Modelling and System Identification – Final Exam

Prof. Dr. Moritz Diehl, IMTEK, Universität Freiburg, and OPTEC/ESAT-STADIUS, KU Leuven

March 13, 2014, 13:15-14:45, Freiburg, Georges-Koehler-Allee 101, SR 00-10/14

page	0	1	2	3	4	5	6	7	8	9
points on page (max)	4	11	10	5	8	5	5	11	4	0
points obtained										
intermediate sum										

Project Mark (if any)	Mark Part A	Mark Part B	Mark Part C		Mark Part D	Mark

Note:　　　　Klausur eingesehen am:　　　　Unterschrift des Prüfers:

Nachname:　　　　　　Vorname:　　　　　　Matrikelnummer:

Fach:　　　　　　Studiengang:　Bachelor ☐　Master ☐　Lehramt ☐　Sonstiges ☐

Please fill in your name above. For the multiple choice questions, tick exactly one box for the right answer. For the text questions, give a short formula or text answer just below the question in the space provided, and, if necessary, write on the backpage of the **same** sheet where the question appears, and add a comment "see backpage". Do not add extra pages (for fast correction, all pages will be separated for parallelization). The exam is a closed book exam, i.e. no books or other material are allowed. The exam is divided into four parts with equal weight, and the mark of one part can be replaced by the project mark. Some legal comments are found in a footnote (in German).[1]

Part A: Concepts from Probability, Statistics, and Estimators

1. Give the probability density function (PDF) $f(x)$ for a **normally distributed** random variable with mean μ and variance σ^2.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \left(\right).$$ Please add a **sketch**:

2	

2. What is the PDF of a random variable with **uniform distribution** on the interval $[a, b]$? For $x \in [a, b]$ it has the value:

$$f(x) = .$$ Please add a **sketch**:

2	

points on page: 4	

---

3. What is the PDF of an n-dimensional normally distributed variable z with zero mean and covariance matrix $P \succ 0$? The answer is $f_z(x) = \frac{1}{\sqrt{(2\pi)^n \det(P)}} \cdots$

(a) ☐ $e^{-\frac{1}{2}x^\top P x}$	(b) ☐ $e^{\frac{1}{2}x^\top P x}$	(c) ☐ $e^{-\frac{1}{2}x^\top P^{-1} x}$	(d) ☐ $e^{\frac{1}{2}x^\top P^{-1} x}$

4. A scalar random variable has the variance v. What is its standard deviation?

(a) ☐ v^{-1}	(b) ☐ $\sqrt{v}$	(c) ☐ v	(d) ☐ v^2

5. Regard a random variable $x \in \mathbb{R}^n$ with mean $c = \mathbb{E}\{x\}$, where $\mathbb{E}$ is the expectation operator. How is the covariance matrix $P \in \mathbb{R}^{n \times n}$ defined?

$$P =$$

1	

6. Regard random $x \in \mathbb{R}^n$ with zero mean and covariance P. What is the mean of the matrix valued variable $Z = xx^\top$?

(a) ☐ $P \operatorname{trace}(P)$	(b) ☐ P	(c) ☐ P^{-1}	(d) ☐ $P^{-1} \det(P)$

7. What does "i.i.d." stand for?

(a) ☐ infinite identically disturbed	(b) ☐ independent identically distributed
(c) ☐ independent identically disturbed	(d) ☐ infinite identically dependent

8. Given a sequence of i.i.d. scalar random variables $x(1), \ldots, x(N)$, each with mean μ and variance σ^2, what is the **variance** of the variable y_N defined by $y_N = \frac{1}{N} \sum_{k=1}^{N} x(k)$ (the arithmetic mean)?

(a) ☐ $\frac{\sigma^2}{N^2}$	(b) ☐ $\frac{\sigma}{N-1}$	(c) ☐ $\frac{\sigma^2}{N}$	(d) ☐ $\frac{\sigma}{N}$

9. What is the covariance matrix of $z = 3x + 2y$ if random variables $x, y \in \mathbb{R}^n$ are independent and have covariance matrices Σ_x, Σ_y ?

