Modeling and System Identification – Microexam 2

Prof. Dr. Moritz Diehl, IMTEK, Universität Freiburg

January 24, 2020, 10:00-12:00, Freiburg

Surname: Name: Matriculation number:

Study: Programm: Bachelor ☐ Master ☐

Please fill in your name above and tick exactly **ONE** box for the right answer of each question below.
You can get a maximum of **10 points** on this microexam.

1. We would like to know the unknown probability θ that a phone breaks when it is dropped. We assume that the phone thrown onto the ground either breaks or has no damage. In an experiment we have dropped 100 smartphones and obtained 19 broken smartphones. What is the negative log likelihood function $f(\theta)$ that we need to minimize in order to obtain the maximum likelihood (ML) estimate of θ ?

(a) ☐ $-\log(81\theta) - \log(19(1-\theta))$	(b) ☐ $-81\log\theta - 19\log(1-\theta)$
(c) ☒ $-19\log\theta - 81\log(1-\theta)$	(d) ☐ $\log(19\theta) + \log(81(1-\theta))$

2. You are given a pendulum which is by nature a nonlinear system and can be modeled by $y(t) = \theta_1\cos(\theta_2 t + \theta_3)$, where $y(t)$ are the measurements. Which of the following algorithms should you use to estimate the parameters θ?

(a) ☐ Weighted Least Squares (WLS)	(b) ☐ Linear Least Squares (LLS)
(c) ☐ Recursive Least Squares (RLS)	(d) ☒ Nonlinear Least Squares (NLS)

3. Consider a model that is linear in parameter (LIP). Which of the following algorithms could you use to estimate the parameters without running into memory problems or high computational costs for a continuous and infinite flow of measurement data?

(a) ☐ LLS	(b) ☐ ML	(c) ☐ WLS	(d) ☒ RLS

4. You are asked to give a computationally efficient approximation of the covariance of the estimate computed in the previous question $\Sigma_{\hat{\theta}}$. The model is given as $y_N = \Phi_N\theta + \epsilon_N$ with $\epsilon_N \sim \mathcal{N}(0, \Sigma_\epsilon)$, $Q_N = \Phi_N^\top \Phi_N$ and $L(\theta, y_N)$ is the negative log likelihood function. The covariance matrix can be approximated by $\Sigma_{\hat{\theta}} \approx \dots$

(a) ☐ $(\Phi_N^\top \Sigma_{\epsilon_N} \Phi_N)^{-1}$	(b) ☒ Q_N^{-1}	(c) ☐ $\Phi_N^+ \Sigma_{\epsilon_N}^{-1} \Phi_N^{+\top}$	(d) ☐ $(\nabla_\theta^2 L^2(\theta, y_N))^{-1}$

5. Let θ_R denote the *regularized* LLS estimator using L_2 regularization. Which of the following is **NOT** true?

(a) ☐ θ_R can be computed analytically.	(b) ☐ θ_R incorporates prior knowledge about θ.
(c) ☒ θ_R is asymptotically biased.	(d) ☐ θ_R is biased.

6. We use the Gauss-Newton (GN) algorithm to solve a nonlinear estimation problem. Which of the following statements is **NOT** true *in general*?

(a) ☐ The idea of GN is to linearize the residual function.	(b) ☐ GN uses a Hessian approximation.
(c) ☒ GN finds the global minimizer of the objective function.	(d) ☐ The inverse of the GN Hessian approximates Σ_θ.

7. 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) ☐ $y(k) = \theta_1 u(k)^4 + \theta_2\exp(u(k))$	(b) ☒ $y(k) = \theta_1\exp(\theta_2 u(k))$
(c) ☐ $y(k) = \theta_1\sqrt{u(k)} + \theta_2 u(k)$	(d) ☐ $y(k) = y(k-1)\cdot(\theta_1 + \theta_2 u(k))$

8. Given is a set of measurements $y_N = [y(1), y(2), \dots, y(N)]^\top$ and the linear model $y_N = \Phi_N\theta + \epsilon_N$ with i.i.d. Gaussian noise ϵ_N, where $\Phi_N = [\varphi(1), \varphi(2), \dots, \varphi(N)]^\top$. Using an RLS algorithm where Q_N is updated recursively with $Q_{N+1} = Q_N + \varphi(N+1)\varphi(N+1)^\top$, which of the following minimisation problems is solved at each iteration step to estimate the parameter $\hat{\theta}(N+1)$ after $N+1$ measurements? $\hat{\theta}(N+1) = \arg\min_\theta \frac{1}{2}\dots$

(a) ☐ $\lVert\theta - \hat{\theta}(N)\rVert_{Q_N}^2 + \lVert y(N) - \varphi(N)^\top\theta\rVert_2^2$	(b) ☐ $\lVert y_N - \Phi_N\cdot\theta\rVert_{Q_N}^2$
(c) ☐ $\lVert\theta - \hat{\theta}(N)\rVert_2^2 + \lVert y(N+1) - \varphi(N+1)^\top\theta\rVert_2^2$	(d) ☒ $\lVert y_{N+1} - \Phi_{N+1}\cdot\theta\rVert_2^2$

Points on page (max. 8)

9. Suppose you are given the Fisher information matrix $M = \int_{y_N} \nabla^2_\theta L(\theta_0, y_N) \cdot p(y_N|\theta_0) \mathrm{d}y_N$ of an unbiased estimator, what is the relation with the covariance matrix $\Sigma_{\hat{\theta}}$ of your estimate $\hat{\theta}$? We assume that the true value θ_0, as well as the PDF $p_Y(y)$ of the measurements is known.
$\Sigma_{\hat{\theta}} \succeq M^{-1}$

10. Give the name of the theorem that provides us with the above result.
Cramer-Rao Inequality

11. Given the probability density function $p_X(x) = \theta e^{-\theta x}$ for $x \geq 0$ (and 0 otherwise) with unknown θ and positive i.i.d. measurements $y_N = [y(1), y(2), \dots, y(N)]^T$ that are assumed to follow the above distribution, what is the minimisation problem you need to solve for a ML-estimate of θ? The problem is: $\min_\theta \dots$?

