

## Introduction

**Euklidian Norm:**  $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{x^T x}$   
 $\|x\|_2^2 = x^T \cdot x$

**Weighting Eukl. Norm:**  $\|x\|_Q^2 = x^T Q \cdot x$

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

Jacobian:  $\nabla f(x) = \frac{\partial f}{\partial x}(x)$  in  $\mathbb{R}^{n \times m}$       Hessian:  $\nabla^2 f(x)$

$$(AB)^T = B^T A^T \quad (AB)^{-1} = B^{-1} A^{-1} \quad (A^T)^{-1} = (A^{-1})^T$$

**Matrix derivatives:**  $\frac{d(c^T x)}{dx} = c \quad \frac{d(x^T A x)}{dx} = (A^T + A)x$

**Linear and non-linear models:**

- linear if parameters linear i.e.  $(\theta_1 x^2 + \theta_2 x + \theta_3)$
- nonlinear if i.e  $(\sin(\theta_1)x + \theta_2)$  or derivatives in other orders than 1

**Table of Derivatives:**

f(x)	f'(x)
$g(x) \cdot h(x)$	$g'(x) \cdot h(x) + g(x) \cdot h'(x)$
$g(h(x))$	$g'(h(x)) \cdot h'(x)$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\tan(x) = \frac{\sin(x)}{\cos(x)}$	$\frac{1}{\cos^2(x)} = \sec^2(x)$
$e^{kx}$	$\frac{1}{k} e^{kx}$
$\ln(x)$	$\frac{1}{x}$
$\log_a x$	$\frac{1}{\ln a} (x \ln x - x)$
$Ax$	$A$
$x^T A$	$A^T$
$x^T B x$	$x^T (B^T + B)$

## Random Variables and Probability

**Dependent Probability:**  $P(A \vee B) = P(A) + P(B)$

**Independent Prob.:**  $P(A, B) = P(A \wedge B) = P(A) \cdot P(B)$

**Conditional Prob.:**  $P(A|B) = \frac{P(A \wedge B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}$  (Bayes' theorem)

$$P(X \in [a, b]) = \int_a^b p_X(x) dx \quad p(x|y) = \frac{p(x, y)}{p(y)}$$

**Mean/Expectation value:**  $\mathbb{E}\{\mu_X\} := \mu_X = \int_{-\infty}^{\infty} x \cdot p_X(x) dx$

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

$$\mathbb{E}\{XY\} = \mathbb{E}\{X\} \cdot \mathbb{E}\{Y\} \Leftrightarrow X, Y \text{ independent}$$

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

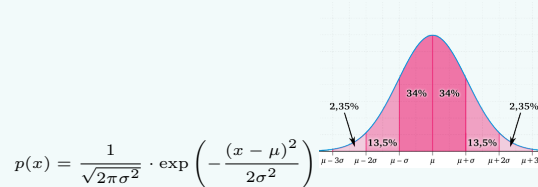
**Standard deviation:**  $\sigma_X = \sqrt{\sigma_X^2}$

## Distributions

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

**Mean:**  $\mu_X = \int_{-\infty}^{\infty} x p_X(x) dx = \int_a^b \frac{1}{b-a} \cdot x dx = \frac{a+b}{2} =: \mu_X$

**Normal distribution:**  $X \sim \mathcal{N}(\mu, \sigma^2) \quad \hat{\theta}_{LS} \sim \mathcal{N}(\theta_0, \Sigma_{\hat{\theta}})$



**Multidimensional Normal Distribution:**

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

**Weibull distribution:**  $F(x) = 1 - \exp(-(\lambda \cdot x)^k)$

**Laplace distribution:**  $f(x|\mu, b) = \frac{1}{2b} \cdot \exp\left(-\frac{|x - \mu|}{b}\right)$

## Useful statistic definitions

**Covariance and Correlaton:**  $\sigma(X, Y) := \mathbb{E}(X - \mu_X)(Y - \mu_Y)$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot p_{X,Y}(x, y) dx dy$$

**Covariance Matrix:**  $\Sigma_x = \text{cov}(X) = \mathbb{E}\{XX^T\} - \mu_x \mu_x^T$  is PSD

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{yx} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \quad \sigma_{xy} = \sigma_{yx} = \rho \sigma_x \cdot \sigma_y \text{ where } \rho \text{ is correlation}$$

