Estimation Theory

Fredrik Rusek

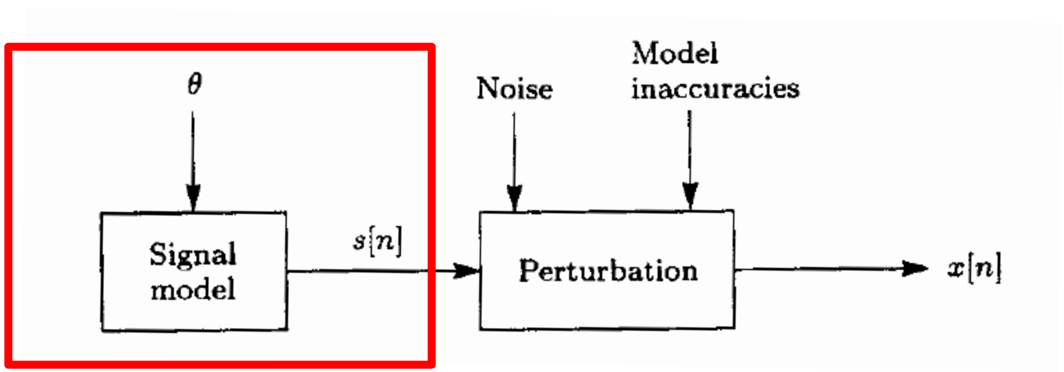
Chapters 8-9 + a bit more

Chapter 8 – Least squares

Previous chapters: Probabilistic model of observations x[n]

Least squares:

- **No probabilistic assumption**
- **Signal model needed**



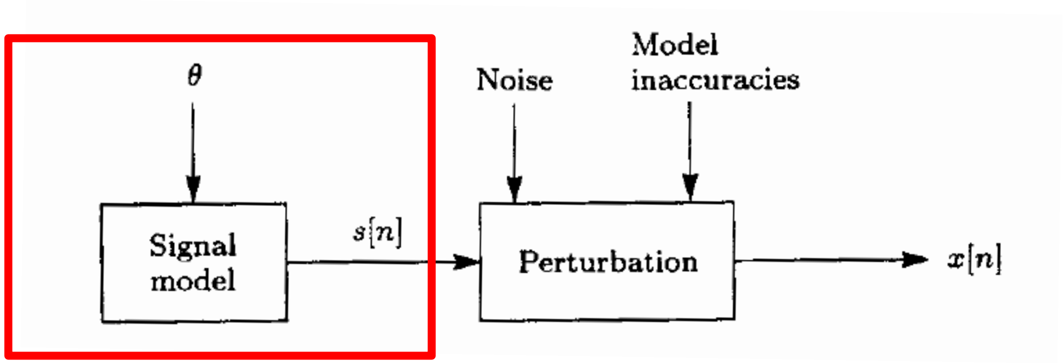
required

Not required

Chapter 8 – Least squares

Definition of least squares

Choose θ so that we minimize $$J(\theta) = \sum_{n=0}^{N-1} (x[n] - s[n])^2$$



required

Not required

Chapter 8 – Least squares

Application areas of least squares

$\boxed{\text{Observations are: deterministic signal + zero mean noise}}$

Chapter 8 – Least squares

Application areas of least squares

Observations are: deterministic signal + zero mean noise

$\boxed{\text{Linear Least squares vs non-linear least squares}}$

A linear LS problem is one where the s[n] is linear in the parameter θ to be estimated

Chapter 8 – Least squares

Application areas of least squares

Observations are: deterministic signal + zero mean noise

Linear Least squares vs non-linear least squares

A linear LS problem is one where the s[n] is linear in the parameter θ to be estimated

$$s[n] = A\cos 2\pi f_0 n$$

Estimation of:

- A is a linear LS problem
- f_0 is a non-linear estimation problem

Chapter 8 – Least squares

Application areas of least squares

Observations are: deterministic signal + zero mean noise

Linear Least squares vs non-linear least squares

A linear LS problem is one where the s[n] is linear in the parameter θ to be estimated

$$s[n] = A\cos 2\pi f_0 n$$

Estimation of:

- A is a linear LS problem
- f_0 is a non-linear estimation problem
- Both A and f_0 is linear in A, but non-linear in f_0 (More about this later)

Chapter 8 – Least squares

Linear least squares: scalar case

Signal model

$$\boxed{s[n] = \theta h[n]}$$

Optimization problem

$$\boxed{\min \quad J(\theta) = \sum_{n=0}^{N-1} (x[n] - \theta h[n])^2}$$

Chapter 8 – Least squares

Linear least squares: scalar case

Signal model

$$\boxed{s[n] = \theta h[n]}$$

Optimization problem

$$\boxed{\min \quad J(\theta) = \sum_{n=0}^{N-1} (x[n] - \theta h[n])^2}$$

Solution $\quad \hat{\theta} = \frac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]} \quad J_{\min} = J(\hat{\theta}) \quad = \sum_{n=0}^{N-1} x^2[n] - \hat{\theta} \sum_{n=0}^{N-1} x[n]h[n]$

Chapter 8 – Least squares

Linear least squares: scalar case

Signal model

$$\boxed{s[n] = \theta h[n]}$$

Optimization problem

$$\boxed{\min \quad J(\theta) = \sum_{n=0}^{N-1} (x[n] - \theta h[n])^2}$$

Clearly:

$$\boxed{0 \leq J_{\min} \leq \sum_{n=0}^{N-1} x^2[n]}$$

Solution $\hat{\theta} = \dfrac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]}$ $\quad J_{\min} = J(\hat{\theta}) = \sum_{n=0}^{N-1} x^2[n] - \hat{\theta} \sum_{n=0}^{N-1} x[n]h[n]$

Chapter 8 – Least squares

Linear least squares: scalar case

Signal model

$$\boxed{s[n] = \theta h[n]}$$

Optimization problem

$$\boxed{\min \quad J(\theta) = \sum_{n=0}^{N-1} (x[n] - \theta h[n])^2}$$

Solution $\hat{\theta} = \frac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]}$ $\quad J_{\min} = J(\hat{\theta}) \quad = \sum_{n=0}^{N-1} x^2[n] - \hat{\theta} \sum_{n=0}^{N-1} x[n]h[n]$

Lots of noise

No noise

Clearly:

$$\boxed{0 \leq J_{\min} \leq \sum_{n=0}^{N-1} x^2[n]}$$

Chapter 8 – Least squares

Linear least squares: multivariate case

Signal model

$$\boxed{\mathbf{s} = \mathbf{H}\boldsymbol{\theta}}$$

Optimization problem

$$\boxed{\min \quad J(\boldsymbol{\theta}) \quad = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})}$$

Chapter 8 – Least squares

Linear least squares: multivariate case

Signal model

$$\boxed{\mathbf{s} = \mathbf{H}\boldsymbol{\theta}}$$

Optimization problem

$$\boxed{\min \quad J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})}$$

Gradient $\dfrac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$

Solution $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$

Chapter 8 – Least squares

Linear least squares: multivariate case

Signal model

$$\boxed{\mathbf{s} = \mathbf{H}\boldsymbol{\theta}}$$

Optimization problem

$$\boxed{\min \quad J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})}$$

When is the LS
The BLUE?
Efficient?

