Introduction

Euklidische Norm:

$$\| x \|_2 \quad = \quad \sqrt{\sum_{i=1}^{n} x_i^2} \quad = \quad \sqrt{x^T x}$$

$$\| x \|_2^2 = x^T \cdot x$$

Weighting Eukledian Norm:

$$\| x \|_Q^2 = x^T Q \cdot x$$

Frobenius Norm:

$$\| x \|_F^2 = trace(AA^T) = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij} A_{ij}$$

$$\begin{aligned} \nabla f(x) &\quad Jacobian \\ \nabla^2 f(x) &\quad Hessian \end{aligned}$$

Error in variables

$$\hat{R}_{ev}(N) \quad = \quad \frac{\frac{1}{N}\sum_{k=1}^{N} u(k)}{\frac{1}{N}\sum_{k=1}^{N} i(k)}$$

Probablility and Statistics

Random Variables and Probability

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$P(A|B) = \frac{P(A,B)}{P(B)} \rightarrow \frac{P(B,A) \cdot P(A)}{P(B)}$$

$$P(X \in [a,b]) = \int_a^b p_X(x) dx$$

Mean

$$\mu_X = \mathbb{E}\{f(x)\} := \int_{-\infty}^{\infty} f(x) \cdot p_X(x) dx$$

$$\mathbb{E}\{a + bX\} := a + b\mathbb{E}\{X\}$$

Variance

$$\sigma_X^2 := \mathbb{E}\{(X - \mu_X)^2\} = \mathbb{E}\{X^2\} - \mu_X^2$$

$$stddev\, \sigma_X = \sqrt{variance\, \sigma_X^2}$$

Verteilungen

Uniform distribution:

$$P_y(x) = \begin{cases} \frac{1}{b-a} & if \quad x \in [a,b] \\ 0 & else \end{cases}$$

Normal distribution:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

Multidimensional Normal Distribution:

$$p(x) = \frac{1}{\sqrt{(2\pi)^n \cdot det(\Sigma)}} \cdot exp(-\frac{1}{2} \cdot (x-\mu)^T \cdot \Sigma^{-1} \cdot (x-\mu))$$

Weibull distribuation:

$$F(x) \, = \, 1 - e^{-(\lambda \cdot x)^k}$$

Laplace distribuation:

$$f(x|\mu, b) = \frac{1}{2b} \cdot exp(-\frac{|x-\mu|}{b})$$

Covariance and Correlaton:

$$\sigma(Y,Z) := \mathbb{E}(Y - \mu_Y)(Z - \mu_Z) =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_Y)(z - \mu_Z) \cdot p_{Y,Z}(y,z)\, dy\, dz$$

Multidimensional Random Variables:

$$\mathbb{E} f(X) = \int_{\mathbb{R}^n} f(x) p_X(x) d^n x$$

$$cov(X) = \mathbb{E}\{(X - \mu_X)(X - \mu_X)^T\}$$

$$cov(X) = \mathbb{E}\{XX^T\} - \mu_X \mu_X^T$$

$$cov(Y) = \Sigma_y = A \Sigma_x A^T \quad for \quad y = A \cdot x$$

$$\mathbb{E}\{AX\} = A \cdot \mathbb{E}\{X\}$$

Rules for variance:

$$var(X+Y) = var(X) + var(Y) + 2 \cdot cov(X,Y)$$

$$var(aX) = a^2 \cdot var(X)$$

Verschiebesatz:

$$var(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

unit Variance is variance = 1

Statistical estimators:
Biased- and unbiasedness $\rightarrow$ an estimator $\hat{\theta}_N$ is called unbiased iff $\mathbb{E}\{\hat{\theta}_N(y_N)\} = \theta_0$, where θ_0 is the true value of a parameter. Otherwise, is called biased.

Asymptotic Unbiasedness $\rightarrow$ An estimator $\hat{\theta}_N$ is called asymptotically unbiased iff $\lim_{n\to\infty} \mathbb{E}\{\hat{\theta}_N(y_N)\} = \theta_0$

Consistency $\rightarrow$ An estimator $\hat{\theta}_N(y_N)$ is called consistent if, for any $\epsilon > 0$, the probability $P(\hat{\theta}_N(y_N) \in [\theta_0 - \epsilon, \theta_0 + \epsilon])$ tends to one as $N \rightarrow \infty$.