$x$  and  $y$  are i.i.d.  $\Rightarrow \Sigma$  is diagonal

**Multidimensional Random Variables:**

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

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

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

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

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

**Rules for variance:**

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

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

**Formula for variance:**  $\text{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

**Correlation:**

uncorrelated if  $\rho(X, Y) = 0$ ,  $\rho(X, Y) := \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$

**Statistical estimators:**

**Biased- and Unbiasedness** An estimator  $\hat{\theta}_N$  is unbiased  $\Leftrightarrow \mathbb{E}\{\hat{\theta}_N(y_N)\} = \theta_0$ , where  $\theta_0 \equiv$  "true" value of  $\theta$ . Otherwise: biased.

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

**Consistency** 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 1 for  $N \rightarrow \infty$ .

## Unconstrained Optimization

**Theorem 1:** (First Order Necessary Conditions)

If  $x^* \in D$  is local minimizer of  $f : D \rightarrow \mathbb{R}$  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  $f : D \rightarrow \mathbb{R}$  and  $f \in C^2$  then  $\nabla^2 f(x^*) \succeq 0$

**Theorem 3:** (Second Order Sufficient Conditions and Stability under Perturbations)

Assume that  $f : D \rightarrow \mathbb{R}$  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

$$\|x^* - \underset{x}{\operatorname{argmin}}(f(x) + p^T x)\| \leq C \|p\|$$

## Linear Least Squares Estimation

**Preliminaries:** i.i.d. and Gaussian noise

**Overall Model:**  $y(k) = \phi(k)^T \theta + \varepsilon(k)$

**LS cost function as sum:**  $\sum_{k=1}^N (y(k) - \phi(k)^T \theta)^2$

**LS cost function:**  $f(\theta) = \|y_N - \Phi_N \theta\|_2^2$

**Unique minimizers:**  $\hat{\theta}_{LS} = \underset{\theta \in \mathbb{R}}{\operatorname{argmin}} f(\theta) \theta^* = \underbrace{(\Phi^T \Phi)^{-1} \Phi^T y}_{:= \Phi^+}$

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

## Weighted Least Squares (unitless)

**For i.i.d noise:** Unweight Least Squares is optimal:  $W = I$

$$f_{WLS}(\theta) = \sum_{k=1}^N \frac{(y(k) - \phi(k)^T \theta)^2}{\sigma_{\varepsilon}^2(k)} = \|y_N - \Phi_N \theta\|_W^2$$

$$= \|W^{\frac{1}{2}} y - W^{\frac{1}{2}} \Phi_N \theta\|_2^2 = (y_N - \Phi \cdot \theta)^T \cdot W \cdot (y_N - \Phi \cdot \theta)$$

**Solution for WLS:**

$$\hat{\theta}_{WLS} = \tilde{\Phi}^+ \tilde{y} \quad \text{mit } \tilde{\Phi} = W^{\frac{1}{2}} \Phi \text{ und } \tilde{y} = W^{\frac{1}{2}} y$$

$$= \underset{\theta \in \mathbb{R}}{\operatorname{argmin}} f_{WLS}(\theta) = (\Phi^T W \Phi)^{-1} \Phi^T W y$$

## Ill-Posed Least Squares

**Singular Value Decomposition:**  $A = USV^T \in \mathbb{R}^{m \times n}$  with  $U \in \mathbb{R}^{m \times m}$ ,  $V \in \mathbb{R}^{n \times n}$  and  $S \in \mathbb{R}^{m \times n}$  where  $S$  is a diagonal Matrix with non-negative elements  $(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$

**Moore Penrose Pseudo Inverse:**

$$\Phi^+ = V S^+ U^T = V (S^T S + \alpha I)^{-1} S^T U^T$$

$\Phi^+$  therefore selects  $\theta^* \in S^*$  with minimal norm

**Regularization for Least Squares:**

$$\lim_{\alpha \rightarrow 0} (\Phi^T \Phi + \alpha I)^{-1} \Phi^T = \Phi^+ \quad \text{with } \Phi^+ \text{ MPPI}$$