Gradient $\quad \dfrac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$

$$\boxed{\text{Solution } \hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}}$$

Chapter 8 – Least squares

Linear least squares: multivariate case

Signal model

$$\boxed{\mathbf{s} = \mathbf{H}\boldsymbol{\theta}}$$

Optimization problem

$$\boxed{\min \quad J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})}$$

When is the LS
The BLUE? Noise-cov is identity
Efficient? Noise is Gaussian

Gradient $$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

Solution $$\boxed{\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}}$$

Chapter 8 – Least squares

Linear least squares: multivariate case

Minimum value

$$\begin{aligned} J_{\min} &= J(\hat{\boldsymbol{\theta}}) \\ &= (\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}})^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}) \\ &= \left(\mathbf{x}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}\right)^T\left(\mathbf{x}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}\right) \\ &= \mathbf{x}^T\left(\mathbf{I}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right)\left(\mathbf{I}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right)\mathbf{x} \\ &\boxed{= \mathbf{x}^T\left(\mathbf{I}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right)\mathbf{x}.} \end{aligned}$$

Optimization problem

$$\boxed{\min \quad J(\boldsymbol{\theta}) = (\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})}$$

When is the LS
The BLUE? **Noise-cov is identity**
Efficient? **Noise is Gaussian**

Gradient $\dfrac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$

$\boxed{\text{Solution } \hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}}$

Chapter 8 – Least squares

Weighted Linear least squares: multivariate case

Optimization problem

$$\boxed{\min \quad J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{W} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})}$$

solution

$$\boxed{\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{x}}$$

Is the BLUE if **W** is taken as the noise covariance matrix

Chapter 8 – Least squares

Summary of linear LS

- **Needs no probabilistic knowledge**
- **Needs observations that are linear in parameter + zero mean noise**
- **Is the BLUE if noise is uncorrelated**
- **If efficient if noise is white Guassian**
- **Weighted linear LS estimators are BLUE if the weights are properly selected**

Almost nothing new

Chapter 8 – Least squares

Tikhonov regularization

- If the matrix H is badly conditioned, the estimator tends to ”blow up”

- To penalize large values of θ (x below) Tikhonov regularization is commonly applied

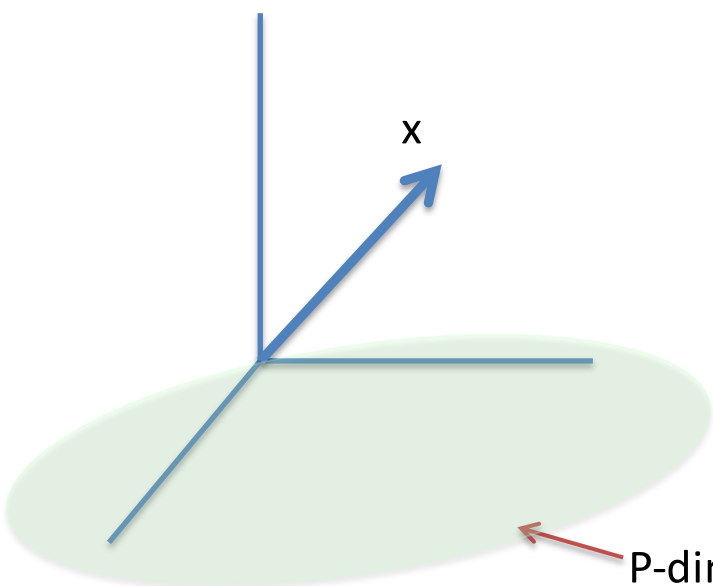
$$\|A\mathbf{x} - \mathbf{b}\|^2 + \|\Gamma\mathbf{x}\|^2$$

Chapter 8 – Least squares

Geometrical interpretation

H is (N x p) matrix

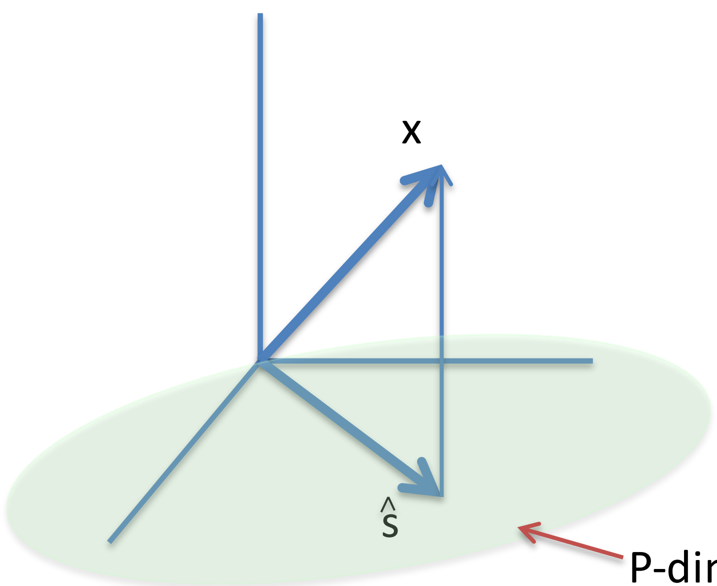
How to best approximate the N-dim observation x in the p-dimensional space



P-dimensional Subspace spanned by H
X (observations) lie in an N-dimensional space

Chapter 8 – Least squares

Geometrical interpretation



How to best approximate the N-dim observation x in the p-dimensional space

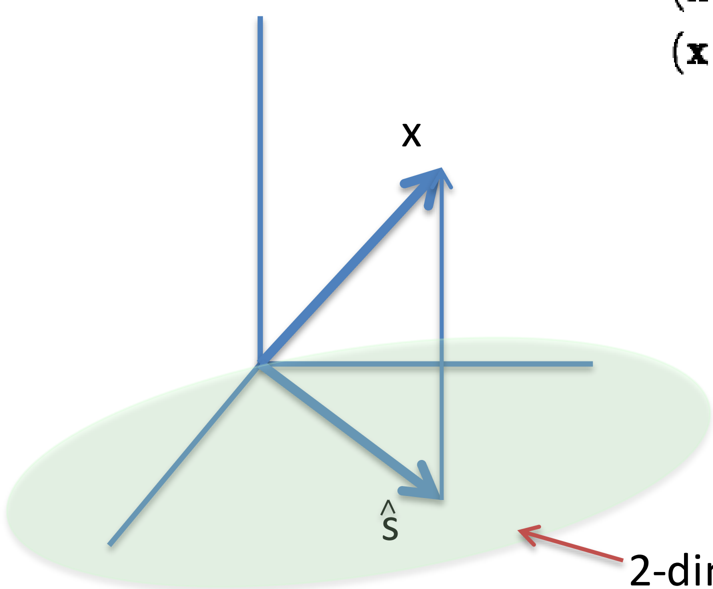
P-dimensional Subspace spanned by H
X (observations) lie in an N-dimensional space

