Statistical Signal Processing

Lecture 5: The Linear Least Squares Problem and Solution

Carlas Smith & Peyman Mohajerin Esfahani

Delft Center for Systems and Control

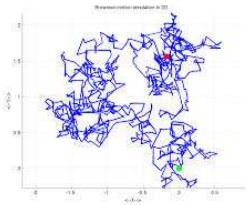
Delft University of Technology

Outline of Lecture 5

1. **Recall the Scientific Challenge 1827 (in a modern setting)**
2. Problem Formulation of estimating the parameters of an AR(p) model
3. Generalization: the Linear Least Squares (LLSQ) problem
4. Solution to LLSQ: (1) mathematically
5. Solution to LLSQ: (2) geometrically
6. Properties of the LLSQ solution

Recall the Scientific Challenge 1827

The “stochastic” nature of the movement of particles (pollen) suspended in fluid.



Paul Langevin (1872-1946)

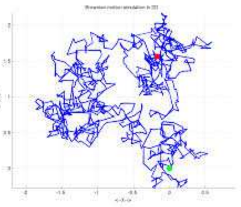
Postulated (based on Newton mechanics) a stochastic differential equation to model the motion of a single particle.

$$m\frac{d^2x(t)}{dt^2} + \gamma\frac{dx(t)}{dt} = \sqrt{2k_B T\gamma}w(t)$$

`Brownian.m`

Recall the Scientific Challenge 1827

The “stochastic” nature of the movement of particles (pollen) suspended in fluid.



Paul Langevin (1872-1946)

A discretization (see Matlab Ex-1) of the stochastic differential equation to model the motion of a single particle is

$$x(n) + \beta_1 x(n-1) + \beta_2 x(n-2) = \beta_3 w(n)$$

`Brownian.m`

The inverse problem we want to address now is to “estimate” the parameters β_i in this AR(2) model from just a single realization $x_i(n)$ (denoted in brief by $x(n)$).

Outline of Lecture 5

1. Recall the Scientific Challenge 1827 (in a modern setting)
2. **Problem Formulation of estimating the parameters of an AR(p) model**
3. Generalization: the Linear Least Squares (LLSQ) problem
4. Solution to LLSQ: (1) mathematically
5. Solution to LLSQ: (2) geometrically
6. Properties of the LLSQ solution

AR(1) Example

Consider the AR(1) model:

$$x(n) - ax(n-1) = v(n) \quad |a| < 1$$

and $v(n)$ ZMWN(σ_v^2), that has "generated" $x(n)$. With an observation (of a single realization) $x(n)$ for $n = 1 : N$, and an estimate a_N we can define the "error" signal:

$$e(n; a_N) = x(n) - a_N x(n-1)$$

AR(1) Example

Consider the AR(1) model:

$$x(n) - ax(n-1) = v(n) \quad |a| < 1$$

and $v(n)$ ZMWN(σ_v^2), that has "generated" $x(n)$. With an observation (of a single realization) $x(n)$ for $n = 1 : N$, and an estimate a_N we can define the "error" signal:

$$e(n; a_N) = x(n) - a_N x(n-1)$$

Gauss' "breakthrough" was to minimize the "sum of squared errors" w.r.t. unknown parameter(s) (a_N):

$$\sum_{n=2}^{N} |e(n; a_N)|^2 \overset{\text{Ergodicity}}{\Longrightarrow} \lim_{N \to \infty} \frac{1}{N} \sum_{n=2}^{N} |e(n; a_N)|^2 = E[|e(n; a_N)|^2]$$

AR(1) Example (Ct'd)

Gauss' Approach:

$$\min_{a_N} E[|e(n; a_N)|^2]$$

AR(1) Example (Ct'd)

Gauss' Approach:

$$\min_{a_N} E[|e(n;a_N)|^2]$$

$$\begin{aligned} e(n;a_N) &= \underbrace{x(n)}_{} - a_N x(n-1) \\ &= \\ &= \end{aligned}$$