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

## Statistical Analysis of WLS

### 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:

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

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

## Example LLS

### Example of the Linear Least Square Estimator for: $N = 2$

$$\epsilon(1) \sim \mathcal{N}(0|\sigma_1^2)$$

$$\epsilon(2) \sim \mathcal{N}(0|\sigma_2^2)$$

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

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

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

### Measuring the goodness of Fit using: $R^2$ ( $0 \leq R^2 \leq 1$ )

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

$$\text{Residual: } \epsilon_N \uparrow \rightarrow R^2 \rightarrow 0 \ (\Rightarrow \text{bad})$$

### Estimating the Covariance with the Single Experiment:

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

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

## Bayesian Estimation and the Maximum a Posteriori Estimate

### Assumptions:

- Measurement:  $y_N \in \mathbb{R}^N$  has i.i.d. noise
- Linear Model:  $M(\theta) = \phi_N \cdot \theta$  and  $\theta \in \mathbb{R}$

$$p(\theta|y_N) = \frac{p(y_N, \theta)}{p(y_N)} = \frac{p(y_N|\theta) \cdot p(\theta)}{p(y_N)}$$

$$\hat{\theta}_{MAP} = \underset{\theta \in \mathbb{R}}{\text{argmin}} \left\{ \underbrace{-\log(p(y_N|\theta))}_{\text{Max. Likelihood prev. knowledge}} \underbrace{-\log(p(\theta))}_{\text{prior}} \right\}$$

### MAP Example: Regularised Least Squares

$$\theta = \hat{\theta} \pm \sigma_{\theta} \quad \text{with} \quad \hat{\theta} = \theta_{\text{a-priori}}$$

$$\hat{\theta}_{MAP} = \underset{\theta \in \mathbb{R}}{\text{argmin}} \frac{1}{2} \cdot \frac{1}{\sigma_{\epsilon}^2} \cdot \|y_N - \Phi_N \cdot \theta\|_2^2 + \frac{1}{2} \cdot \frac{1}{\sigma_{\theta}^2} \cdot (\theta - \bar{\theta})^2$$

## Maximum Likelihood Estimation

### $L_2$ Estimation: Maximum Likelihood Estimation (ML):

- Measurement Errors assumed to be Normally distributed
- Model described by a non-linear function  $M(\theta)$
- Every unbiased estimator needs to satisfy the Cramer-Rao inequality, which gives a lower bound on the covariance matrix.

$$\text{Model: } y = M(\theta) + \epsilon$$

$$p(y|\theta) = C \prod_{i=1}^N \exp\left(\frac{-(y_i - M_i(\theta))^2}{2 \cdot \sigma_i^2}\right) \quad C = \prod_{i=1}^N \frac{1}{\sqrt{2 \cdot \pi \sigma_i^2}}$$

### Positive log-Likelihood: (log changes products into sums)

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

### Negative log-Likelihood:

$$\hat{\theta}_{ML} = \underset{\theta \in \mathbb{R}^d}{\text{argmax}} p(y|\theta) = \underset{\theta \in \mathbb{R}^d}{\text{argmin}} \sum_{i=1}^N \frac{(y_i - M_i(\theta))^2}{2\sigma_i^2}$$
$$= \underset{\theta \in \mathbb{R}^d}{\text{argmin}} \frac{1}{2} \sum_{i=1}^N \left( \frac{y_i - M_i(\theta)}{\sigma_i} \right)^2$$

$$= \underset{\theta \in \mathbb{R}^d}{\text{argmin}} \frac{1}{2} \|S^{-1}(y - M(\theta))\|_2^2 \quad \text{with: } S = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_N \end{bmatrix}$$

### $L_1$ Estimation:

- Measurement Errors assumed to be Laplace distributed and more robust against outliers.

$$\min_{\theta} \|y - M(\theta)\|_1 = \min_{\theta} \sum_{i=1}^N |y_i - M_i(\theta)|$$
$$\Rightarrow \text{median of } \{Y_1, \dots, Y_N\}$$

## Recursive Linear Least Squares

$$Q_{N+1} = \alpha \cdot Q_N + \varphi(N+1) \cdot \varphi(N+1)^T, \quad \alpha \triangleq \text{"forgettingfactor"}$$

$$\hat{\theta}_{ML}(N+1) = \hat{\theta}_{ML}(N) + Q_{N+1}^{-1} \cdot \varphi(N+1)$$
$$\cdot \left[ \underbrace{y(N+1)}_{\text{new measurement}} - \underbrace{\varphi(N+1)^T \cdot \hat{\theta}_{ML}(N)}_{\text{old prediction}} \right]$$

$Q_0$  and  $\hat{\theta}_0$  have to be chosen.