Chapter 8 – Least squares

Geometrical interpretation

$$
\begin{aligned}
(\mathbf{x} - \hat{\mathbf{s}}) &\perp \mathbf{h}_1 \\
(\mathbf{x} - \hat{\mathbf{s}}) &\perp \mathbf{h}_2
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{x} - \hat{\mathbf{s}})^T \mathbf{h}_1 &= 0 \\
(\mathbf{x} - \hat{\mathbf{s}})^T \mathbf{h}_2 &= 0
\end{aligned}
$$



2-dimensional Subspace spanned by H
X (observations) lie in an 3-dimensional space

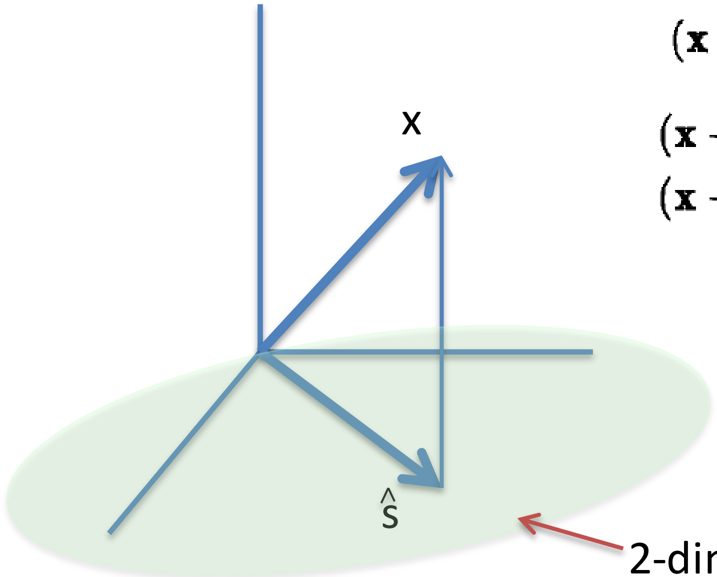
Chapter 8 – Least squares

Geometrical interpretation

$$
\begin{aligned}
(\mathbf{x}-\hat{\mathbf{s}}) &\perp \mathbf{h}_1 \\
(\mathbf{x}-\hat{\mathbf{s}}) &\perp \mathbf{h}_2
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{x}-\hat{\mathbf{s}})^T\mathbf{h}_1 &= 0 \\
(\mathbf{x}-\hat{\mathbf{s}})^T\mathbf{h}_2 &= 0
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{h}_1 &= 0 \\
(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{h}_2 &= 0
\end{aligned}
$$



$$\hat{\mathbf{s}} = \theta_1\mathbf{h}_1 + \theta_2\mathbf{h}_2$$

2-dimensional Subspace spanned by H

X (observations) lie in an 3-dimensional space

Chapter 8 – Least squares

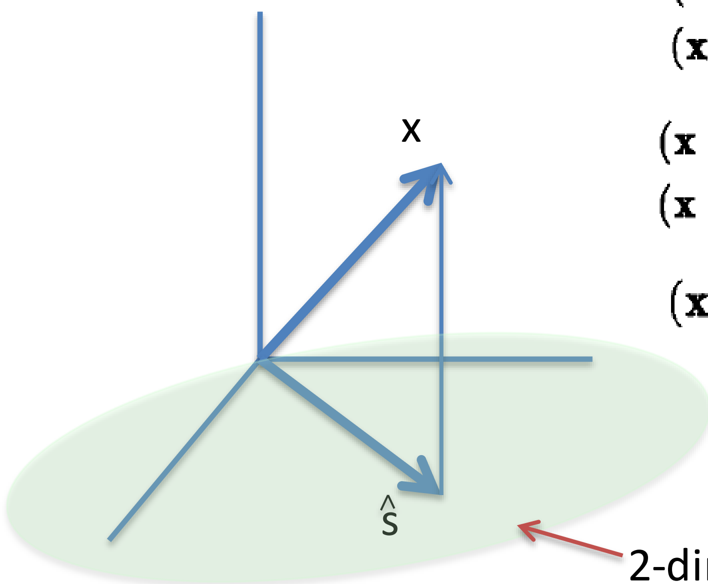
Geometrical interpretation

$$(\mathbf{x}-\hat{\mathbf{s}}) \perp \mathbf{h}_1$$
$$(\mathbf{x}-\hat{\mathbf{s}}) \perp \mathbf{h}_2$$

$$(\mathbf{x}-\hat{\mathbf{s}})^T\mathbf{h}_1 = 0$$
$$(\mathbf{x}-\hat{\mathbf{s}})^T\mathbf{h}_2 = 0$$

$$(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{h}_1 = 0$$
$$(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{h}_2 = 0$$

$$(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{H} = \mathbf{0}^T$$



$$\hat{\mathbf{s}} = \theta_1\mathbf{h}_1 + \theta_2\mathbf{h}_2$$

2-dimensional Subspace spanned by H

X (observations) lie in an 3-dimensional space

Chapter 8 – Least squares

Geometrical interpretation

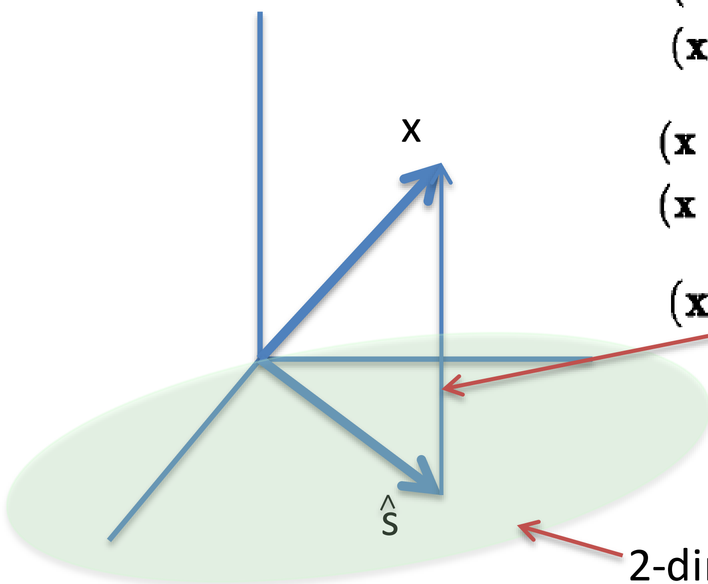
$$
\begin{aligned}
(\mathbf{x}-\hat{\mathbf{s}}) &\perp \mathbf{h}_1 \\
(\mathbf{x}-\hat{\mathbf{s}}) &\perp \mathbf{h}_2
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{x}-\hat{\mathbf{s}})^T\mathbf{h}_1 &= 0 \\
(\mathbf{x}-\hat{\mathbf{s}})^T\mathbf{h}_2 &= 0
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{h}_1 &= 0 \\
(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{h}_2 &= 0
\end{aligned}
$$