AR(1) Example (Ct'd)

Gauss' Approach:

$$\min_{a_N} E[|e(n;a_N)|^2]$$

$$\begin{aligned} e(n;a_N) &= \underbrace{x(n)} - a_N x(n-1) \\ &= \underbrace{ax(n-1)+v(n)} - a_N x(n-1) \\ &= \end{aligned}$$

AR(1) Example (Ct'd)

Gauss' Approach:

$$\min_{a_N} E[|e(n;a_N)|^2]$$

$$\begin{aligned} e(n;a_N) &= \underbrace{x(n)} - a_N x(n-1) \\ &= \underbrace{ax(n-1) + v(n)} - a_N x(n-1) \\ &= (a - a_N)x(n-1) + v(n) \end{aligned}$$

AR(1) Example (Ct'd)

Gauss' Approach:

$$\min_{a_N} E[|e(n;a_N)|^2]$$

$$\begin{aligned} e(n;a_N) &= \underbrace{x(n)} - a_N x(n-1) \\ &= \underbrace{ax(n-1) + v(n)} - a_N x(n-1) \\ &= (a - a_N)x(n-1) + v(n) \end{aligned}$$

Then we can evaluate the variance of this error signal as:

$$E[|e(n;a_N)|^2] = |a - a_N|^2 r_x(0) + \sigma_v^2 \qquad \text{(Exercise 6.1)}$$

AR(1) Example (Ct'd)

Gauss' Approach:

$$\min_{a_N} E[|e(n;a_N)|^2]$$

$$\begin{aligned} e(n;a_N) &= \underbrace{x(n)}\, - a_N x(n-1) \\ &= \underbrace{ax(n-1) + v(n)}\, - a_N x(n-1) \\ &= (a - a_N)x(n-1) + v(n) \end{aligned}$$

Then we can evaluate the variance of this error signal as:

$$E[|e(n;a_N)|^2] = |a - a_N|^2 r_x(0) + \sigma_v^2 \qquad \text{(Exercise 6.1)}$$

What is then the answer to the question:

$$\min_{a_N} E[|e(n;a_N)|^2] = ? \quad \text{for} \quad \hat{a}_N = ?$$

Estimating the parameters of an AR(p) model

Recall the AR(p) model:

$$x(n) + a(1)x(n-1) + \cdots + a(p)x(n-p) = v(n)$$

for $v(n)$ ZMWN(σ_v^2). Then define the error signal $e(n; a_N)$:

$$x(n) + a_N(1)x(n-1) + \cdots + a_N(p)x(n-p) = e(n; a_N)$$

for the "unknown" parameter vector $a_N = \begin{bmatrix} a_N(1) & a_N(2) & \cdots & a_N(p) \end{bmatrix}^T$

Then we consider the following problem:

> Given the data $x(n)$ for $n = 0 : N-1$, then determine an estimate of the AR parameters $a_N(i)$ as follows:
>
> $$\min_{a_N} \frac{1}{N} \sum_{n=p}^{N-1} |e(n; a_N)|^2$$

Outline of Lecture 5

1. Recall the Scientific Challenge 1827 (in a modern setting)
2. Problem Formulation of estimating the parameters of an AR(p) model
3. **Generalization: the Linear Least Squares (LLSQ) problem**
4. Solution to LLSQ: (1) mathematically
5. Solution to LLSQ: (2) geometrically
6. Properties of the LLSQ solution