$Q_0$  should be **non-singular, small and positive definite**. (e.g.  $10^{-3} \cdot I$ )

$$Q_N \approx \Sigma_{\hat{\theta}_{ML}(N)}^{-1}$$

## Cramer-Rao-Inequality (Fisher information Matrix M)

$$\text{cov}(\hat{\theta}(y_N)) = \Sigma_{\hat{\theta}} \succeq M^{-1} = \underbrace{(\Phi_N^T \cdot \Sigma^{-1} \cdot \Phi_N)^{-1}}_{\text{for lin. model w. } \epsilon \sim \mathcal{N}(\mu, \sigma)}$$

$$M = \int_{y_N} \nabla_{\theta}^2 L(\theta_0, y_N) \cdot p(y_N|\theta_0) dy_N$$

### Assumptions:

- Minimising a Linear Model
- Gaussian Noise:  $X \sim \mathcal{N}(0, \Sigma)$

$$L(\theta, y_N) = -\log(p(y_N|\theta))$$

$$(\text{if lin. model, etc.}) = \frac{1}{2} \cdot (\Phi_N \cdot \theta - y_N)^T \cdot \Sigma^{-1} \cdot (\Phi_N \cdot \theta - y_N)$$

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

$$\Rightarrow W = \Sigma^{-1} \text{ is the optimal weighting Matrix for WLS.}$$

## Continuous Time Systems

### Ordinary Differential Equations (ODE):

$$\dot{x} = f(x(t), u(t), \epsilon(t), p)$$

### Differential Algebraic Equations(DAE):

$$\dot{x} = f(x(t), u(t), \epsilon(t), p)$$

$$0 = g(x, z).$$

### LTI Sytem (ODE):

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

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

## Numerical Integration Methods

### Euler Integration Step

$$\hat{x}(t; x_0, u_{\text{const}}) = x_0 + t f(x_0, u_{\text{const}}), \quad t \in [0, \Delta t]$$

$$\hat{x}_{j+1} = \hat{x}_j + h f(\hat{x}_j, u_{\text{const}}), \quad j = 0, \dots, M-1$$

- Approximation becomes better by decreasing the step size  $h$ .
- Consistency Error:  $h^2$
- Total Number of steps:  $\Delta t/h$
- Error in the final step of order  $h\Delta t$
- Linear in step size  $\rightarrow$  order one
- Taking more steps is more accurate but needs more computation

### Runge-Kutta Method of Order Four (RK4)

$$k_1 = f(\tilde{x}_j, u_{\text{const}})$$

$$k_2 = f(\tilde{x}_j, \frac{h}{2} k_1, u_{\text{const}})$$

$$k_3 = f(\tilde{x}_j, \frac{h}{2} k_2, u_{\text{const}})$$

$$k_4 = f(\tilde{x}_j, h k_3, u_{\text{const}})$$

$$\tilde{x}_{j+1} = \tilde{x}_j + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4)$$

One Step of RK4 is thus as expensive as four steps of euler

accuracy of final approximation is of order  $h^4 \Delta t$

$\rightarrow$  RK4 needs fewer functions to obtain the same accuracy level as euler

## Discrete Time Systems

Det. Model as State Space      Stoch. Model as State Space

Det. Model as Input-Output      Stoch. Model as Input-Output

### State Space Model

$x_{k+1} = f_k(x_k, u_k), k = 0, 1, \dots, N-1$  with input vector  $u_k$  and state vector  $x_k$