(a) ☐ $\Sigma_x + \frac{2}{3}\Sigma_y$	(b) ☐ $3\Sigma_x + 2\Sigma_y$	(c) ☐ $9\Sigma_x + 4\Sigma_y$	(d) ☐ $(3\Sigma_x^{-1} + 2\Sigma_y^{-1})^{-1}$

10. What is the mean μ_z and the covariance matrix Σ_z of the random variable $z = b + Ax + 2y$ where $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times m}$ are fixed and $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ are independent and have means μ_x, μ_y and covariance matrices Σ_x, Σ_y ?

$$\mu_z =$$

$$\Sigma_z =$$

2	

11. What is the minimizer x^* of the convex function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \alpha + \beta x + \frac{\gamma}{2}x^2$ with $\gamma > 0$?

$$\frac{\partial f}{\partial x}(x) = \qquad\qquad\qquad x^* =$$

1	

12. Given a sequence of numbers $y(1), \ldots, y(N)$, what is the minimizer θ^* of the function $f(\theta) = \sum_{k=1}^{N} (y(k) - \theta)^2$?

(a) ☐ $\frac{\sum_{k=1}^{N} y(k)}{N}$	(b) ☐ $\frac{1}{N} \sum_{k=1}^{N} y(k)^2$	(c) ☐ $\frac{1}{N^2} \sum_{k=1}^{N} y(k)^2$	(d) ☐ $\frac{N}{\sum_{k=1}^{N} y(k)}$

points on page: 11	

Part B: Linear Least Squares and Dynamic System Models

13. What is the minimizer of the convex function $f : \mathbb{R}^n \to \mathbb{R}$, $\boxed{f(x) = \|y - \Psi x\|_2^2}$ (with Ψ of rank n) ? The answer is $x^* = \ldots$

(a) ☐ $\quad -(\Psi^\top \Psi)^{-1} \Psi^\top y$ (b) ☐ $\quad (\Psi^\top \Psi)^{-1} \Psi^\top y$ (c) ☐ $\quad -(\Psi \Psi^\top)^{-1} \Psi^\top y$ (d) ☐ $\quad (\Psi \Psi^\top)^{-1} \Psi^\top y$

14. Given a prediction model $\boxed{y(k) = \theta_1 + \theta_2 e^{x(k)} + \epsilon(k)}$ with unknown parameter vector $\theta = (\theta_1, \theta_2)^\top$, and assuming i.i.d. noise $\epsilon(k)$ with zero mean, and given a sequence of N scalar input and output measurements $x(1), \ldots, x(N)$ and $y(1), \ldots, y(N)$, we want to compute the linear least squares (LLS) estimate $\hat{\theta}_N$ by minimizing the function $f(\theta) = \|y_N - \Psi_N \theta\|_2^2$. If $y_N = (y(1), \ldots, y(N))^\top$, how do we need to choose the matrix $\Psi_N \in \mathbb{R}^{N \times 2}$?

(a) ☐ $\begin{bmatrix} 1 & -x(1) \\ \vdots & \vdots \\ 1 & -x(N) \end{bmatrix}$ (b) ☐ $\begin{bmatrix} e^{x(1)} & 1 \\ \vdots & \vdots \\ e^{x(1)} & 1 \end{bmatrix}$ (c) ☐ $\begin{bmatrix} 1 & e^{x(1)} \\ \vdots & \vdots \\ 1 & e^{x(N)} \end{bmatrix}$ (d) ☐ $\begin{bmatrix} 1 & x(1) \\ \vdots & \vdots \\ 1 & x(N) \end{bmatrix}$

15. Using linear least squares, you want to fit a quadratic polynomial $y = a_0 + a_1 x + a_2 x^2$ to a set of measurements $x(1), \ldots, x(N)$ and $y(1), \ldots, y(N)$. What is the unknown parameter vector θ, and what objective function $f(\theta)$ do you need to minimize ?