Generalization: the Linear Least Squares (LLSQ) problem

Recall ther AR parameter estimation problem:

$$\min_{a_N} \frac{1}{N} \sum_{n=p}^{N-1} |e(n; a_N)|^2$$

With the definition of the following quatities:

$$X_N = \begin{bmatrix} x(p-1) & x(p-2) & \cdots & x(0) \\ x(p) & x(p-1) & & x(1) \\ \vdots & & \ddots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-p-1) \end{bmatrix} \quad y_N = \begin{bmatrix} x(p) \\ x(p+1) \\ \vdots \\ x(N-1) \end{bmatrix}$$

$$a_N = \begin{bmatrix} a_N(1) & \cdots & a_N(p) \end{bmatrix}^T \quad e_N = \begin{bmatrix} e(p; a_N) & e(p+1; a_N) & \cdots & e(N-1; a_N) \end{bmatrix}^T$$

Then we can write the AR parameter estimation problem as:

$$\boxed{\min_{a_N} \frac{1}{N} \sum_{n=p}^{N-1} |e(n; a_N)|^2 = \min_{a_N} \frac{1}{N} \|e_N\|_2^2 = \min_{a_N} \frac{1}{N} \|y_N + X_N a_N\|_2^2}$$

For general matrix X_N and vectors y_N, a_N this is The Linear Least Squares Problem

Example: Estimating Resistor Value

We attempt to verify Ohms law using a voltage sensor with an unknown offset.

The voltage sensor “artifacts” can be modelled as:

$$u_m(n) = u(n) + u_0 + e(n)$$

with $u(n)$ - the real voltage, $u_0 \in \mathbb{R}$ the unknown offset

and $e(n)$ and unknown ZMWN.

Then according to Ohm’s law the data equation reads:

$$u_m(n) = Ri(n) + u_0 + e(n)$$

With the measurements $\{u_m(n), i(n)\}_{n=0}^{N-1}$ $\rightarrow$ LLSQ problem:

$$X_N = \begin{bmatrix} -i(0) & -1 \\ -i(1) & -1 \\ \vdots & \vdots \\ -i(N-1) & -1 \end{bmatrix} \quad y_N = \begin{bmatrix} u_m(0) \\ u_m(1) \\ \vdots \\ u_m(N-1) \end{bmatrix} \quad a_N = \begin{bmatrix} R_N \\ u_{0,N} \end{bmatrix}$$

Outline of Lecture 5

1. Recall the Scientific Challenge 1827 (in a modern setting)
2. Problem Formulation of estimating the parameters of an AR(p) model
3. Generalization: the Linear Least Squares (LLSQ) problem
4. **Solution to LLSQ: (1) mathematically**
5. Solution to LLSQ: (2) geometrically
6. Properties of the LLSQ solution

Recall A Note on Optimization

Example: Let $e(n,a)$ be affine in a, for example given as:

$$e(n,a) = d(n) + ax(n) \quad d(n), x(n), a \in \mathbb{C},$$

then the necessary condition for solving the following optimization problem,

$$\min_{a^*,(a)} |e(n,a)|^2 = \min_{a^*,(a)} e(n,a)e^*(n,a)$$

Recall A Note on Optimization

Example: Let $e(n,a)$ be affine in a, for example given as:

$$e(n,a) = d(n) + ax(n) \quad d(n), x(n), a \in \mathbb{C},$$

then the necessary condition for solving the following optimization problem,

$$\min_{a^*,(a)} |e(n,a)|^2 = \min_{a^*,(a)} e(n,a)e^*(n,a)$$

is given by,

$$\boxed{e(n,a)\frac{\partial e^*(n,a)}{\partial a^*} = 0} \quad \text{or} \quad \left(\frac{\partial e(n,a)}{\partial a} e^*(n,a) = 0\right)$$

Mathematical Solution of the AR problem

Consider

$$J(a_N) = \frac{1}{N} \sum_{n=p}^{N-1} |e(n; a_N)|^2 = \|X_N a_N + y_N\|_2^2$$

then the necessary and sufficient conditions for minimizing this cost function are:

$$\frac{\partial J(a_N)}{\partial a_N^*(k)} = 0 \quad \text{for} \;\; k = 1 : p$$