### Input-Output Model

$$y(k) = h(u(k), \dots, u(k-n), y(k-1), \dots, y(k-n))$$

### LTI system as State-Space Model:

$$x_{k+1} = Ax_k + Bu_k, \quad k = 0, 1, \dots, N-1.$$

### LTI system as Input-Output Model:

$$G(s) = \frac{b_0 + b_1 s + \dots + b_n s^n}{a_0 + a_1 s + \dots + a_{n-1} s^{n-1} + s^n} \quad | \cdot s = z^{-1}$$

$$G(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_n z^{-n}}{a_0 + a_1 z^{-1} + \dots + a_n z^{-n}}$$
$$= \frac{b_0 z^n + b_n z^{n-1} + \dots + b_n}{a_0 z^n + a_1 z^{n-1} + \dots + a_n} \Rightarrow \text{Also called "polynomial model".}$$

## Deterministic Model

The output of the system can be obtained with absolute certainty. The Output  $y$  or the state  $x$ , depend on the known inputs  $u(1), \dots, u(N)$ , the previous Outputs  $y(1), \dots, y(N)$  or state  $x(n-1)$  and initial conditions. All deterministic models are **time invariant**.

**State Space Model:**

$$x(k+1) = f(x(k), u(k))$$

$$y(k) = g(x(k), u(k))$$

Initial conditions:  $x(1) = x_0$

**Input-Output Model**

$$y(k) = h(u(k), \dots, u(k-n), y(k-1), \dots, y(k-n))$$

Initial conditions:  $y(1) = y_1, \dots, y(n) = y_n$   $u(1) = u_1, \dots, u(n) = u_n$

**Finite Impulse Response (FIR):**

$$y(k) = b_0 u(k) + \dots + b_{n_b} u(k - n_b)$$

$$G(z) = b_0 + b_1 z^{-1} + \dots + b_{n_b} z^{-n_b} \quad | \cdot \frac{z^{n_b}}{z^{n_b}}$$

$$= \frac{b_0 z^{n_b} + b_1 z^{n_b-1} + \dots + b_{n_b}}{z^{n_b}}$$

**Auto Regressive model with eXogenous inputs (ARX/IRR):**

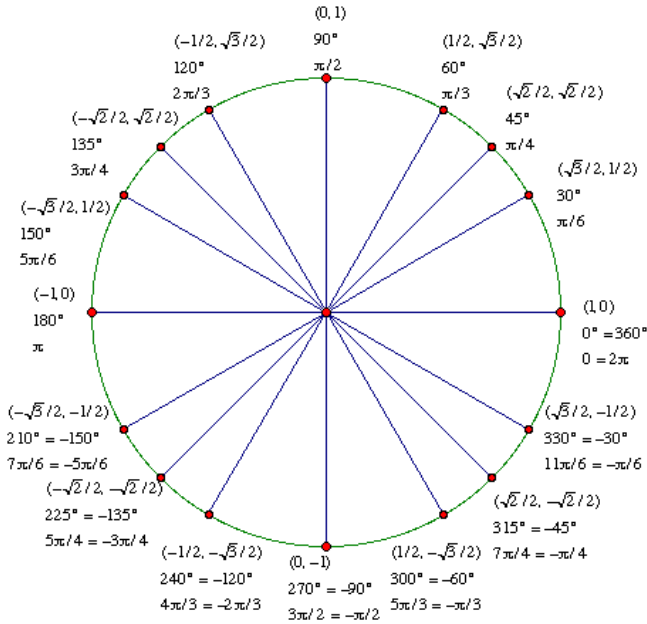
$$a_0 y(k) + \dots + a_{n_a} y(k - n_a) = b_0 u(k) + \dots + b_{n_b} u(k - n_b)$$

$$G(z) = \frac{b_0 z^n + b_1 z^{n-1} + \dots + b_n}{a_0 z^n + a_1 z^{n-1} + \dots + a_n}$$

The next output depends on the previous output. Also called **IIR** (infinite impulse response)

**Auto Regressive model(AR):**

$$y(k) = a_1 y(k-1) - \dots - a_{n_a} y(k - n_a)$$



## Stochastic Model

Real systems are far from deterministic.