Unconstrainded Optimization

Theorem 1 (First Order Necessary Conditions)
If $x^* \in D$ is local minimizer of $f : D \to \mathbb{R}$ and $f \in C^1$ then $\nabla f(x^*) = 0$ Definition (Stationary Point) A point $\bar{x}$ with $\nabla f(\bar{x}) = 0$ is called a stationary point of f.

Theorem 2 (Second Order Necessary Conditions)
If $x^* \in D$ is local minimizer of $f : D \to R$ and $f \in C^2$ then $\nabla^2 f(x^*) \succeq 0$

Theorem 3 (Second Order Sufficient Conditions and Stability under Perturbations)
Assume that $f : D \to R$ is C^2. If $x^* \in D$ is a stationary point and $\nabla^2 f(x^*) \succ 0$ then x^* is a strict local minimizer of f. In addition, this minimizer is locally unique and is stable against small perturbations of f, i.e. there exists a constant C such that for sufficiently small $p \in \mathbb{R}^n$ holds

$$\| x^* - arg\min_x (f(x) + p^T x) \| \leq C \| p \|$$

Linear Least Squares Estimation

Preliminaries: I.I.D and gaussian noise

Overall Model

$$y(k) = \phi(k)^T \theta + \epsilon(k)$$

Least Squares cost function as sum

$$\sum_{k=1}^{N} (y(k) - \phi(k)^T \theta)^2$$

Least Squares cost function

$$f(\theta) = \| y_N - \Phi_N \theta \|_2^2$$

Unique minimizers

$$\hat{\theta}_{LS} = arg\min_{\theta \in \mathbb{R}} f(\theta)$$

$$\theta^* = \underbrace{(\Phi^T \Phi)^{-1} \Phi^T}_{\Phi^+} y$$

Pseudo Inverse: $\quad \Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T$

Weighted Least Squares (unitless)
For I.I.D noise: Unweight Least Squares is optimal: W=I

$$\sum_{k=1}^{N} \frac{(y(k) - \phi(k)^T \theta)^2}{\sigma_\epsilon^2(k)}$$

$$f_{WLS}(\theta) = \| y_N - \Phi_N \theta \|_W^2 = (Y_N - \Phi \cdot \theta)^T \cdot W \cdot (Y_N - \Phi \cdot \theta)$$

$$\hat{\theta}_{WLS} = arg\min_{\theta \in \mathbb{R}} f_{WLS}(\theta) = (\Phi^T W \Phi)^{-1} \Phi^T W y$$

Singular Value Decomposition

$$A = USV^T \quad mit \quad U \in \mathbb{R}^{mxm}, \quad V \in \mathbb{R}^{nxn} und \quad S \in \mathbb{R}^{mxn}$$

Moore Penrose Pseudo Inverse

$$\Phi^+ = VS^+U^T$$

Regularization for Least Squares
$lim_{\alpha \to 0} (\Phi^T \Phi + \alpha I)^{-1} \Phi^T = \Phi^+ \; with \; \Phi^+ \; MPPI$

$$\theta^*(\alpha) = arg\min_{\theta \in R} \frac{1}{2} \| y - \Phi \theta \|_2^2 + \frac{\alpha}{2} \| \theta \|_2^2$$

Expectation of Least Squares Estimator

$$E\{\hat{\theta}_{WLS}\} = E\{(\Phi_N^T W \Phi_N)^{-1} \Phi_N^T W y_N\} = \theta_0$$

Covariance of the least squares estimator

$$cov(\hat{\theta}_{WLS}) = (\Phi_N^T W \Phi_N)^{-1} = (\Phi_N^T \Sigma_{\in N}^{-1} \Phi_N)^{-1}$$

$$cov(\hat{\theta}_{WLS}) \succeq (\Phi_N^T W \Phi_N)^{-1}$$

Example:

$$\varepsilon(1) \sim \mathcal{N}(0|\sigma_1^2) \quad \varepsilon(2) \sim \mathcal{N}(0|\sigma_2^2)$$

$$N = 2 \quad \Sigma_{\varepsilon_N} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$$W^{OPT} = \Sigma_{\varepsilon_N}^{-1} \quad \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix}$$

$$cov(\hat{\theta}_{WLS}) = (Y_N - \Phi_N \theta)^T \cdot W \cdot (Y_N - \Phi_N \theta) =$$

$$\sum_{k=1}^{2} (y(k) - \phi(k)^T \theta) \cdot \frac{1}{\sigma_k^2} \cdot (y(k) - \phi(k)^T \theta)$$