Mathematical Solution of the AR problem

Consider

$$J(a_N) = \frac{1}{N}\sum_{n=p}^{N-1} |e(n;a_N)|^2 = \|X_N a_N + y_N\|_2^2$$

then the necessary and sufficient conditions for minimizing this cost function are:

$$\frac{\partial J(a_N)}{\partial a_N^*(k)} = 0 \quad \text{for} \;\; k = 1 : p$$

This results into,

$$\boxed{\frac{1}{N}\sum_{n=p}^{N-1} \begin{bmatrix} x^*(n-1) \\ x^*(n-2) \\ \vdots \\ x^*(n-p) \end{bmatrix} e(n;\hat{a}_N) = 0} \Leftrightarrow$$

which is the famous Orthogonality Condition

Mathematical Solution of the AR problem

Consider

$$J(a_N) = \frac{1}{N} \sum_{n=p}^{N-1} |e(n; a_N)|^2 = \|X_N a_N + y_N\|_2^2$$

then the necessary and sufficient conditions for minimizing this cost function are:

$$\frac{\partial J(a_N)}{\partial a_N^*(k)} = 0 \quad \text{for} \quad k = 1 : p$$

This results into,

$$\frac{1}{N} \sum_{n=p}^{N-1} \begin{bmatrix} x^*(n-1) \\ x^*(n-2) \\ \vdots \\ x^*(n-p) \end{bmatrix} e(n; \hat{a}_N) = 0$$

which is the famous Orthogonality Condition

$\Leftrightarrow$

$$\left(\frac{1}{N} X_N^H X_N\right) \hat{a}_N + \frac{1}{N} X_N^H y_N = 0$$

which are the famous Normal Equations

Outline of Lecture 5

1. Recall the Scientific Challenge 1827 (in a modern setting)
2. Problem Formulation of estimating the parameters of an AR(p) model
3. Generalization: the Linear Least Squares (LLSQ) problem
4. Solution to LLSQ: (1) mathematically
5. **Solution to LLSQ: (2) geometrically**
6. Properties of the LLSQ solution

Geometrical Solution of the AR problem

Consider the AR(p) LSQ problem for $p = 2$ and $N = 5$,

$$\min_{a_5} \|y_5 - \Big(\underbrace{-a_5(1)x_{5,1} - a_5(2)x_{5,2}}\Big)\|_2^2 \qquad \text{with}$$

$$x_{5,1} = \begin{bmatrix} x(1) \\ x(2) \\ x(3) \end{bmatrix} \quad x_{5,2} = \begin{bmatrix} x(0) \\ x(1) \\ x(2) \end{bmatrix} \quad y_5 = \begin{bmatrix} x(2) \\ x(3) \\ x(4) \end{bmatrix} \quad e_5 = \begin{bmatrix} e(2;a_5) \\ e(3;a_5) \\ e(4;a_5) \end{bmatrix}$$

Optimal?

Geometrical Solution of the AR problem

Consider the AR(p) LSQ problem for $p = 2$ and $N = 5$,

$$\min_{a_5} \|y_5 - \Big(\underbrace{-a_5(1)x_{5,1} - a_5(2)x_{5,2}} \Big)\|_2^2 \qquad \text{with}$$

$$x_{5,1} = \begin{bmatrix} x(1) \\ x(2) \\ x(3) \end{bmatrix} \quad x_{5,2} = \begin{bmatrix} x(0) \\ x(1) \\ x(2) \end{bmatrix} \quad y_5 = \begin{bmatrix} x(2) \\ x(3) \\ x(4) \end{bmatrix} \quad e_5 = \begin{bmatrix} e(2; a_5) \\ e(3; a_5) \\ e(4; a_5) \end{bmatrix}$$

Optimal? $\quad e_5 \perp \text{span}\Big(\begin{bmatrix} x_{5,1} & x_{5,2} \end{bmatrix} \Big)$