$$(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{H} = \mathbf{0}^T$$

$$\boldsymbol{\epsilon}^T\mathbf{H} = \mathbf{0}^T$$



$$\hat{\mathbf{s}} = \theta_1\mathbf{h}_1 + \theta_2\mathbf{h}_2$$

2-dimensional Subspace spanned by H

X (observations) lie in an 3-dimensional space

Orthogonality principle

Chapter 8 – Least squares

Geometrical interpretation

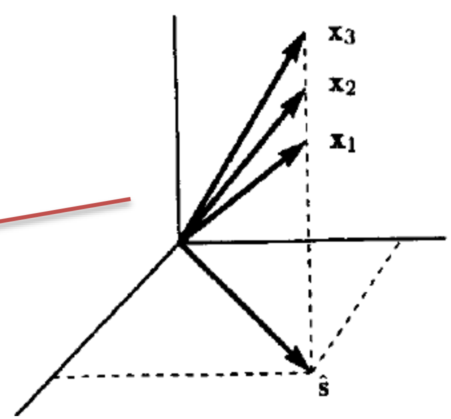
Estimator $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$

Projected signal $\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$

Projection matrix $\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$
(projects a signal onto subspace spanned by H)

Properties:

1. $P^T=P$
2. $P^2=P$
3. P is singular



Chapter 8 – Least squares

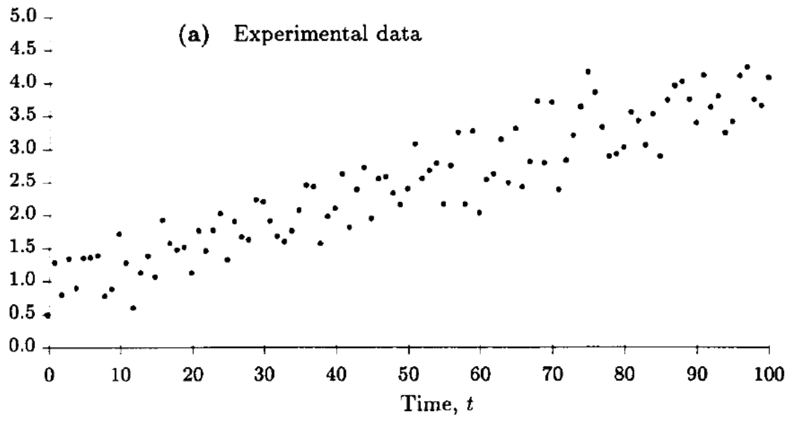
Section 8.6 and 8.7: good to know about, but not much more

Chapter 8 – Least squares

Section 8.6 and 8.7: good to know about, but not much more

Section 8.6

Fit this data to some Polynomial



Should it be 1st, 2nd, 3d, or even higher order??

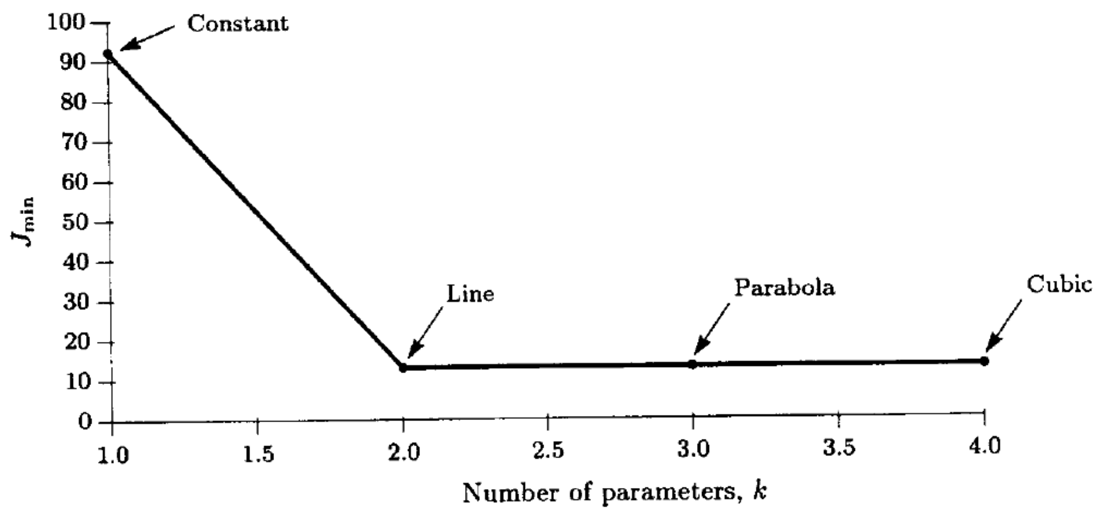
Chapter 8 – Least squares

Section 8.6 and 8.7: good to know about, but not much more

Section 8.6

Result for some

Polynomial orders



There is a recursive method that finds the solution for order n given that n-1 is already done

Hence, one can reuse the computations…not much more in this section

Chapter 8 – Least squares

Section 8.6 and 8.7: good to know about, but not much more

Section 8.7

Given the LS estimate for N-1 observations, how to find the LS estimate with one additional observation?

A toolbox for this exist. This is in Section 8.7….not much more…

Scalar case (DC level in white noise)

$$\hat{A}[N] = \hat{A}[N-1] + \frac{1}{N+1}\left(x[N] - \hat{A}[N-1]\right)$$

$$J_{\min}[N] = J_{\min}[N-1] + \frac{N}{N+1}(x[N] - \hat{A}[N-1])^2$$

Chapter 8 – Least squares

Section 8.8: constrained least squares

Suppose that the parameter vector is constrained to r<p constraints (A is rxp matrix, rank=r)

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$$

Optimization problem

min $\quad J(\boldsymbol{\theta}) \quad = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$

such that $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$

Chapter 8 – Least squares

Section 8.8: constrained least squares

min $J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$

such that $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$

Lagrangian $$J_c = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) + \boldsymbol{\lambda}^T(\mathbf{A}\boldsymbol{\theta} - \mathbf{b})$$

Chapter 8 – Least squares

Section 8.8: constrained least squares

$$\min \quad J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

$$\text{such that} \quad \mathbf{A}\boldsymbol{\theta} = \mathbf{b}$$

Lagrangian

$$\boxed{J_c = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) + \boldsymbol{\lambda}^T(\mathbf{A}\boldsymbol{\theta} - \mathbf{b})}$$

$$\frac{\partial J_c}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} + \mathbf{A}^T\boldsymbol{\lambda}$$

