Parameter Estimation Lecture 15



Least Squares Exercises Recursive

► Recursive Least Squares with Forgetting Factor

Least Squares Lecture 15



A data model with minples and p parameters, one atput a set of n measurements with the cares principles and all function to minimize:

- $\mathcal{Y} \subseteq \mathbb{R}^m \times \overline{\mathbb{R}}$ $\mathcal{Y} = \{(x_{1,1}, \dots, x_{m,1}, \tilde{y}_1), \dots, (x_{1,n}, \dots, x_{m,n}, \tilde{y}_n)\}$
- $J(\beta_1,\ldots,\beta_p) = \sum_{k=1}^n \left(f(x_{1,k},\ldots,x_{m,k},\beta_1,\ldots,\beta_p) \tilde{y}_k \right)^2$ $(\beta_1^*,\ldots,\beta_p^*) = \arg\min J$

Least Squares Lecture 15



Since the mode has
$$p$$
 pall metric here. graded actions:
$$\frac{\partial J}{\partial \beta_j} = 2\sum_k \left(f(x_{1,k}, \dots, x_{-k}, \beta_1, \dots, \beta_p) \right) \frac{\partial f}{\partial \beta_j} = \quad ; \quad j = \overline{1, p}$$

The gradient equations apply to all least squares problems. Each particular problem requires particular expressions for the model function *f* and thus its partial derivatives.



► The data medel is the ar in the parameters:

$$J(\beta_1,\ldots,\beta_i) = \sum_{i=1}^n J(\beta_i)$$

▶ The minimization of the residual function *J* is equivalent to:

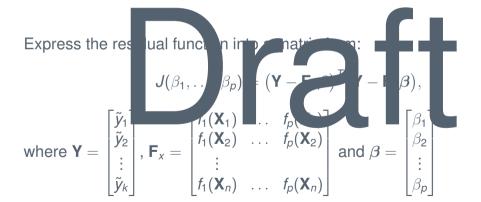
$$\frac{\partial J}{\partial \beta_j} = -2 \sum_{k=1}^n \left(\tilde{y}_k - \sum_{i=1}^p f_i(\mathbf{X}_k) \cdot \beta_i \right) f_j(\mathbf{X}_k) = 0$$



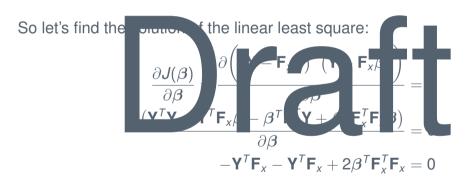
We will use:

 $J(\beta_1,\ldots,\beta_n) = \sum_{k=1}^{n} \left(\mathbf{X}_k, \beta_i \right)$









$$\beta = \left(\mathbf{F}_{x}^{T}\mathbf{F}_{x}\right)^{-1}\mathbf{F}_{x}^{T}\mathbf{Y}$$

Matrix/Multivariate Calculus Lecture 15



https://en.wikip

iki/Matrix_calculus

Identities: vector-by-vector $\frac{-b}{\partial \mathbf{x}}$								
Condition	Expression	Numerator layout, i.e. by y and x ^T	Denominator layout, i.e. by y ^T and x					
a is not a function of x	$\frac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	0						
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	ı						
A is not a function of x	$\frac{\partial \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} =$	A	\mathbf{A}^{T}					
A is not a function of x	$\frac{\partial \mathbf{x}^{T} \mathbf{A}}{\partial \mathbf{x}} =$	\mathbf{A}^{\top}	A					
a is not a function of x , $u = u(x)$	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$						
$v = v(\mathbf{x}),$ a is not a function of x	$\frac{\partial va}{\partial x} =$	$a \frac{\partial v}{\partial x}$	$\frac{\partial v}{\partial \mathbf{x}} \mathbf{a}^{\top}$					
$\vee = \vee(\mathbf{x}), \ \mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial v\mathbf{u}}{\partial \mathbf{x}} =$	$v \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u} \frac{\partial v}{\partial \mathbf{x}}$	$v \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}} \mathbf{u}^{\top}$					
A is not a function of x , u = u(x)	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$A \frac{\partial u}{\partial x}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{A}^{\top}$					
u = u(x), v = v(x)	$\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$						
u = u(x)	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$					
u = u(x)	∂f(g(u)) _	$\frac{\partial f(g)}{\partial g(u)} \frac{\partial g(u)}{\partial u}$	$\underline{\partial \mathbf{u}} \; \underline{\partial \mathbf{g}(\mathbf{u})} \; \underline{\partial \mathbf{f}(\mathbf{g})}$					