- there is stochastic noise  $\varepsilon(k)$
- there are constant and unknown parameters  $p$
- measured outputs  $y(k)$  depend in both,  $\varepsilon(k)$  and  $p$

Assumptions: noise is **i.i.d** and enters the model like a normal input, but as a random variable

**State Space Model**

$$x(k+1) = f(x(k), u(k), \varepsilon(k))$$

$$y(k) = g(x(k), u(k), \varepsilon(k))$$

**Input-Output Model**

Only interested in input and output, not the whole model state

$$y(k) = h(u(k), \dots, u(k-n), y(k-1), \dots, y(k-n), \varepsilon(k), \dots, \varepsilon(k-n))$$

for  $k = n+1, n+2, \dots$

**Measurement Noise (Output Error Model)**

$$y(k) = M(k; U, x_0, p) + \varepsilon(k)$$

**Stochastic Disturbance (Equation Errors)**

$$y(k) = h(p, u(k), \dots, u(k-n), y(k-1), \dots, y(k-n)) + \varepsilon(k)$$

for  $k = n+1, n+2, \dots$

**Linear In the Parameters models (LIP):**

$$y(k) = \sum_{i=1}^d \theta_i \phi_i(u(k), \dots, y(k-1), \dots) + \varepsilon(k)$$

$$y(k) = \varphi(k)^T \theta + \varepsilon(k) \quad \text{where } \varphi = (\phi_1(\cdot), \dots, \phi_d(\cdot))$$

**LIP-LTI Models with Equation Errors (ARX)<sup>a</sup>**

- Combining best of two worlds (LTI and LIP)

$$a_0 y(k) + \dots + a_{n_a} y(k - n_a) = b_0 u(k) + \dots + b_{n_b} u(k - n_b) + \varepsilon(k)$$

**Auto-Regressive Moving Average with eXogeneous input (ARMAX):**

$$a_0 y(k) + \dots + a_{n_a} y(k - n_a) = b_0 u(k) + \dots + b_{n_b} u(k - n_b) + \varepsilon(k) +$$

$$c_1 \varepsilon(k-1) + \dots + c_{n_c} \varepsilon(k - n_c)$$

**Auto-Regressive Moving Average without inputs (ARMA):**

$$a_0 y(k) + \dots + a_{n_a} y(k - n_a) = \varepsilon(k) + c_1 \varepsilon(k-1) + \dots + c_{n_c} \varepsilon(k - n_c)$$

Where  $c_i$  represent the noise coefficient, we have to use non-linear least squares with the unknown noise terms  $\varepsilon(k-i)$

**Difference between Deterministic and Stochastic Models**

- stochastic noise  $\varepsilon(k)$
- unknown but constant parameter  $p$
- measured output  $y(k)$  depend on both,  $\varepsilon(k)$  and  $p$

<sup>a</sup>additive noise is a special case

## Example for State Space Model

$$\dot{a} = m \cdot \dot{a} + g \cdot a + c \cdot u$$

$$y = \dot{a}$$

$$x = \begin{bmatrix} a \\ \dot{a} \end{bmatrix} \quad \dot{x} = \begin{bmatrix} \dot{a} \\ \ddot{a} \end{bmatrix} \quad \dot{x} = Ax + Bu \quad y = Cx + Du$$

$$A = \begin{bmatrix} 0 & 1 \\ g & m \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ c \end{bmatrix} \quad C = \begin{bmatrix} 0 & 1 \end{bmatrix} \quad D = \begin{bmatrix} 0 \end{bmatrix}$$

## Pure Output Error (OE) Minimization

Assume: i.i.d. gaussian noise only affecting output using non-linear least squares

$$\theta_{ML} = \min_{\theta} \sum_{k=1}^N (y(k) - M(k; U, x_0 p))^2$$

**Output Error Minimization for FIR Models:** lead to convex problems, therefore global minimum can be found

$$y(k) = (u(k), u(k-1), \dots, u(k-n_b)) \cdot \theta + \varepsilon(k)$$

$$= \min_{\theta} \sum_{k=n_b+1}^N (y(k) - \underbrace{(u(k), u(k-1), \dots, u(k-n_b)) \cdot \theta}_{\text{det. part is also } M(k; U, x_0, p)})^2$$

They often need a very high dimension  $n_b$  to obtain a reasonable fit. As a consequence ARX models are usually used instead.