$\theta = \qquad\qquad\qquad\qquad f(\theta) =$

$\boxed{2 \quad\quad}$

16. Given an autoregressive nonlinear dynamic system model $\boxed{y(k) = \theta_1 y(k-1)^2 + \theta_2 y(k-2) + \epsilon(k)}$ with i.i.d. noise $\epsilon(k)$ with zero mean and unknown parameter vector $\theta = (\theta_1, \theta_2)^\top$, and given a sequence of N output measurements $y(1), \ldots, y(N)$, we want to compute the estimate $\hat{\theta}_N$ by minimizing the prediction error function $f(\theta) = \sum_{k=3}^{N} \epsilon(k)^2 = \|y - \Psi \theta\|_2^2$. How do we need to choose the vector y and matrix Ψ ?

$$y = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} \quad , \qquad \Psi = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

$\boxed{2 \quad\quad}$

17. Which of the following dynamic models with inputs $u(t)$ and outputs $y(t)$ is **not** linear or affine?

(a) ☐ $t^2 \ddot{y}(t) = 2u(t)$ (b) ☐ $\dot{y}(t) = u(t) + \cos(t)$ (c) ☐ $\dot{y}(t) = y(t)u(t) + 1$ (d) ☐ $\ddot{y}(t) = t^2 u(t) + 1$

18. Which of the following dynamic models with inputs $u(t)$ and outputs $y(t)$ is **not** time varying ?

(a) ☐ $\dot{y}(t)^2 = u(t)$ (b) ☐ $\ddot{y}(t) = tu(t)^2$ (c) ☐ $\dot{y}(t) = u(t) + \sin(t)$ (d) ☐ $t^2 \ddot{y}(t) = u(t)$

19. Which of the following dynamic models with inputs $u(t)$ and outputs $y(t)$ is a linear time invariant (LTI) system ?

(a) ☐ $\ddot{y}(t) = u(t) + 4$ (b) ☐ $\ddot{y}(t) = \dot{u}(t)$ (c) ☐ $\dot{y}(t) = u(t) + \sin(t)$ (d) ☐ $\dot{y}(t)^3 = u(t)$

20. Which of the following models with input $u(k)$ and output $y(k)$ is **not** linear-in-the-parameters w.r.t. $\theta \in \mathbb{R}^2$?

(a) ☐ $\quad y(k) = \theta_1 \sin(y(k-1)) + \theta_2 u(k-1)$ (b) ☐ $\quad y(k) = \exp(\theta_1 y(k-1) + \theta_2 u(k-1))$

(c) ☐ $\quad y(k) = \theta_1 \exp(y(k-1)y(k-2)) + \theta_2 u(k-1)$ (d) ☐ $\quad y(k) = \theta_1 \exp(y(k-1)) + \theta_2 u(k-1)$

$\boxed{\text{points on page: } 10 \quad\quad}$

21. Which system is described by the transfer function $G(s) = \frac{s^2+2s}{s^2+1}$?

(a) ☐ $\ddot{y} + y = \ddot{u} + 2\dot{u}$	(b) ☐ $\ddot{y} + 2\dot{y} = \ddot{u} + u$	(c) ☐ $\ddot{y} + 2y = \dot{u} + u$	(d) ☐ $\ddot{y} + y = \dot{u} + 2u$

22. Compute the transfer function $G(s) = \frac{Y(s)}{U(s)}$ that describes the LTI-system $\ddot{x}(t) = -x(t) + u(t),\ y(t) = \dot{x}(t) + 2u(t)$.

$G(s) =$

1	

23. Sketch the step response $h(t)$ of the system with transfer function $G(s) = \frac{2}{1+Ts}$

1	

24. Modelling the temperature of water in a washbasin: regard a washbasin with two separate water taps, a hot and a cold one (as typical in the UK). The water temperatures of the inflowing water streams are constant and given by T_h and T_c. The incoming mass flows, which we can control with the taps, are given by $u_h(t)$ and $u_c(t)$ and are non-negative. We neglect heat losses to the environment, assume a constant specific heat capacity for water, and assume that the plug is closed, i.e. no water flows out of the basin. We have two states, the total water mass $m(t)$ in the washbasin, and its temperature $T(t)$. We assume that $m(t)$ is strictly positive from the beginning and that the warm and cold waters mix immediately in the basin. Using mass and energy conservation, derive an ordinary differential equation that describes the evolution of $m(t)$ and $T(t)$.