Excercise 1 Lecture 15



Linear Least Schares on 10 data:

Given the times of the last ix not and some Olyr pic witners of the 100 meters sprint, predict the time of the next short pion at Pa is 2024.

Year	S2000	A200	B2 \18	12	R201	T2020
Women	11.12	10.93	10.78	10.75	10.71	10.61
Men	9.87	09.85	9.69	9.63	9.81	9.80

Data from olympics.com

Exercise 1 Lecture 15



- ► Choose inputs, cnowe the data model, e.g.
- $(4)^m$ of n=2 pa ► There are

There are
$$\beta = 6$$
 meas remaints, $n = 2$

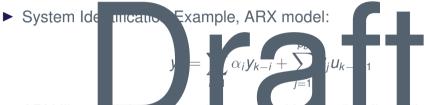
$$\begin{bmatrix} 11.12 \\ 10.95 \\ 10.75 \\ 10.71 \\ 10.61 \end{bmatrix}, \mathbf{F}_{x} = \begin{bmatrix} 1 \\ 21 \\ 22 \\ 1 \\ 23 \\ 1 \\ 24 \end{bmatrix}, \beta = \begin{bmatrix} 1 \\ \beta_{2} \\ \beta_{2} \end{bmatrix}$$

$$\Rightarrow \beta = (\mathbf{F}_{x}^{T} \mathbf{F}_{x})^{-1} \mathbf{F}_{x}^{T} \mathbf{Y}$$

$$\blacktriangleright \ \beta = \left(\mathbf{F}_{x}^{T} \mathbf{F}_{x} \right)^{-1} \mathbf{F}_{x}^{T} \mathbf{Y}$$

Exercise 2 Lecture 15





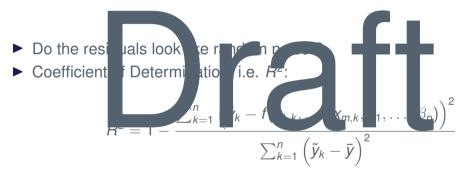
► ARX-like mannear i state (b) line parameters), e.g:

$$y_k = \beta_1 y_{k-1} y_{k-2} + \beta_2 u_k \sin(y_{k-1}) + \dots$$

► Total Least-Squares, when there are measurement errors in the "inputs" (matrix **F**_x) - Discussion

Quality of the Estimation Lecture 15





Evaluate the quality of the estimation for the two examples



Let us assume we we just calculated the solution of LS problem, from a give dataset:

where
$$\mathbf{Y} = \begin{bmatrix} y_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix}$$
, $\mathbf{F}_X = \begin{bmatrix} f_1(\mathbf{x}_1) & \vdots & f_p(\mathbf{X}_n) \\ f_1(\mathbf{x}_2) & \dots & f_p(\mathbf{X}_n) \\ \vdots & \vdots & \vdots \\ f_1(\mathbf{X}_n) & \dots & f_p(\mathbf{X}_n) \end{bmatrix}$ and $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$