Chapter 8 – Least squares

Section 8.8: constrained least squares

min $\quad J(\boldsymbol{\theta}) \quad = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$

such that $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$

Lagrangian $\quad J_c = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) + \boldsymbol{\lambda}^T(\mathbf{A}\boldsymbol{\theta} - \mathbf{b})$

$$\frac{\partial J_c}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} + \mathbf{A}^T\boldsymbol{\lambda}$$

$$\begin{aligned}\hat{\boldsymbol{\theta}}_c &= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} - \frac{1}{2}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\boldsymbol{\lambda} \\ &= \hat{\boldsymbol{\theta}} - (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\frac{\boldsymbol{\lambda}}{2}\end{aligned}$$

Chapter 8 – Least squares

Section 8.8: constrained least squares

$$\boxed{\begin{aligned} &\text{min} \quad J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \\ &\text{such that} \quad \mathbf{A}\boldsymbol{\theta} = \mathbf{b} \end{aligned}}$$

Lagrangian $$\boxed{J_c = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) + \boldsymbol{\lambda}^T(\mathbf{A}\boldsymbol{\theta} - \mathbf{b})}$$

$$\frac{\partial J_c}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} + \mathbf{A}^T\boldsymbol{\lambda}$$

$$\boxed{\begin{aligned} \hat{\boldsymbol{\theta}}_c &= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} - \frac{1}{2}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\boldsymbol{\lambda} \\ &= \hat{\boldsymbol{\theta}} - (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\frac{\boldsymbol{\lambda}}{2} \end{aligned}}$$

$$\mathbf{A}\boldsymbol{\theta}_c = \mathbf{A}\hat{\boldsymbol{\theta}} - \mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\frac{\boldsymbol{\lambda}}{2} = \mathbf{b}$$

$$\frac{\boldsymbol{\lambda}}{2} = \left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}} - \mathbf{b}).$$

Chapter 8 – Least squares

Section 8.8: constrained least squares

min $\quad J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$

such that $\quad \mathbf{A}\boldsymbol{\theta} = \mathbf{b}$

Lagrangian $\quad J_c = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) + \boldsymbol{\lambda}^T(\mathbf{A}\boldsymbol{\theta} - \mathbf{b})$

$$\frac{\partial J_c}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} + \mathbf{A}^T\boldsymbol{\lambda}$$

$$\begin{aligned} \hat{\boldsymbol{\theta}}_c &= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} - \frac{1}{2}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\boldsymbol{\lambda} \\ &= \hat{\boldsymbol{\theta}} - (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}} - \mathbf{b}) \end{aligned}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

If the signal is not linear in θ a non-quadratic optimization results, termed non-linear LS

$$\boxed{J = (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))^T (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))}$$

Check-point: When is the non-linear LS also the MLE?

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Apart from numerical search, two methods are presented

1. **Transformation of parameters**

We try to find a fiunction $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$ such that

$$\boxed{\mathbf{s}(\boldsymbol{\theta}(\boldsymbol{\alpha})) = \mathbf{H}\boldsymbol{\alpha}}$$

Solve the linear LS problem $\hat{\boldsymbol{\alpha}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$

and plug back $\hat{\boldsymbol{\theta}} = \mathbf{g}^{-1}(\hat{\boldsymbol{\alpha}})$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Example 8.9

Find A and ϕ, f_0 is known

$$\boxed{s[n] = A\cos(2\pi f_0 n + \phi) \qquad n = 0, 1, \ldots, N-1}$$

Problem is non-linear in the phase $J = \sum_{n=0}^{N-1} \left(x[n] - A\cos(2\pi f_0 n + \phi)\right)^2$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Example 8.9

Find A and ϕ , f_0 is known

$$\boxed{s[n] = A\cos(2\pi f_0 n + \phi) \qquad n = 0, 1, \ldots, N-1}$$

Problem is non-linear in the phase $J = \sum_{n=0}^{N-1} (x[n] - A\cos(2\pi f_0 n + \phi))^2$

but $A\cos(2\pi f_0 n + \phi) = A\cos\phi\cos 2\pi f_0 n - A\sin\phi\sin 2\pi f_0 n$

$$\boxed{\begin{aligned} \alpha_1 &= A\cos\phi \\ \alpha_2 &= -A\sin\phi \end{aligned} \qquad s[n] = \alpha_1\cos 2\pi f_0 n + \alpha_2 \sin 2\pi f_0 n}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Example 8.9

Find A and ϕ , f_0 is known

$$\boxed{s[n] = A\cos(2\pi f_0 n + \phi) \qquad n = 0, 1, \ldots, N-1}$$

Problem is non-linear in the phase $J = \sum_{n=0}^{N-1} (x[n] - A\cos(2\pi f_0 n + \phi))^2$

but $A\cos(2\pi f_0 n + \phi) = A\cos\phi\cos 2\pi f_0 n - A\sin\phi\sin 2\pi f_0 n$

$$\boxed{\begin{aligned}\alpha_1 &= A\cos\phi \\ \alpha_2 &= -A\sin\phi\end{aligned} \qquad s[n] = \alpha_1\cos 2\pi f_0 n + \alpha_2 \sin 2\pi f_0 n}$$

Signal is linear in α_1 α_2

$$\mathbf{s} = \mathbf{H}\boldsymbol{\alpha}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \cos 2\pi f_0(N-1) & \sin 2\pi f_0(N-1) \end{bmatrix}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Example 8.9

Find A and ϕ , f_0 is known

$$\boxed{s[n] = A\cos(2\pi f_0 n + \phi) \qquad n = 0, 1, \ldots, N-1}$$

Problem is non-linear in the phase $J = \sum_{n=0}^{N-1} \left(x[n] - A\cos(2\pi f_0 n + \phi)\right)^2$

but $A\cos(2\pi f_0 n + \phi) = A\cos\phi\cos 2\pi f_0 n - A\sin\phi\sin 2\pi f_0 n$

$$\boxed{\begin{aligned} \alpha_1 &= A\cos\phi \\ \alpha_2 &= -A\sin\phi \end{aligned} \qquad s[n] = \alpha_1\cos 2\pi f_0 n + \alpha_2 \sin 2\pi f_0 n}$$

Estimator is $\hat{\boldsymbol{\alpha}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$

$$\boxed{\hat{\boldsymbol{\theta}} = \begin{bmatrix} \sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2} \\ \arctan\left(\frac{-\hat{\alpha}_2}{\hat{\alpha}_1}\right) \end{bmatrix}}$$

Inverse mapping is

$$\begin{aligned} A &= \sqrt{\alpha_1^2 + \alpha_2^2} \\ \phi &= \arctan\left(\frac{-\alpha_2}{\alpha_1}\right) \end{aligned}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Apart from numerical search, two methods are presented

1. Transformation of parameters
2. **Separable problems (half-linear, half non-linear)**

With a signal model according to

Matrix, non-linear in α

$$\boxed{\mathbf{s} = \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta}}$$

We get a linear problem in β, but non-linear in α.