$\dot{m}(t) =$

$\dot{T}(t) =$

2	

points on page: 5	

13

Part C: Maximum Likelihood, Bayesian Estimation, Cramer-Rao, and Nonlinear Optimization

25. Maximum Likelihood Estimator (MLE): Assume a nominal model $h_i(\theta)$ and given measurements y_i, $i = 1, \ldots, N$. The PDF to obtain a measurement value y_i for a parameter value θ is known to be $\frac{1}{\sqrt{2\pi}\sigma_i} \exp(-\frac{1}{2\sigma_i^2}(y_i - h_i(\theta))^2)$ and measurement noises are uncorrelated. What function of θ does the MLE minimize in this case?

(a) ☐ $\sum_{i=1}^{N} \frac{1}{\sigma_i^2}(y_i - h_i(\theta))^2$	(b) ☐ $\|\sum_{i=1}^{N} \frac{y_i}{\sigma_i} - \sum_{i=1}^{N} \frac{1}{\sigma_i}h_i(\theta)\|$
(c) ☐ $\frac{1}{2}\|h(\theta) - y\|_2^2$	(d) ☐ $\sum_{i=1}^{N} \frac{1}{\sigma_i}|y_i - h_i(\theta)|$

26. Compute the (positive) minimizer x^* of the convex function $f : \mathbb{R}_{++} \to \mathbb{R}$, $\boxed{f(x) = \frac{V}{x^2} + \log x}$, and sketch it.

$$\frac{\partial f}{\partial x}(x) = \qquad\qquad , x^* = \qquad\qquad , \text{Sketch:}$$

$\boxed{2}$

27. Maximum Likelihood Estimator (MLE): we want to measure the height h of a room and have two devices (say, a laser distance measurement tool, and a yardstick). For each device, we assume a constant Gaussian distribution for the measurement errors, with zero mean and known, but different standard deviations σ_1 and σ_2. With each device, we have taken several measurements $y_1(1), \ldots, y_1(N_1)$ and $y_2(1), \ldots, y_2(N_2)$. We want to use all $N_1 + N_2$ measurements to get a MLE of the height h.

   (a) Write down the likelihood function $f(y|\theta)$, i.e. the probability density function of obtaining all measurements $y = (y_1(1), \ldots, y_1(N_1), y_2(1), \ldots, y_2(N_2))^\top$ for a given value of $\theta = h$.

   $$f(y|\theta) =$$

   $\boxed{1}$

   (b) Taking the negative logarithm (and neglecting constant terms in θ, if you like), write down the function which the MLE minimizes.

   $$-\log f(y|\theta) = \qquad\qquad\qquad\qquad\qquad\qquad + \text{const}$$

   $\boxed{1}$

   (c) What is the explicit formula for the MLE solution $\hat{\theta}_{\text{MLE}}$?

   $$\hat{\theta}_{\text{MLE}} =$$

   $\boxed{1}$

   (d) (*) Let us now assume that we do not know the standard deviations σ_1 and σ_2, i.e. that we have three unknown parameters $\theta = (h, \sigma_1, \sigma_2)^\top$. We still assume Gaussian and zero mean measurement errors. Using the formula from (a), and taking negative logarithms, what function would we have to minimize in a MLE to obtain estimates for h, σ_1, σ_2 together?

   $$-\log f(y|\theta) = -\log f(y|h, \sigma_1, \sigma_2) =$$

   $\boxed{1}$

   (e) (*) To see if the result from (d) makes sense, sketch $-\log f(y|h, \sigma_1, \sigma_2)$ as a function of σ_1 for fixed h and σ_2 (Tip: get inspiration from Question 26).

   $\boxed{1}$

   points on page: $\boxed{8}$

28. Which condition is necessary for a point x^* to be a local minimizer of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$?