(a) [ ] $\|y(k) - \theta e^{-\theta}\|_2^2$	(b) [ ] $-\log \sum_{k=1}^N \theta e^{-\theta y(k)}$
(c) [ ] $\|\theta e^{-\theta y(k)}\|_2^2$	(d) [x] $-N \log(\theta) + \theta \sum_{k=1}^N y(k)$

12. For the problem in the previous question, what is a lower bound on the covariance $\Sigma_{\hat{\theta}}$ for any unbiased estimator $\hat{\theta}(y_N)$, assuming that θ_0 is the true value? $\Sigma_{\hat{\theta}} \succeq \dots$

(a) [ ] N/θ^2	(b) [x] θ_0^2/N
(c) [ ] $\int_{y_N} N\theta_0^{N-2} \exp[-\theta \sum_k y_k] \mathrm{d}y_N$	(d) [ ] $\left(\int_{y_N} N\theta^{N-2} \exp[-\theta \sum_k y_k] \mathrm{d}y_N\right)^{-1}$

13. Which of the following models is time invariant?

(a) [ ] $t \cdot \ddot{y}(t) = \sqrt{u(t)}$	(b) [x] $\dot{y}(t) = u(t)^2 + 1$	(c) [ ] $\ddot{y}(t)^2 = u(t)^t + e^{u(t)}$	(d) [ ] $\dot{y}(t) = t^4 - u(t)$

14. In L_1 estimation the measurement errors are assumed to follow a ... distribution and it is generally speaking more ... to outliers compared to L_2 estimation.

(a) [ ] Laplace, sensitive	(b) [ ] Gaussian, robust	(c) [ ] Gaussian, sensitive	(d) [x] Laplace, robust

15. The PDF of a random variable Y is given by $p_Y(y) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{\|y-\theta\|_2^2}{2}\right)$, with unknown $\theta \in \mathbb{R}$. We obtained three measurements, $y(1) = 2$, $y(2) = 2$, and $y(3) = 5$. What is the minimizer θ^* of the negative log-likelihood function ?

(a) [ ] 5	(b) [x] 3	(c) [ ] 4	(d) [ ] 2

16. Which of the following statements is **NOT** correct. Recursive Least Squares (RLS):

(a) [ ] implicitly assumes that there is only i.i.d. and Gaussian measurement noise	(b) [ ] computes an estimation with a computational cost independent of the number of past measurements
(c) [x] can be used as an alternative to Maximum Likelihood Estimation	(d) [ ] can use prior knowledge on the estimated parameter θ

17. Which of the following model equations describes a FIR system with input u and output y? $y(k+1) = \dots$

(a) [ ] $u(k) + e^{i\pi \cdot k}$	(b) [x] $u(k) - \pi^2 u(k-2)$	(c) [ ] $\frac{1}{2}u(k+1) + y(k)$	(d) [ ] $u(k) \cdot y(k)$

18. In practice, how do we estimate the covariance matrix of a parameter estimate θ^* with the objective $f(\theta) = \|R(\theta)\|_2^2$ and $R(\theta)$ being a possibly nonlinear residual function with Jacobian $J(\theta) = \frac{\partial R(\theta)}{\partial \theta} \in \mathbb{R}^{N \times d}$? $\Sigma_{\hat{\theta}} = \frac{\|R(\theta^*)\|_2^2}{N-d} \cdot (\dots)$

(a) [ ] $R(\theta^*)R(\theta^*)^\top$	(b) [ ] $\nabla f(\theta^*)^\top \nabla f(\theta^*)$	(c) [ ] $J(\theta^*)J(\theta^*)^\top$	(d) [x] $(J(\theta^*)^\top J(\theta^*))^{-1}$

19. You want to estimate the paramters θ of a linear model $y_N = \Phi\theta$. For this you minimze the objective $f(\theta) = \|y_N - \Phi\theta\|_2^2$, but unfortunately your minimization problem $\min_\theta f(\theta)$ turns out to be ill-posed. Which of the following statements is **NOT** true:

(a) [x] Regularized LLS can find a unique minimizer of $f(\theta)$	(b) [ ] the set of solutions is $\theta^* = \{\theta \| \nabla f(\theta) = 0\}$
(c) [ ] the set of solutions is $\theta^* = \{\theta \| \Phi^\top \Phi\theta - \Phi^\top y = 0\}$	(d) [ ] $\Phi^\top \Phi$ is not invertible

20. Suppose you are fitting a model to 500 **noisy** measurements using MAP. Afterwards you compute the R-Squared value of the fit. Which of the following values suggests a meaningful fit?

(a) [ ] 3.23	(b) [ ] -1	(c) [x] 0.86	(d) [ ] 1

Points on page (max. 12)	