$$\boxed{J(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta})^T (\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta})}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Apart from numerical search, two methods are presented

1. Transformation of parameters
2. **Separable problems (half-linear, half non-linear)**

With a signal model according to

Matrix, non-linear in α

$$\boxed{\mathbf{s} = \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta}}$$

We get a linear problem in β, but non-linear in α.

$$\boxed{J(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta})^T (\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta})}$$

For given H(α), optimal β (in a LS sense) is

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha})\right)^{-1} \mathbf{H}^T(\boldsymbol{\alpha})\mathbf{x}}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Apart from numerical search, two methods are presented

1. Transformation of parameters
2. **Separable problems (half-linear, half non-linear)**

With a signal model according to

Matrix, non-linear in α

$$\boxed{\mathbf{s} = \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta}}$$

We get a linear problem in β, but non-linear in α.

$$\boxed{J(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}) = \mathbf{x}^T \left[\mathbf{I} - \mathbf{H}(\boldsymbol{\alpha}) \left(\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha})\right)^{-1} \mathbf{H}^T(\boldsymbol{\alpha})\right] \mathbf{x}}$$

For given H(α), optimal β (in a LS sense) is

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha})\right)^{-1} \mathbf{H}^T(\boldsymbol{\alpha})\mathbf{x}}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Apart from numerical search, two methods are presented

1. Transformation of parameters
2. **Separable problems (half-linear, half non-linear)**

With a signal model according to

Matrix, non-linear in α

$$\boxed{\mathbf{s} = \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta}}$$

We get a linear problem in β, but non-linear in α.

$$\boxed{J(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}) = \mathbf{x}^T \left[\mathbf{I} - \mathbf{H}(\boldsymbol{\alpha}) \left(\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha})\right)^{-1} \mathbf{H}^T(\boldsymbol{\alpha})\right] \mathbf{x}}$$

For given H(α), optimal β (in a LS sense) is

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha})\right)^{-1} \mathbf{H}^T(\boldsymbol{\alpha})\mathbf{x}}$$

$$\boxed{max \quad \mathbf{x}^T\mathbf{H}(\boldsymbol{\alpha}) \left(\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha})\right)^{-1} \mathbf{H}^T(\boldsymbol{\alpha})\mathbf{x}}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Example 8.10

Estimate $\{A_1, A_2, A_3, r\}$

$$\boxed{s[n] = A_1 r^n + A_2 r^{2n} + A_3 r^{3n}}$$

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Example 8.10

Estimate $\{A_1, A_2, A_3, r\}$

$$\boxed{s[n] = A_1 r^n + A_2 r^{2n} + A_3 r^{3n}}$$

Signal is linear in A_1, A_2, A_3, non-linear in r

Chapter 8 – Least squares

Section 8.9: non-linear Least squares

Example 8.10

Estimate $\{A_1, A_2, A_3, r\}$

$$\boxed{s[n] = A_1 r^n + A_2 r^{2n} + A_3 r^{3n}}$$

Signal is linear in A_1, A_2, A_3, non-linear in r

$$\mathbf{H}(r) = \begin{bmatrix} 1 & 1 & 1 \\ r & r^2 & r^3 \\ \vdots & \vdots & \vdots \\ r^{N-1} & r^{2(N-1)} & r^{3(N-1)} \end{bmatrix}$$

LSE for the amplitudes

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{H}^T(\hat{r})\mathbf{H}(\hat{r})\right)^{-1}\mathbf{H}^T(\hat{r})\mathbf{x}$$

LSE for r

$$\max \quad \mathbf{x}^T\mathbf{H}(r)\left(\mathbf{H}^T(r)\mathbf{H}(r)\right)^{-1}\mathbf{H}^T(r)\mathbf{x}$$

Chapter 9 – method of moments

Consider a Gaussian mixture pdf

$$p(x[n];\epsilon) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_1^2}\right) + \frac{\epsilon}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_2^2}\right)$$

$$p(x[n];\epsilon) = (1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n])$$

$$\phi_i(x[n]) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_i^2}\right).$$

$\mathcal{N}(0,\sigma_1^2)$ PDF with probability $1-\epsilon$ and from a $\mathcal{N}(0,\sigma_2^2)$ PDF with probability ϵ

Chapter 9 – method of moments

Consider a Gaussian mixture pdf

$$p(x[n];\epsilon) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_1^2}\right) + \frac{\epsilon}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_2^2}\right)$$

$$p(x[n];\epsilon) = (1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n])$$

$$\phi_i(x[n]) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_i^2}\right).$$

$\mathcal{N}(0,\sigma_1^2)$ PDF with probability $1-\epsilon$ and from a $\mathcal{N}(0,\sigma_2^2)$ PDF with probability ϵ.

Our methods to find MVU estimator for ε will all fail

Options:

1. MLE
2. Method of moments (much simpler)

Chapter 9 – method of moments

Consider a Gaussian mixture pdf

$$p(x[n];\epsilon) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_1^2}\right) + \frac{\epsilon}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_2^2}\right)$$

$$p(x[n];\epsilon) = (1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n])$$

$$\phi_i(x[n]) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_i^2}\right).$$

Second moment

$$\begin{aligned} E(x^2[n]) &= \int_{-\infty}^{\infty} x^2[n]\left[(1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n])\right] dx[n] \\ &= (1-\epsilon)\sigma_1^2 + \epsilon\sigma_2^2 \end{aligned}$$

Chapter 9 – method of moments

Consider a Gaussian mixture pdf

$$p(x[n];\epsilon) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_1^2}\right) + \frac{\epsilon}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_2^2}\right)$$

$$p(x[n];\epsilon) = (1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n])$$

$$\phi_i(x[n]) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_i^2}\right).$$

Second moment. Replace with its estimator

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] = (1-\epsilon)\sigma_1^2 + \epsilon\sigma_2^2$$

$$\hat{\epsilon} = \frac{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - \sigma_1^2}{\sigma_2^2 - \sigma_1^2}$$

Not optimal, not unbiased in general, but simple to find

Chapter 9 – method of moments

Consider a Gaussian mixture pdf

$$p(x[n];\epsilon) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_1^2}\right) + \frac{\epsilon}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_2^2}\right)$$

$$p(x[n];\epsilon) = (1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n])$$

$$\phi_i(x[n]) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\frac{x^2[n]}{\sigma_i^2}\right).$$

Second moment. Replace with its estimator

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] = (1-\epsilon)\sigma_1^2 + \epsilon\sigma_2^2$$

$$\hat{\epsilon} = \frac{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - \sigma_1^2}{\sigma_2^2 - \sigma_1^2}$$

As $N \to \infty$, $\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \to E(x^2[n])$ so $\hat{\epsilon} \to \epsilon$ as $N \to \infty$.