(a) ☐ $f(x^*) = 0$	(b) ☐ $\nabla f(x^*) = 0$	(c) ☐ $\nabla^2 f(x^*) = 0$	(d) ☐ $\nabla^2 f(x^*) \succ 0$

29. Gauss-Newton: we want to solve the nonlinear least squares problem $\boxed{\min_\theta \dfrac{1}{2}\|h(\theta) - y\|_2^2}$, and have a current parameter guess $\theta_{[k]}$. If the Jacobian of h at $\theta_{[k]}$ is denoted by $J_{[k]} := \frac{\partial h}{\partial \theta}(\theta_{[k]}) \in \mathbb{R}^{N \times n}$, which linear least squares problem needs to be solved to obtain the next Gauss-Newton iterate $\theta_{[k+1]}$, and what is its explicit solution ? (Tip: the first order Taylor series of h at $\theta_{[k]}$ is given by $h(\theta_{[k]}) + J_{[k]}(\theta - \theta_{[k]})$,

$$\theta_{[k+1]} = \arg\min_\theta$$

| 1 | |

$$\theta_{[k+1]} = \theta_{[k]} +$$

| 1 | |

(*) If $\theta_{[k]}$ was already a (local) minimum of the nonlinear least squares objective function $f(\theta) = \frac{1}{2}\|h(\theta) - y\|_2^2$, prove that the Gauss-Newton method will stay there, i.e. that $\theta_{[k+1]} = \theta_{[k]}$.

| 1 | |

30. Bayesian estimation: we have a priori information about a parameter in form of a PDF $g(\theta)$ and know that the PDF to obtain measurements y given θ is given by $f(y, \theta)$. What function is minimized by a Bayesian estimator in this context?

(a) ☐ $g(\theta) + f(y, \theta)$	(b) ☐ $-g(\theta) + f(y, \theta)$
(c) ☐ $-\log g(\theta) - \log f(y, \theta)$	(d) ☐ $\log g(\theta) - \log f(y, \theta)$

points on page: 5 | |

15

31. Comparing MLE and Bayes: we regard a linear model $y = C\theta + \epsilon$ with unknown θ, fixed $C \in \mathbb{R}^{N \times n}$ and i.i.d. normally distributed noise $\epsilon \in \mathbb{R}^N$ with unit variance and zero mean (both known). Its negative log likelihood $\ell(\theta, y)$ is given by $\ell(\theta, y) = \frac{1}{2}\|C\theta - y\|_2^2 + \text{const}$. The unknown but true value of θ is θ_0.

(a) Give the Fisher Information Matrix $M = \mathbb{E}\{\nabla_\theta^2 \ell(\theta, y)\}|_{\theta=\theta_0}$:

$M =$

| 1 | |

(b) The MLE is given by $\hat{\theta}_{\mathrm{MLE}}(y) = (C^\top C)^{-1} C^\top y$. What is its covariance matrix P_{MLE} ?

$P_{\mathrm{MLE}} =$

| 1 | |

(c) Does the MLE estimator have the smallest possible covariance among all unbiased estimators? Justify your answer by citing the name and result of a theorem.

| 1 | |

(d) A Bayesian estimator uses in addition some prior knowledge on the parameters θ, e.g. that they are normally distributed around an a-priori guess θ_{prior}. The estimator will with some $\alpha > 0$ be given by

$$\hat{\theta}_{\mathrm{Bayes}}(y) = \arg\min_\theta \frac{1}{2}\|C\theta - y\|_2^2 + \frac{\alpha}{2}\|\theta - \theta_{\mathrm{prior}}\|_2^2 = (C^\top C + \alpha I)^{-1}(C^\top y + \alpha\theta_{\mathrm{prior}})$$

What is its covariance matrix P_{Bayes} ?

$P_{\mathrm{Bayes}} =$

| 1 | |

(e) (*) Find out if the covariance matrix of the MLE or the Bayesian estimator is smaller (in the matrix sense). Do your observations contradict the theorem cited in (c)? Justify.

| 1 | |

| points on page: 5 | |

16

Part D: Frequency Domain Identification, Recursive Identification, and the Kalman Filter

32. What quantity of a continuous time transfer function $G(s)$ shows the Bode amplitude diagram in doubly logarithmic scale?