Measuring the goodness of Fit using $R^2 \quad 0 \leq R^2 \leq 1$

$$R^2 = 1 - \frac{\| y_N - \Phi_N \hat{\theta} \|_2^2}{\| y_N \|_2^2} = 1 - \frac{\| \epsilon_N \|_2^2}{\| y_N \|_2^2} =$$

$$\frac{\| y_N \|_2^2 - \| \epsilon_N \|_2^2}{\| y_N \|_2^2} = \frac{\| \hat{y}_N \|_2^2}{\| y_N \|_2^2}$$

residual $\epsilon_N \uparrow \quad \to \quad R^2 \to 0 \; (= bad)$

Estimating the Covariance with the Single Experiment

$$\hat{\sigma}_\varepsilon^2 := \frac{1}{N-d} \sum_{k=1}^{N} (y(k) - \phi(k)^T \hat{\theta}_{LS})^2 = \frac{\| y_N - \phi_N \hat{\theta}_{LS} \|_2^2}{N-d}$$

$$\hat{\Sigma}_{\hat{\theta}} := \hat{\sigma}_\varepsilon^2 (\phi_N^T \phi_N)^{-1} = \frac{\| y_N \; - \phi_N \hat{\theta}_{LS} \|_2^2}{N-d} \cdot (\phi_N^T \phi_N)^{-1}$$

Maximum Likelihood Estimation

Maximum Likelihood Estimation (ML) L_2 Estimation:
Measurement Errors assumed to be Normally distributed

$$P(y|\theta) = C \prod_{i=1}^{N} exp(\frac{-(y_i - M_i(\theta))^2}{2 \cdot \sigma_i^2})$$

Positive log-Likelihood. Logarithm makes from products a sum!

$$log\, p(y|\theta) = log(C) + \sum_{i=1}^{N} \frac{-(y_i - M_i(\theta))^2}{2 \cdot \sigma_i^2}$$

Negative log-Likelihood:

$$\hat{\theta}_{ML} = arg \max_{\theta \in \mathbb{R}^d} p(y|\theta) = arg \min_{\theta \in \mathbb{R}^d} \sum_{i=1}^{m} \frac{(y_i - M_i(\theta))^2}{2\, {\sigma_i}^2}$$

$$arg \max_{\theta \in \mathbb{R}^d} p(y|\theta) = arg \min_{\theta \in R^d} \quad \frac{1}{2} \parallel S^{-1} \cdot (y - M(\theta)) \|_2^2$$

L_1 Estimation:
Measurement Errors assumed to be Laplace distributed.
$Median(x) = \lceil \frac{x+1}{2} \rceil$
Robust against outliers

$$\min_{\theta} \| \ y - M(\theta) \ \|_1 = \min_{\theta} \sum_{i=1}^{N} |y_i - M_i \theta| =$$

$$= median\, of \, \{Y_1, ..., Y_N\}$$

$$P(y|\theta) = C \prod_{i=1}^{N} exp(\frac{-|y_i - \theta|}{2 \cdot a_i})$$

Bayesian Estimation and the Maximum a Posteriori Estimate
Assumptions: i.i.d noise and linear model

$$p(\theta|y_N) \cdot p(y_N) = p(y_N|\theta) \cdot p(\theta)$$

$$\hat{\theta}_{MAP} = arg \min_{\theta \in \mathbb{R}} \{-log(p(y_N|\theta)) - log(p(\theta))\}$$

MAP Example: Regularised Least Squares

$$\theta = \overline{\theta} \pm \sigma_\theta \quad with \quad \overline{\theta} = \theta_{apriori}$$

$$\hat{\theta}_{MAP} = arg \min_{\theta \in R} \frac{1}{2} \cdot \frac{1}{\sigma_\epsilon^2} \cdot \parallel y_N - \Phi_N \cdot \theta \parallel_2^2 + \frac{1}{2} \cdot \frac{1}{\sigma_\theta^2} \cdot (\theta - \overline{\theta})^2$$