**Equation Error Minimization:** Assume: i.i.d.  $\varepsilon(k)$  noise enters the input-output equation as additive disturbance

$$y(k) = h(p, u(k), \dots, u(k-n), y(k-1), \dots, y(k-n)) + \varepsilon(k)$$

for  $k = n+1, n+2$

if the i.i.d noise is gaussian, a maximum likelihood formulation to estimate the unknown parameter vector  $\theta = p$  is given:

$$\theta_{ML} = \min_{\theta} \sum_{k=n+1}^N (y(k) - h(p, u(k), \dots, y(k-1), \dots))^2$$

$u$  and  $k$  are known input and output measurements, and the algorithm minimises the so called **equation errors** or **prediction errors**.

This problem is also known as **Prediction error minimisation(PEM)** Such a problem is convex if  $p$  enters linearly in  $f$ , i.e. if the model is **linear-in-the-parameters (LIP)**

**PEM of LIP Models**

$$y(k) = \varphi(k)^T \theta + \varepsilon(k)$$

where  $\varphi = (\phi_1(\cdot), \dots, \phi_d(\cdot))^T$  are the regressor variables

considering this last expression, the prediction error minimisation(PEM) problem can be formulated as:

$$\min_{\theta} \sum_{k=1}^N (y(k) - \varphi(k)^T \theta)^2 = \|y_N - \Phi_N \theta\|_2^2$$

Which can be solved using LLS  $\theta^* = \Phi_N^+ y_N$

**Special Case: PEM of LIP-LTI Models with Equation Errors(ARX)** General ARX model equation

$$a_0 y(k) + \dots + a_{n_a} y(k - n_a) = b_0 u(k) + \dots + b_{n_b} u(k - n_b) + \varepsilon(k)$$

In order to have a determined estimation problem,  $a_0$  has to be fixed, otherwise the number of optimal solutions would be infinite. Therefore we usually fix  $a_0 = 1$  and use  $\theta = (a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b})^T$  as the parameter estimation vector. The regressor vector is given by:

$$\varphi = (-y(k-1), \dots, -y(k-n_a), u(k), \dots, u(k-n_b))^T$$

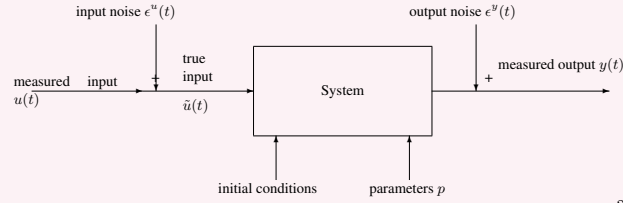
leading to the optimal solution provided by LLS:

$$y(k) = \varphi(k)^T \theta + \varepsilon(k)$$

## Pure Output Error (OE) Minimization

### Models with Input and Output Errors:

$$y(k) = M(k; U + \varepsilon_N^u, x_0, p) + \varepsilon^y(k)$$



Assume: i.i.d. gaussian noise on both input and output with variance  $\sigma_u^2$  for the input and  $\sigma_y^2$  for the output

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{k=1}^N \frac{1}{\sigma_y^2} (y(k) - M(k; U + \varepsilon_N^u, x_0, p))^2 + \frac{1}{\sigma_u^2} (\varepsilon_u(k))^2$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{k=1}^N \frac{1}{\sigma_y^2} (y(k) - M(k; \tilde{U}, x_0, p))^2 + \frac{1}{\sigma_u^2} (u(k) - \tilde{u}(k))^2$$

$$\tilde{U} := U + \varepsilon_N^u$$

## Fourier Transformation

### FT:

$$F\{F\}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt$$

$$G(j\omega) = \frac{Y_0}{U_0} e^{j\omega t}$$

### iFT:

$$f(t) = F^{-1}\{F\}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{-j\omega t} d\omega$$

### DFT:

$$U(m) := \sum_{k=0}^{N-1} u(t) e^{-j \frac{2\pi m k}{N}}$$

### iDFT:

$$u(n) := \sum_{k=0}^{N-1} U(k) e^{j \frac{2\pi k n}{N}}$$

## Useful frequency things

$$\omega = 2\pi f = \frac{2\pi}{T}, \quad f_s > 2f_{\max}, \quad T = N\Delta t = \frac{N}{f_s} = (f_{\text{res}})^{-1}$$

$$\sin(\varphi) = \frac{e^{j\varphi} - e^{-j\varphi}}{2}, \quad \cos(\varphi) = \frac{e^{j\varphi} + e^{-j\varphi}}{2}$$