(a) ☐ $|G(e^{j\omega})|$	(b) ☐ $\arg G(j\omega)$	(c) ☐ $G(j\omega)$	(d) ☐ $|G(j\omega)|$

33. Sketch (very roughly and without numbers) the Bode amplitude diagram of $G(s) = \frac{1}{1+s}$

1	

34. Which phase shows the Bode diagram of $G(s) = \frac{1}{1+T^2 s^2}$ for high frequencies?

(a) ☐ 90 deg	(b) ☐ 0 deg	(c) ☐ -90 deg	(d) ☐ -180 deg

35. Which phase shows the Bode diagram of $G(s) = \frac{1}{1+T^2 s^2}$ for very low frequencies?

(a) ☐ 90 deg	(b) ☐ 0 deg	(c) ☐ -90 deg	(d) ☐ -180 deg

36. Which slope has the Bode amplitude diagram of $G(s) = \frac{1}{1+T^2 s^2}$ for high frequencies?

(a) ☐ 20 dB/decade	(b) ☐ 0 dB/decade	(c) ☐ -20 dB/decade	(d) ☐ -40 dB/decade

37. Regard a sampled periodic signal with total period duration T (in seconds) that contains N samples (i.e. $\Delta t = \frac{T}{N}$). We assume N to be an even number. List all frequencies (in Hz) that can be present in this signal. What is its lowest, what its highest frequency, and how many different frequencies can it contain?

3	

38. Describe in words the procedure that one uses when identifying an LTI-SISO system with multi-sine excitation.

3	

points on page: 11	

39. Regard a recursive algorithm to compute $\hat{\theta}_n = \arg\min_\theta \sum_{k=1}^n (y(k) - \theta)^2$ for a sequence of incoming measurements $y(k) \in \mathbb{R}$, starting with $\hat{\theta}_1 = y(1)$. Which recursion formula describes its solution for $n > 1$?

(a) ☐	$\hat{\theta}_n = \hat{\theta}_{n-1} + \frac{1}{n}(y(n) - \hat{\theta}_{n-1})$	(b) ☐	$\hat{\theta}_n = \hat{\theta}_{n-1} - \frac{1}{n+1}(y(n) - \hat{\theta}_{n-1})$
(c) ☐	$\hat{\theta}_n = \hat{\theta}_{n-1} - \frac{1}{n}(y(n) - \hat{\theta}_{n-1})$	(d) ☐	$\hat{\theta}_n = \hat{\theta}_{n-1} + \frac{1}{n+1}(y(n) - \hat{\theta}_{n-1})$

40. Regard a recursive algorithm to compute $\hat{\theta}_n = \arg\min_\theta \epsilon\|\theta\|_2^2 + \sum_{k=1}^n \|y(k) - C\theta\|^2$ for a sequence of incoming measurements $y(k) \in \mathbb{R}^m$, starting with $\hat{\theta}_0 = 0$ and $Q_0 = \epsilon\mathbb{I}$. If the recursion for Q_n is given by $Q_n = Q_{n-1} + C^\top C$ for $n > 1$, which recursion formula holds for $\hat{\theta}_n$?

$$\hat{\theta}_n = \hat{\theta}_{n-1} +$$

2	

41. Regard a Kalman filter to estimate the state of a discrete time linear time invariant system $x(k+1) = Ax(k) + w(k)$ and $y(k) = Cx(k) + v(k)$, where we assume i.i.d. white noises w and v with variances Σ_w and Σ_v. These two matrices are tuning parameters, and we often choose them as diagonal matrices with larger or smaller entries. What choice of entries in matrices Σ_w, Σ_v should we use if we have not very accurate measurements of $y(k)$, but trust our linear model very well ?

(a) ☐	large Σ_w, small Σ_v	(b) ☐	large Σ_w, large Σ_v
(c) ☐	small Σ_w, small Σ_v	(d) ☐	small Σ_w, large Σ_v

points on page: 4	