Recursive Linear Least Squares

$$\theta_{ML}(N) = arg \min_{\theta \, \in \, R} \frac{1}{2} \parallel y_N - \Phi_N \cdot \theta \|_2^2$$

$$\hat{\theta}_{ML}\,(N+1) = arg \min_{\theta \, \in \, R^d} (\alpha \cdot \frac{1}{2} \cdot \parallel \theta - \hat{\theta}_{ML}(N) \|_{Q_N}^2 +$$

$$\frac{1}{2} \cdot \parallel y(N+1) - \varphi(N+1)^T \cdot \theta \|_2^2)$$

$$Q_0 \quad given, \quad and \quad \hat{\theta}_{ML}(0) \quad given,$$

$$Q_{N+1} = \alpha \cdot Q_N + \varphi(N+1) \cdot \varphi(N+1)^T,$$

$$\hat{\theta}_{ML}(N+1) = \hat{\theta}_{ML}(N) + Q_{N+1}^{-1} \cdot \varphi(N+1) \cdot [y(N+1) - \varphi(N+1)^T \cdot \hat{\theta}_{ML}(N)]$$

Cramer-Rao-Inequality

$$\Sigma_\theta \succeq M^{-1} = (\Phi_N^T \cdot \Sigma^{-1} \cdot \Phi)^{-1}$$

$$L(\theta, y_N) = \frac{1}{2} \cdot (\Phi_N \cdot \theta - y_N)^T \cdot \Sigma^{-1} \cdot (\Phi_N \cdot \theta - y_N) = log\,(p(y_N|\theta))$$

$$M = E\{\nabla_\theta^2 \, L(\theta, y_N)\} = \nabla_\theta^2 \, L(\theta, y_N) = \Phi_N^T \cdot \Sigma^{-1} \cdot \Phi_N$$

Dynamic Models

Linear Time Invariant (LTI) Systems
with A, B, C, D are matrices

$$\dot{x} = Ax + Bu \quad y = Cx + Du$$

$$G(s) = C(sI - A)^{-1}B + D$$

LTI sytems as Input-Output Models

$$G(S) = \frac{b_0 + b_1 s + ... + b_n s^n}{a_0 + a_1 s + ... + a_{n-1} s^{n-1} + s^n}$$

Different Models
Deterministic Model: $y(k) = M(k; U, x_{init}, p)$
Model with measurement Noise:
$y(k) = M(k; U, x_{init}, p) + \varepsilon(k)$
Model with Input and Output Errors:
$y(k) = M(k; U + \varepsilon_N^u, x_{init}, p) + \varepsilon^y(k)$

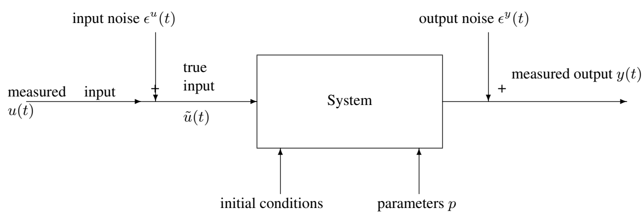
Pure Output Error (OE) Minimization

$$\theta_{ML} = arg \min_{\theta} \sum_{k=1}^{N} (y(k) - M(k; U, \, x_{init}\,, \, p)\,)^2$$

Output Error Minimization for FIR Models

$$y(k) = (u(k), \, u(k-1), \, ..., \, u(k - n_{n_b})) \cdot \theta + \varepsilon(k)$$

$$\min_{\theta} \sum_{k=n_b+1}^{N} (\quad y(k) - (u(k), \, u(k-1), \, ..., \, u(k - n_{n_b})) \cdot \theta \quad)^2$$



Models with Input and Output Errors

$$arg \min_{\theta} \sum_{k-1}^{N} \frac{1}{\sigma_y^2} (y(k) - M(k; U + \epsilon_N^u, x_{init}, p))^2 + \frac{1}{\sigma_u^2} (\epsilon_u(k))^2$$

$$arg \min_{\theta} \sum_{k-1}^{N} \frac{1}{\sigma_y^2} (y(k) - M(k; \tilde{U}, x_{init}, p))^2 + \frac{1}{\sigma_u^2} (u(k) - \tilde{u}(k))^2$$

Fourier Transformation

How to compute FT? By DFT, which solves the problem of finite time and discrete values.

Can we use an input with many frequencies to get many FRF (Frequency Response Function) values in a single experiment? So far only frequency sweeping (high comp. times due to repetition for each frequency). We should use multisines!