Geometrical Solution of the AR problem

Consider the AR(p) LSQ problem for $p = 2$ and $N = 5$,

$$\min_{a_5} \|y_5 - \Big(\underbrace{-a_5(1)x_{5,1} - a_5(2)x_{5,2}} \Big)\|_2^2 \qquad \text{with}$$

$$x_{5,1} = \begin{bmatrix} x(1) \\ x(2) \\ x(3) \end{bmatrix} \quad x_{5,2} = \begin{bmatrix} x(0) \\ x(1) \\ x(2) \end{bmatrix} \quad y_5 = \begin{bmatrix} x(2) \\ x(3) \\ x(4) \end{bmatrix} \quad e_5 = \begin{bmatrix} e(2;a_5) \\ e(3;a_5) \\ e(4;a_5) \end{bmatrix}$$

Optimal? $\quad e_5 \perp \text{span}\Big(\begin{bmatrix} x_{5,1} & x_{5,2} \end{bmatrix} \Big)$

This is equivalent to,

$$\frac{1}{5} \begin{bmatrix} x_{5,1}^T \\ x_{5,2}^T \end{bmatrix} e_5 = 0 \quad OC$$

Outline of Lecture 5

1. Recall the Scientific Challenge 1827 (in a modern setting)
2. Problem Formulation of estimating the parameters of an AR(p) model
3. Generalization: the Linear Least Squares (LLSQ) problem
4. Solution to LLSQ: (1) mathematically
5. Solution to LLSQ: (2) geometrically
6. **Properties of the LLSQ solution**

Properties of a LLSQ estimate

Consider the LLSQ problem:

$$\min_{a_N} \frac{1}{N} \sum_{n=p}^{N-1} |e(n;a_N)|^2 = \min_{a_N} \frac{1}{N} \|e_N\|_2^2 = \min_{a_N} \frac{1}{N} \|y_N + X_n a_N\|_2^2$$

with data equation given as,

$$y_N + X_N a = v_N \quad E[v_N v_N^H] = \sigma_v^2 I_N$$

That is v_N is ZMWN(σ_v^2) and X_N is known (and hence deterministic).

Properties of a LLSQ estimate

Consider the LLSQ problem:

$$\min_{a_N} \frac{1}{N} \sum_{n=p}^{N-1} |e(n;a_N)|^2 = \min_{a_N} \frac{1}{N} \|e_N\|_2^2 = \min_{a_N} \frac{1}{N} \|y_N + X_n a_N\|_2^2$$

with data equation given as,

$$y_N + X_N a = v_N \quad E[v_N v_N^H] = \sigma_v^2 I_N$$

That is v_N is ZMWN(σ_v^2) and X_N is known (and hence deterministic).

Property 1: The LSQ solution to the normal equations

$$\left(\frac{1}{N} X_N^H X_N\right) \hat{a}_N + \frac{1}{N} X_N^H y_N = 0$$

is a random variable.

Properties of a LLSQ estimate (Ct'd)

Property 2: [(Un-)Bias of the LLSW estimate?] Let the conditions on the previous slide hold, and let the matrix $\left(\frac{1}{N} X_N^H X_N\right)$ be invertible, then,

$$E[\hat{a}_N] = a$$

Properties of a LLSQ estimate (Ct'd)

Property 2: [(Un-)Bias of the LLSW estimate?] Let the conditions on the previous slide hold, and let the matrix $\left(\frac{1}{N}X_N^H X_N\right)$ be invertible, then,

$$E[\hat{a}_N] = a$$

Property 3: [Covariance matrix of the LLSW estimate] Let the conditions on the previous slide hold, and let the matrix $\left(\frac{1}{N}X_N^H X_N\right)$ be invertible, then,

$$E[(\hat{a}_N - E[\hat{a}_N])(\hat{a}_N - E[\hat{a}_N])^H] = \sigma_v^2 \left(X_N^H X_N\right)^{-1}$$