## Aliasing and Leakage Errors

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

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

**Leakage Error:** Due to windowing.

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

## Crest Factor (ger. = Scheitelfaktor)

$$\text{Crest Factor} = \frac{u_{\max}}{u_{\text{rms}}} \quad \text{with: } u_{\text{rms}} := \sqrt{\frac{1}{T} \int_0^T u(t)^2 dt} \quad \left( = \sqrt{\frac{u_1^2 + u_2^2}{2}} \right)$$

symm. square wave

$$\text{and } u_{\max} := \max_{t \in [0, T]} |u(t)|$$

## Optimising Multisine for optimal crest factor

**Frequency:** Choose frequencies in logarithmic manner as multiples of the base frequency.  $\omega_{k+1}/\omega_k \approx 1.05/1.01/1.03$  (round to  $n \cdot \omega_{\text{base}}!$ )

**Phase:** To prevent high peaks (Crest Factor): random algorithm for phase shifts

## Multisine Identification Implementation procedure

**Window Length:** Integer multiple of sampling time:  $T = N \cdot \Delta t$

**Harmonics of base frequency:** Are contained in multisine

$$\omega_{\text{base}} = \frac{2\pi}{T}$$

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

**Experiment and Analysis:** (Step 2):

1. Insert Multisine periodically
2. Drop first Periods (till transients died out)
3. Record  $M$  Periods, with  $N$  samples, of input and output data
4. Average all windows and apply DFT (or vice versa)
5. 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:

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

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

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

### Block diagram from frequency sweeps:

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

## Block Diagramm

Magnitude  $\hat{=}$  Amplitude  $|G(j\omega)|$

Phase  $\hat{=}$  arg  $G(j\omega)$

## When to use what in frequency response

- In general: Multisines are a good approach
- Very fast system and transients  $\Rightarrow$  Frequency sweep
- **Very** slow system  $\Rightarrow$  Step response
- In the middle  $\Rightarrow$  Multisines

## Kalman Filter

### Valid for Discrete and Linear!

If recursive least squares:  $x_{k+1} = A_k \cdot x_k$

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

### 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 RLS  $\Rightarrow$  no  $W_{k-1}$

#### 2 Innovation update

$$P_{[k|k]} = (P_{[k|k-1]}^{-1} + C_k^T \cdot V^{-1} \cdot C_k)^{-1}$$

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

## Useful hints and practices

- Time invariant:  $t$  is only an argument of  $u(*)$  or  $y(*)$  i.e. **not**:

$$\dot{y}(t) = u(t) + \cos(t), \quad \dot{y}(t) = t \cdot u(t) + y(t)$$

- Linear: highest exponent of  $u$  and  $y$  is 1, i.e. **not**:

$$\dot{y}(t) = u(t)^2$$

- Affine: linear and has additive term **independent** from  $u$  and  $y$ , i.e.:

$$\dot{y}(t) = y(t) + \cos(t), \quad \dot{y}(t) = y(t) + C$$

- Sample Mean:  $\hat{\theta}_N(Y_N) = \frac{1}{N} \sum_{k=1}^N Y(k)$

- Sample Variance:  $S^2 = \frac{1}{N-1} \sum_{k=1}^N (Y(k) - \mathbb{E}\{Y_N\})^2$

### Magic Matrix for 8 point DFT:

$$W_8 = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 1.0 & 0.7 & 0.0 & -0.7 & -1.0 & -0.7 & 0.0 & 0.7 \\ 1.0 & 0.0 & -1.0 & 0.0 & 1.0 & 0.0 & -1.0 & 0.0 \\ 1.0 & -0.7 & 0.0 & 0.7 & -1.0 & 0.7 & 0.0 & -0.7 \\ 1.0 & -1.0 & 1.0 & -1.0 & 1.0 & -1.0 & 1.0 & -1.0 \\ 1.0 & -0.7 & 0.0 & 0.7 & -1.0 & 0.7 & 0.0 & -0.7 \\ 1.0 & 0.0 & -1.0 & 0.0 & 1.0 & 0.0 & -1.0 & 0.0 \\ 1.0 & 0.7 & 0.0 & -0.7 & -1.0 & -0.7 & 0.0 & 0.7 \end{bmatrix}$$