► And we just received another measurement point:

$$\tilde{y}_{n+1}, f_1(\mathbf{X}_{n+1}), \dots, f_p(\mathbf{X}_{n+1}),$$



- So what is \mathbf{e} solution? $\mathbf{B} = \left(\mathbf{F}_{x}^{T}(t+1) \mathbf{x}(n+1)^{-1} \mathbf{x}(n+1)\mathbf{Y}(n-1), \right)$ where $\mathbf{Y} = \begin{bmatrix} \mathbf{I}_{x}(n) \\ \tilde{y}_{n+1} \end{bmatrix}$, $\mathbf{F}_{x} = \begin{bmatrix} \mathbf{I}_{x}(n) \\ f_{1}(\mathbf{X}_{n+1}) & \dots & f_{p}(\mathbf{X}_{n+1}) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{x}(n) \\ \mathbf{I}_{x}(n+1) \end{bmatrix}$
- ▶ But let's work on this a bit, because we do not want to do the inversion again from scratch



$$\beta = \left(\begin{bmatrix} \mathbf{F}_{x}^{T}(& \mathbf{f}(n+1) & \mathbf{f}(n+1) \end{bmatrix} \right) \begin{bmatrix} \mathbf{f}(n) & \mathbf{f}(n+1) \end{bmatrix} \begin{bmatrix} \mathbf{f}(n) & \mathbf{f}(n) & \mathbf{f}(n+1) \end{bmatrix} \begin{bmatrix} \mathbf{f}(n) & \mathbf{f}(n) & \mathbf{f}(n) \end{bmatrix} \begin{bmatrix} \mathbf{f}(n) & \mathbf{f}(n) & \mathbf{f}(n) \end{bmatrix} \begin{bmatrix} \mathbf{f}(n) & \mathbf{f}(n) & \mathbf{f}(n) & \mathbf{f}(n) \end{bmatrix} \begin{bmatrix} \mathbf{f}(n) & \mathbf{f}(n) & \mathbf{f}(n) & \mathbf{f}(n) & \mathbf{f}(n) \end{bmatrix} \begin{bmatrix} \mathbf{f}(n) & \mathbf{f}(n) &$$

► (Woodbury) Matrix Inversion Lemma:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1},$$

where **A** is $n \times n$, **C** is $k \times k$, **U** is $n \times k$ and **V** is $k \times n$ and **A** and **C** are invertible.

Recursive Least Squares



$$\mathbf{P}_{x}(n+1) \triangleq \left(\mathbf{F}_{x}(n+1)\mathbf{F}_{x}\right) + \mathbf{F}_{x}(n+1)\mathbf{F}_{x}(n+1)\mathbf{f}^{T}(n+1)\right)^{-1} = \\
= \mathbf{P}_{x}(n) - \mathbf{P}_{x}(n)\mathbf{f}(n+1)(n+1)\mathbf{f}^{T}(n+1)\mathbf{P}_{x}(n)(n+1) + \mathbf{F}_{x}(n)(n+1)\mathbf{f}^{T}(n+1)\mathbf{F}_{x}(n) = \\
= \mathbf{P}_{x}(n) - \frac{\mathbf{P}_{x}(n)\mathbf{f}(n+1)\mathbf{f}^{T}(n+1)\mathbf{P}_{x}(n)}{1 + \mathbf{f}^{T}(n+1)\mathbf{P}_{x}(n)\mathbf{f}(n+1)}$$

where $\mathbf{P}_{x}(n)$ is of course $\left(\mathbf{F}_{x}^{T}(n)\mathbf{F}_{x}(n)\right)^{-1}$.



► So,

$$\beta(n) = \mathbf{P}_{x}(n) \quad (n); \quad \mathbf{F}(n) = \left(\mathbf{F}_{x}^{T}(n)\mathbf{F}_{x}(n)\right)^{-1}$$

$$(n+1) = \mathbf{P}_{x}(n+1)\mathbf{F}^{T}(n+1) \quad (n+1), \text{ where}$$

$$\mathbf{P}_{x}(n+1) = \mathbf{P}_{x}(n) - \frac{\mathbf{P}_{x}(n+1)\mathbf{F}^{T}(n+1)\mathbf{F}^{T}(n+1)}{1+\mathbf{f}^{T}(n+1)\mathbf{P}_{x}(n)\mathbf{f}(n+1)}$$

