(0, \Sigma_\nu) & \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \Sigma_\epsilon) \end{array}$$

- denoting $\|\mathbf{a}\|_P = (\mathbf{a}^T P \mathbf{a})^{\frac{1}{2}}$, $P$-quadratic norm for a positive semidefinite matrix $P$, the resulting estimation problem is:

$$\underset{\mathbf{b}_1,\ldots,\mathbf{b}_m}{\text{minimise}} \quad \textstyle\sum_{i=1}^{m} \|X_i\mathbf{b}_i - \mathbf{y}_i\|^2_{\Sigma_\epsilon^{-1}} + \sum_{i=1}^{m-1} \|\mathbf{b}_{i+1} - F\mathbf{b}_i\|^2_{\Sigma_\nu^{-1}}$$

## Non-Gaussian observations

- we can use any convex loss for the observations, such as quantile, logistic, Poisson, Huber etc:

$$\underset{\mathbf{b}_1,\ldots,\mathbf{b}_m}{\text{minimise}} \quad \textstyle\sum_{i=1}^{m} \mathcal{L}(X_i\mathbf{b}_i, \mathbf{y}_i) + \sum_{i=1}^{m-1} \|\mathbf{b}_{i+1} - \mathbf{b}_i\|_2^2$$

- for example for the quantile loss, where $\tau$ is the quantile of interest we would have:

$$\mathcal{L}(X_i\mathbf{b}_i, \mathbf{y}_i) = \psi(\mathbf{y}_i - X_i\mathbf{b}_i), \quad \psi(\mathbf{a}) = (\tau - 1)\textstyle\sum_{a_j<0} a_j + \tau \sum_{a_j \geq 0} a_j$$

# Non-Gaussian state noise

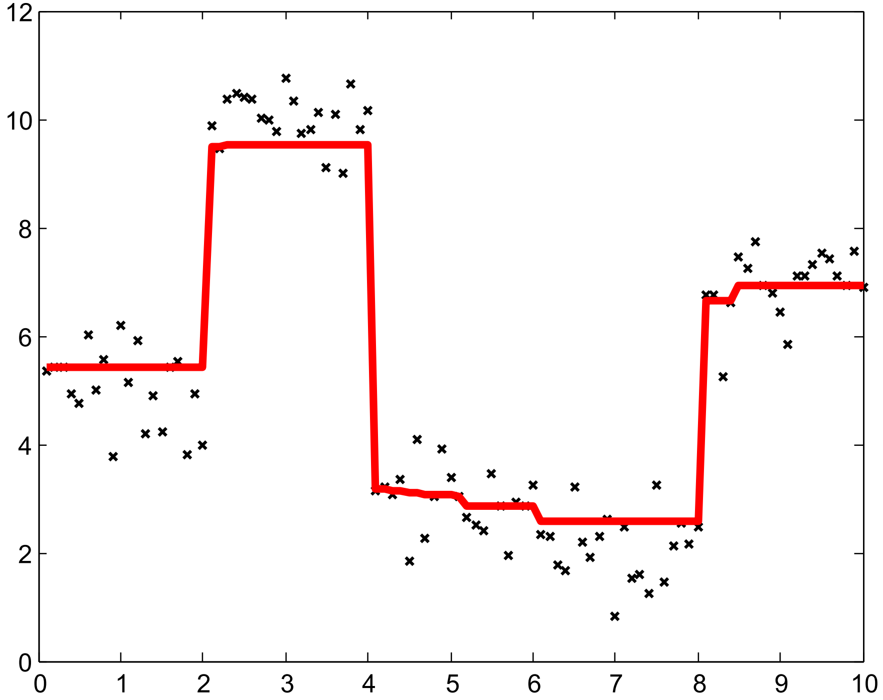
- it may be beneficial to apply $\ell_1$ norm penalty to state changes provided most of the time parameters stay constant with occasional large jumps:

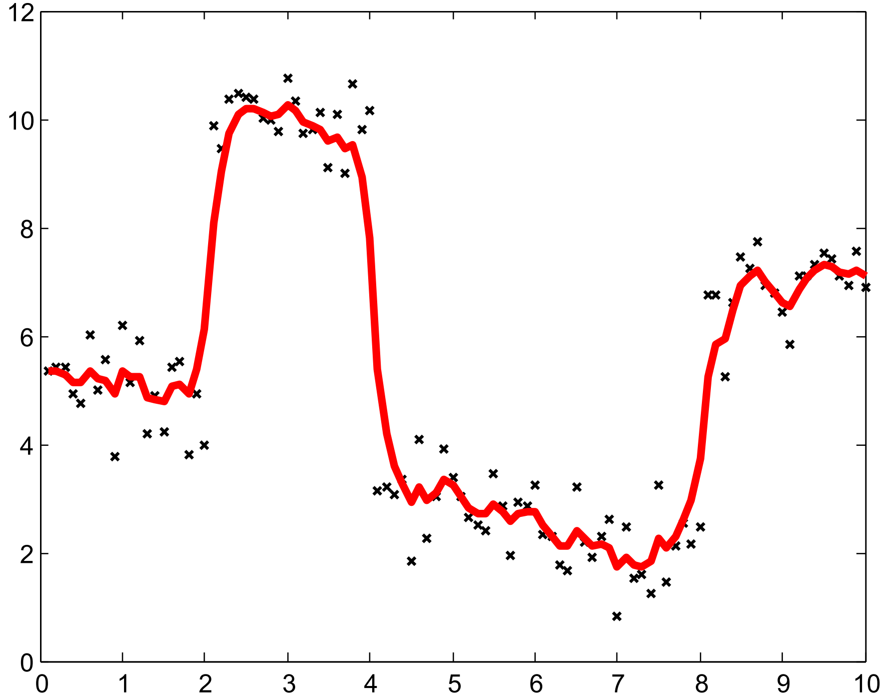
$$\underset{\mathbf{b}_1,\ldots,\mathbf{b}_m}{\text{minimise}} \quad \sum_{i=1}^{m} \mathcal{L}(X_i \mathbf{b}_i, \mathbf{y}_i) + \sum_{i=1}^{m-1} \|\mathbf{b}_{i+1} - \mathbf{b}_i\|_1$$

- another possibility is a combination of norms - this will attempt to decompose the state trajectory into a smooth and a piecewise constant component:

$$\underset{\mathbf{b}_1,\ldots,\mathbf{c}_m,}{\text{minimise}} \quad \sum_{i=1}^{m} \mathcal{L}(X_i(\mathbf{b}_i + \mathbf{c}_i), \mathbf{y}_i) + \lambda \sum_{i=1}^{m-1} \|\mathbf{b}_{i+1} - \mathbf{b}_i\|_1 + \mu \sum_{i=1}^{m-1} \|\mathbf{c}_{i+1} - \mathbf{c}_i\|_2^2$$

# $\ell_1$-norm vs. squared Euclidean norm regularisation



# Other modifications

- we can allow linear trends in the parameters (this formulation can be reduced to the standard state space model by expanding the state vector):

$$\underset{\mathbf{b}_1,\ldots,\mathbf{b}_m}{\text{minimise}} \quad \sum_{i=1}^{m} \mathcal{L}(X_i \mathbf{b}_i, \mathbf{y}_i) + \sum_{i=1}^{m-2} \|\mathbf{b}_{i+2} - 2\mathbf{b}_{i+1} + \mathbf{b}_i\|_1$$

- seasonality adjustments can be handled through the introduction of some equality constraints:

$$\begin{array}{ll} \underset{\mathbf{b}_1,\ldots,\mathbf{c}_m,}{\text{minimise}} & \sum_{i=1}^{m} \mathcal{L}(X_i(\mathbf{b}_i + \mathbf{c}_i), \mathbf{y}_i) + \sum_{i=1}^{m-1} \|\mathbf{b}_{i+1} - \mathbf{b}_i\|_2^2 \\ \text{subject to} & \mathbf{c}_i = \mathbf{c}_{i+k}, \quad i = 1, \ldots, m-k \\ & \sum_{i=1}^{m} \mathbf{c}_i = 0 \end{array}$$

# Conclusions

- for many modelling tasks it may be worthwhile to try to formulate a convex optimisation problem and use existing modelling software for prototyping

- formulating state space models as regularised regression can make them more intuitive for people without background in control theory (e.g. actuaries)