Chapter 9 – method of moments

Method of moments, scalar case

kth ***moment***

$\mu_k = E(x^k[n]) = h(\theta)$

Chapter 9 – method of moments

Method of moments, scalar case

***k*th moment**

$\mu_k = E(x^k[n]) = h(\theta)$

Find inverse function

$\theta = h^{-1}(\mu_k)$

Chapter 9 – method of moments

Method of moments, scalar case

kth ***moment***

$$\mu_k = E(x^k[n]) = h(\theta)$$

Find inverse function

$$\theta = h^{-1}(\mu_k)$$

Replace moment with estimate

$$\hat{\mu}_k = \frac{1}{N}\sum_{n=0}^{N-1} x^k[n]$$

Chapter 9 – method of moments

Method of moments, scalar case

***k*th moment**

$$\mu_k = E(x^k[n]) = h(\theta)$$

Find inverse function

$$\theta = h^{-1}(\mu_k)$$

Replace moment with estimate

$$\hat{\mu}_k = \frac{1}{N}\sum_{n=0}^{N-1} x^k[n]$$

conclude

$$\hat{\theta} = h^{-1}\left(\frac{1}{N}\sum_{n=0}^{N-1} x^k[n]\right)$$

Chapter 9 – method of moments

Example 9.2: exponential pdf

Exponential pdf $p(x[n];\lambda) = \begin{cases} \lambda \exp(-\lambda x[n]) & x[n] > 0 \\ 0 & x[n] < 0. \end{cases}$

$$\mu_1 = E(x[n]) = \frac{1}{\lambda}$$

Chapter 9 – method of moments

Example 9.2: exponential pdf

Exponential pdf $p(x[n];\lambda) = \begin{cases} \lambda \exp(-\lambda x[n]) & x[n] > 0 \\ 0 & x[n] < 0. \end{cases}$

$$\mu_1 = E(x[n]) = \frac{1}{\lambda}$$

$$\hat{\lambda} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}$$

According to discussion yesterday, this is not MVU

Chapter 9 – method of moments

Method of moments, vector parameter

Find p moments

$$\begin{aligned}\mu_1 &= h_1(\theta_1, \theta_2, \ldots, \theta_p) \\ \mu_2 &= h_2(\theta_1, \theta_2, \ldots, \theta_p) \\ \vdots \;\; & \;\; \vdots \quad\;\; \vdots \\ \mu_p &= h_p(\theta_1, \theta_2, \ldots, \theta_p)\end{aligned}$$

$$\boldsymbol{\mu} = \mathbf{h}(\boldsymbol{\theta})$$

find inverse

$$\boldsymbol{\theta} = \mathbf{h}^{-1}(\boldsymbol{\mu})$$

conclude

$$\hat{\boldsymbol{\theta}} = \mathbf{h}^{-1}(\hat{\boldsymbol{\mu}})$$

Chapter 9 – method of moments

Method of moments, Performance

Definition of method of moments $\hat{\theta} = g(\mathbf{T})$ ***where*** $E(\mathbf{T}) = \boldsymbol{\mu}$

Taylor expand around $\boldsymbol{\mu}$

$$\boxed{\hat{\theta} = g(\mathbf{T}) \approx g(\boldsymbol{\mu}) + \sum_{k=1}^{r} \left.\frac{\partial g}{\partial T_k}\right|_{\mathbf{T}=\boldsymbol{\mu}} (T_k - \mu_k)}$$

Chapter 9 – method of moments

Method of moments, Performance

Definition of method of moments $\hat{\theta} = g(\mathbf{T})$ ***where*** $E(\mathbf{T}) = \boldsymbol{\mu}$

Taylor expand around $\boldsymbol{\mu}$

$$\boxed{\hat{\theta} = g(\mathbf{T}) \approx g(\boldsymbol{\mu}) + \sum_{k=1}^{r} \left.\frac{\partial g}{\partial T_k}\right|_{\mathbf{T}=\boldsymbol{\mu}} (T_k - \mu_k)}$$

To the precision of the linearization $E(\hat{\theta}) = g(\boldsymbol{\mu})$

$$\operatorname{var}(\hat{\theta}) = E\left\{\left[g(\boldsymbol{\mu}) + \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} (\mathbf{T} - \boldsymbol{\mu}) - E(\hat{\theta})\right]^2\right\} = E\left\{\left[g(\boldsymbol{\mu}) + \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} (\mathbf{T} - \boldsymbol{\mu}) - g(\boldsymbol{\mu})\right]^2\right\}$$

$$= E\left\{\left[\left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} (\mathbf{T} - \boldsymbol{\mu})\right]^2\right\} = \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} \mathbf{C}_T \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}$$

Chapter 9 – method of moments

Method of moments, exponential distribution continued

Recall estimator

$$\hat{\lambda} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}$$

$$\operatorname{var}(\hat{\theta}) = \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} \mathbf{C}_T \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}} \qquad E(\hat{\theta}) = g(\boldsymbol{\mu})$$

Chapter 9 – method of moments

Method of moments, exponential distribution continued

Recall estimator

$$\hat{\lambda} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}$$

Identify

$$\hat{\lambda} = g(T_1)$$

$$T_1 = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad g(T_1) = \frac{1}{T_1}$$

$$\operatorname{var}(\hat{\theta}) = \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} \mathbf{C}_T \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}} \qquad E(\hat{\theta}) = g(\boldsymbol{\mu})$$

Chapter 9 – method of moments

Method of moments, exponential distribution continued

Recall estimator

$$\hat{\lambda} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}$$

Identify $\hat{\lambda} = g(T_1)$

We should do T.S.E at mean of T

$$T_1 = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad g(T_1) = \frac{1}{T_1}$$

$$\mu_1 = E(T_1) = \frac{1}{\lambda}$$

$$\operatorname{var}(\hat{\theta}) = \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} \mathbf{C}_T \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}} \qquad E(\hat{\theta}) = g(\boldsymbol{\mu})$$

Chapter 9 – method of moments

Method of moments, exponential distribution continued

Recall estimator $$\hat{\lambda} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}$$

Identify $\hat{\lambda} = g(T_1)$

We should do T.S.E at mean of T

$$T_1 = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad g(T_1) = \frac{1}{T_1}$$

$$\mu_1 = E(T_1) = \frac{1}{\lambda}$$

Now, to the precision of the linear approximation

$$E(\hat{\lambda}) = g(\mu_1) = \frac{1}{\frac{1}{\lambda}} = \lambda$$

$$\boxed{\operatorname{var}(\hat{\theta}) = \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} \mathbf{C}_T \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}} \qquad E(\hat{\theta}) = g(\boldsymbol{\mu})}$$

Chapter 9 – method of moments

Method of moments, exponential distribution continued

Recall estimator

$$\hat{\lambda} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}$$

Identify $\hat{\lambda} = g(T_1)$

$$T_1 = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad g(T_1) = \frac{1}{T_1}$$