Aliasing and Leakage Errors

Aliasing Error due to sampling of continous signal to discrete signal. Avoid with Nyquist Theoreme:

$$f_{Nyquist} = \frac{1}{2\Delta t}[Hz] \quad or \quad \omega_{Nyquist} = \frac{2\pi}{2\Delta t}[rad/s]$$

Leackage Error due to windowing.

$$\omega_{base} := \frac{2\pi}{N \cdot \Delta t} = \frac{2\pi}{T} \rightarrow \omega = m\frac{2\pi}{N \cdot \Delta t}$$

Crest Factor = Scheitelfaktor

$$CrestFactor = \frac{u_{max}}{u_{rms}} \quad with:$$

$$u_{rms} := \sqrt{\frac{1}{T}\int_0^T u(t)^2\,dt} \quad and \quad u_{max} := \max_{t\in[0,T]}|u(t)|$$

Optimising Multisine for optimal crest factor

Frequency: Choose frequencies in logarithmic manner as multiples of the base frequency. $\omega_{k+1}/\omega_k \approx 1.05$

Phase: To prevent high peaks (Crest Factor) in the Signal, the phases of the different frequencies are modulated accordingly. (Positive interference)

Multisine Identification Implementation procedure

Window Length integer multiple of sampling time: $T = N \cdot \delta t$

Harmonics of base frequency are contained in multisine $\omega_{base} = \frac{2\pi}{T}$

Highest contained Frequency is **half** of Nyquist frequency: $\omega_{Nyquist} = \frac{2\pi}{4\Delta T}$

Experiment and Analysis (step 2): Insert Multisine periodically. Drop first Periods (till transients died out). Record M Periods, each with N samples, of input and output data. Average all the M periods and make the DFT (or vice versa). Finally build transfer function.

$$\hat{G}_{j\omega_k} = \frac{\hat{Y}(k(p))}{\hat{U}(k(p))}$$

Nonparametric and Frequency Domain Identification Models

Impulse response and transfer function

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)\delta t$$

$$Y(s) = G(s) \cdot U(s)$$

$$G(s) = \int_0^\infty \epsilon^{-st} g(t)dt$$

Bode diagram from frequency sweeps

$$u(t) = A \cdot sin(\omega \cdot t), \quad y(t) = \| G(j\cdot\omega) \| \, A \cdot sin(\omega \cdot t + \alpha)$$

Online estimation for dynamic systems

Recursive Least Squares

New Inverse Covariance:

$$Q_K \;=\; Q_{k-1} + \phi_K\,\phi_K^T$$

Innovation update:

$$\hat{\theta}_k \;=\; \hat{\theta}_{k-1} + \underbrace{Q_k^{-1}\,\phi_k\,(y_k \;-\; \phi_k^T\,\hat{\theta}_{k-1})}_{``innovation''}$$

General Optimization Problem:

$$\hat{\theta}_k \;=\; arg\min_{\theta}\,(\theta - \hat{\theta}_0)^T \cdot Q_0 \cdot (\theta - \hat{\theta}_0) + \sum_{i=1}^{k}(y_k - \phi_k^T \;\cdot\theta)^2$$

Kalman Filter

Valid for Discrete and Linear!
(If recursive least squares: $x_{k+1} \;=\; A_k \cdot x_k$

$$x_{k+1} \;=\; A_k \cdot x_k + \omega_k \quad and \quad y_k = C_k \cdot x_k + v_k$$

Steps of Kalman Filter

1 Prediction $\hat{x}_{[k\,|\,k-1]} \;=\; A_{k-1} \cdot \hat{x}_{[k-1\,|\,k-1]}$
$P_{[k\,|\,k-1]} \;=\; A_{k-1} \cdot P_{[k-1\,|\,k-1]} \;\cdot\; A_{k-1}^T \cdot W_{k-1}$
if recursive linear least squares without W_{k-1}.

2 Innovation update $P_{[k\,|\,k]} \;=\; (P_{[k\,|\,k-1]}^{-1} + C_k^T \;\cdot\; V^{-1} \;\cdot\; C_k)^{-1}$
$\hat{x}_{[k\,|\,k]} \;=\; \hat{x}_{[k\,|\,k-1]} + P_{[k\,|\,k]} \cdot C_k^T \cdot V^{-1} \cdot (y_k - C_k \cdot \hat{x}_{[k\,|\,k-1]})$

Bode Diagram:

Magnitude = Amplitude $|G(j\omega)|$
Phase $arg\,G(j\omega)$