▶ More over, we can write $\beta(n+1)$ as:

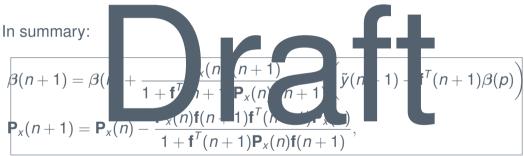
$$\beta(n+1) = \beta(n) + \frac{\mathbf{P}_{x}(n)\mathbf{f}(n+1)}{1 + \mathbf{f}^{T}(n+1)\mathbf{P}_{x}(n)\mathbf{f}(n+1)} \left(\tilde{y}(n+1) - \mathbf{f}^{T}(n+1)\beta(n)\right)$$



This form is important. It call
$$\mathbf{K} = \frac{\mathbf{P}_X(n)\mathbf{f}(n+1)}{1+\mathbf{f}^T(n+1)\mathbf{P}_X(n)\mathbf{f}(n+1)}$$
 and $\hat{y}(n+1) = \mathbf{f}^T(n+1)\beta(n)$. Then
$$\mathbf{f}(n+1) = \beta(n+1) + \mathbf{K}(\hat{y}(n+1) - \hat{y}(n+1))$$

- ▶ The equation can be though of as a correction of the $\beta(n)$ estimate
- ► K is called the correction gain
- $\left(\tilde{y}(n+1) \hat{y}(n+1)\right)$ is the innovation (or prediction error), measurement minus the prediction.





for any $n \ge 0$, with initializers $\beta(0)$ arbitrary, and $\mathbf{P}_x(0)$ arbitrary but positive definite e.g. the identity matrix \mathbf{I} . It should/will converge to good values.

RLS with Forgetting Factor



We modify the pective notion, such that we introduce a conential forgetting with a actor $\alpha \in [0,1]$. $J(\beta) = \sum_{k=1}^{n} e^{-k} \left(\tilde{y} - \sum_{i=1}^{p} |\mathbf{x}_k|_i \right)^2$

- Newer measurements have more impact on the estimation than old ones.
- ▶ Use large α for slowly changing parameters, and small α for rapidly changing parameters.

RLS with Forgetting Factor Lecture 15



The optimization of the respective function and additional the following recursive relations:

$$\beta(n+1) = \beta(1 + \frac{1}{\alpha + 1}) \left(\frac{(n+1)}{(n+1)} \left(\frac{\tilde{y}(n+1)}{(n+1)\beta(n)} \right) + \frac{1}{\alpha} \left[\mathbf{P}_{x}(n) - \frac{\mathbf{P}_{x}(n)\mathbf{f}(n+1)\mathbf{f}^{T}(n+1)\mathbf{P}_{x}(n)}{\alpha + \mathbf{f}^{T}(n+1)\mathbf{P}_{x}(n)\mathbf{f}(n+1)} \right],$$

However, a drawback that needs to be addressed

RLS with Forgetting Factor Lecture 15



- When the sestem enters steady-state, the matrix $\mathbf{P}_x(n)\mathbf{f}(n+1)\mathbf{f}^T(n+1)$ and $\mathbf{P}_x(n)\mathbf{f}(n+1)$ when the sestem enters steady-state, the matrix $\mathbf{P}_x(n)\mathbf{f}(n+1)$ and $\mathbf{P}_x(n)\mathbf{f}(n+1)$ and $\mathbf{P}_x(n)\mathbf{f}(n+1)$
- As $\alpha < 1$, $\frac{1}{\alpha}$ makes $\mathbf{P}_x(n)$ bigger than $\mathbf{P}_x(n-1)$. Although we reached steady state, $\mathbf{P}_x(n)$ begins to increase exponentially.
- Covariance Resetting: occasionally reset the matrix P(n), for example to the value $k \cdot I$, k > 0.

Well done so far!

Dal