We should do T.S.E at mean of T

$$\mu_1 = E(T_1) = \frac{1}{\lambda}$$

Now, to the precision of the linear approximation

$$\mathrm{var}(T_1) = \frac{1}{N\lambda^2}$$

$$\left.\frac{\partial g}{\partial T_1}\right|_{T_1=\mu_1} = -\frac{1}{\mu_1^2} = -\lambda^2$$

$$\boxed{\mathrm{var}(\hat{\theta}) = \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} \mathbf{C}_T \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}} \qquad E(\hat{\theta}) = g(\boldsymbol{\mu})}$$

Chapter 9 – method of moments

Method of moments, exponential distribution continued

Recall estimator $$\hat{\lambda} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}$$

Identify $\hat{\lambda} = g(T_1)$

$$T_1 = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad g(T_1) = \frac{1}{T_1}$$

We should do T.S.E at mean of T

$$\mu_1 = E(T_1) = \frac{1}{\lambda}$$

Now, to the precision of the linear approximation

$$\mathrm{var}(T_1) = \frac{1}{N\lambda^2}$$

$$\left.\frac{\partial g}{\partial T_1}\right|_{T_1=\mu_1} = -\frac{1}{\mu_1^2} = -\lambda^2$$

$$\boxed{\mathrm{var}(\hat{\lambda}) = (-\lambda^2)\frac{1}{N\lambda^2}(-\lambda^2) = \frac{\lambda^2}{N}}$$

$$\boxed{\mathrm{var}(\hat{\theta}) = \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} \mathbf{C}_T \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}} \qquad E(\hat{\theta}) = g(\boldsymbol{\mu})}$$

Chapter 9 – method of moments

Method of moments, exponential distribution continued

Recall estimator

$$\hat{\lambda} = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} x[n]}$$

Identify $\hat{\lambda} = g(T_1)$

We should do T.S.E at mean of T

$$T_1 = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad g(T_1) = \frac{1}{T_1}$$

$$\mu_1 = E(T_1) = \frac{1}{\lambda}$$

Now, to the precision of the linear approximation

$$\mathrm{var}(T_1) = \frac{1}{N\lambda^2}$$

$$\left.\frac{\partial g}{\partial T_1}\right|_{T_1=\mu_1} = -\frac{1}{\mu_1^2} = -\lambda^2$$

$$\boxed{\mathrm{var}(\hat{\lambda}) = (-\lambda^2)\frac{1}{N\lambda^2}(-\lambda^2) = \frac{\lambda^2}{N}}$$

$$\boxed{\mathrm{var}(\hat{\theta}) = \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}}^{T} \mathbf{C}_T \left.\frac{\partial g}{\partial \mathbf{T}}\right|_{\mathbf{T}=\boldsymbol{\mu}} \qquad E(\hat{\theta}) = g(\boldsymbol{\mu})}$$

Fisher information
Per sample: λ^2

Chapter 9 – method of moments

In practice it is common to use cumulants rather than moments.

Cumulants are compressing the moments effectively

Cumulant generating function is log of moment generating function, thus, a a bit smaller in size and fluctuation" than the moments

Cumulants

$$\begin{aligned}
\kappa_2 &= \mu_2 \\
\kappa_3 &= \mu_3 \\
\kappa_4 &= \mu_4 - 3{\mu_2}^2 \\
\kappa_5 &= \mu_5 - 10\mu_3\mu_2 \\
\kappa_6 &= \mu_6 - 15\mu_4\mu_2 - 10{\mu_3}^2 + 30{\mu_2}^3
\end{aligned}$$

moments

Chapter 9 – method of moments

Example, PIMRC 2014

System model: y=x+n,
X is M-QAM, n is noise.

Solution: linear combination of cumulants

Task: estimate M

Used cumulants

$$C_{20} = M_{20}$$

$$C_{40} = M_{40} - 3{M_{20}}^2$$

$$C_{41} = M_{40} - 3M_{20}M_{21}$$

$$C_{42} = M_{42} - abs(M_{20})^2 - 2{M_{21}}^2$$

$$C_{60} = M_{60} - 15M_{20}M_{40} + 30{M_{20}}^3$$

$$C_{61} = M_{61} - 5M_{21}M_{40} - 10M_{20}M_{41} + 30{M_{20}}^2M_{21}$$

$$C_{62} = M_{62} - 6M_{20}M_{42} - 8M_{21}M_{41} - M_{22}M_{40} + 6{M_{20}}^2M_{22} + 24{M_{21}}^2M_{20}$$

$$C_{63} = M_{63} - 9M_{21}M_{42} + 12{M_{21}}^3 - 3M_{20}M_{43} - 3M_{22}M_{41} + 18M_{20}M_{21}M_{22} \quad (3)$$

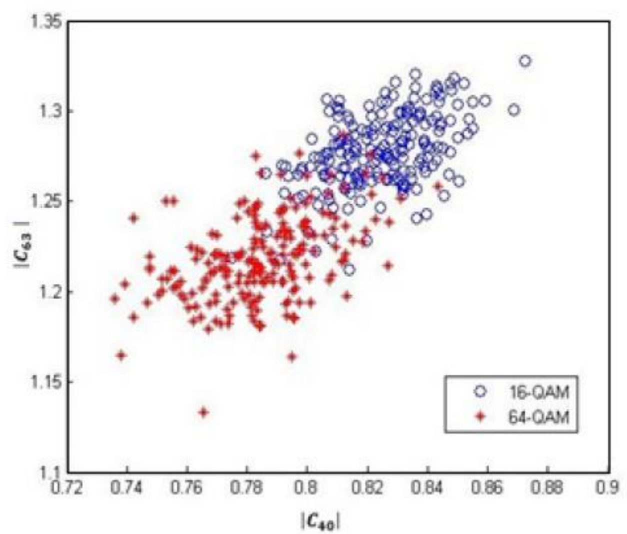


Figure 3: 200 realizations of 16-QAM and 64-QAM

Generalized method of moments

Fairly new method. Nobel prize in economics 2013.

Find function g(,) so that $m(\theta_0) \equiv \mathrm{E}[\,g(Y_t, \theta_0)\,] = 0$

Replace function by its estimator $\hat{m}(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T} g(Y_t, \theta)$

Minimize the norm of $\hat{m}(\theta)$ $\|\hat{m}(\theta)\|_W^2 = \hat{m}(\theta)' W \hat{m}(\theta),$

Example: reciprocity calibration of MaMIMO……

Generalized method of moments

Fairly new method. Nobel prize in economics 2013.

Find function g(,) so that $m(\theta_0) \equiv \mathrm{E}[\, g(Y_t, \theta_0) \,] = 0$

Replace function by its estimator $\hat{m}(\theta) \equiv \frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta)$

Minimize the norm of $\hat{m}(\theta)$ $\qquad \|\hat{m}(\theta)\|_W^2 = \hat{m}(\theta)' W \hat{m}(\theta),$

Optimal weights: $W \propto \Omega^{-1}$

$$\Omega = \mathrm{E}[\, g(Y_t, \theta_0) g(Y_t, \theta_0)' \,]$$