1. DERIVATIVES OF VECTORS & MATRICES
5. MICROEXAM 1 LAST YEAR
3. SINGULAR VALUE DECOMPOSITION
4. (LINEAR) REGRESSION VECTORS : LINEARLY INDEPENDENT?
2. POSITIVE DEFINITENESS
   EXAMPLE FOR COVAR. MATRIX COMP.
   HOW TO COMPUTE EXPECT. OF MULTI-DIM. VARIABLE

$$f(x) = \frac{1}{2} x^T B x$$
$$\nabla f(x) = \frac{1}{2}(B + B^T)x$$
$$= Bx$$

(1) → $f(x) = c^T x$ $\quad \nabla f(x) = c \quad \nabla^2 f(x) = 0$

→ $f(x) = x^T A x$

$$\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_{j,k} x_j A_{jk} x_k = \frac{\partial}{\partial x_i} \sum_j x_j \sum_k A_{jk} x_k = \sum_j x_j \left(\frac{\partial}{\partial x_i} \sum_k A_{jk} x_k\right)$$
$$+ \sum_k x_k \left(\frac{\partial}{\partial x_i} \sum_j A_{jk} x_j\right)$$

$$= \sum_j x_j A_{ji} + \sum_k x_k A_{ik}$$

$$= (A^T x + A x)_i$$

$$\nabla f(x) = (A^T + A)x$$

$$\nabla^2 f(x) = \left(\frac{\partial \nabla f}{\partial x_1} \middle| \frac{\partial \nabla f}{\partial x_2} \middle| \dots \middle| \right) = (A^T + A)$$

$$f(x) = \|Ax + c\|_2^2 = (Ax + c)^T (Ax + c) = x^T A^T A x + x^T A^T c + c^T A x + c^T c$$

$S = S^T$ $\quad \nabla f(x) = 2A^T A x + A^T c \quad \nabla^2 f = 2A^T A$

EUCL. NORM. , POS. DEF. , SVD

$\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum x_i^2}$ $\quad x \in \mathbb{R}^n$

$\|x\|_2^2 = x^T x = \sum x_i^2$

$\arg\min_x \frac{1}{2}\|Ax - b\|_2^2$

$(B \succ 0 \Rightarrow 0 \preceq 0)$

$A^T A x - A^T b = 0 \iff x = (A^T A)^{-1} A^T b$

$B \succeq 0$

B is (SEMI) POS. DEF.

$:\iff (B = B^T)$ $\forall z \in \mathbb{R}^n$ ~~$(z \neq 0)$~~ $z^T B z \geq 0 \iff$ EV B ARE POSITIVE OR ~~ZERO~~

(RECALL. IF $B = B^T$, THEN ALL EVAL OF B ARE REAL, WITH ORTHOGONAL EVEC

$B = T^{-1} D T \qquad T^T T = I$

$= T^T D T$

$B \in \mathbb{R}^{1\times 1}$ $\quad B = \beta$

$z \beta z > 0$ IF $z \neq 0$

$\iff \beta > 0$

SING. VAL DECOMPOSITION

(STEP TOWARDS THE MOORE-PENROSE PSEUDOINVERSE)

THM FOR ANY $A \in \mathbb{R}^{m \times n}$ EXIST U, S, V

WITH $\boxed{A = U S V^T}$ WITH $U \in \mathbb{R}^{m \times m}$, $U^T U = I$

$V \in \mathbb{R}^{n \times n}$, $V^T V = I$

$S \in \mathbb{R}^{m \times n}$, DIAGONAL, ($I$ = UNIT MATRIX)

WITH $\boxed{\sigma_1 \geq \sigma_2 \geq \dots \sigma_r \geq 0}$ $\in \mathbb{R}$

"SINGULAR VALUES"

$$S = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \\ \hline & & 0 & & \end{pmatrix}$$

RANK$(A) = r$

$\longrightarrow$ MOORE PENROSE PSEUDO INVERSE

$$A^+ = V S^+ U^T$$

$$\boxed{S^+ := \begin{pmatrix} \sigma_1^{-1} & & & & \\ & \ddots & & & \\ & & \sigma_r^{-1} & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix} \; 0 }$$

$\boxed{S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}, \; S^+ = \begin{pmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_n^{-1} \end{pmatrix} = S^{-1}}$

IF A is invertible, $A \in \mathbb{R}^{n \times n}$, $r = n$ $\quad A^+ = V S^+ U^T = V S^{-1} U^T = A^{-1}$

IF $(A^T A)$ is INVERTIBLE, $r = n$ $\quad \boxed{A^+ = (A^T A)^{-1} A^T}$

$\boxed{\text{pinv}(A)}$

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \qquad U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

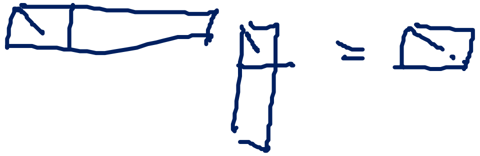
$$S^+ = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix} \qquad S^{-1}$$ DOES NOT EXIST

$$A^+ = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}$$

$$A^+ b = \arg\min \|x\|_2^2 \text{ s.t. } A^T A x - A^T b = 0$$

MINIMUM NORM SOLUTION

NICE CASE $A \in \mathbb{R}^{N \times d}$ $\quad N > d$ $\quad A^T A$ INV. $\iff \operatorname{rank}(A) = d$



LIN. INDEP. OF REGRESSION

LINEAR INDEP. : LINEAR ALGEBRA

INDEP. VARIABLES : PROBABILITY.

$\frac{1}{2} \|\Phi_N \theta - Y_N\|_2^2$: COLUMNS OF Φ_N LIN. INDEP.

$\iff \operatorname{rank} \Phi_N = d$

$\Phi_N = U S V^T \qquad S = \begin{pmatrix} \sigma_1 \dots \sigma_d \\ 0 \end{pmatrix}$

Modelling and System Identification – Microexam 1

Prof. Dr. Moritz Diehl, IMTEK, Universität Freiburg

November 18, 2014, 8:15-9:15, Freiburg

Surname: Name: Matriculation number:

Study: Studiengang: Bachelor ☐ Master ☐

Please fill in your name above and tick exactly one box for the right answer of each question below.

1. What is the probability density function (PDF) $p_X(x)$ for a normally distributed random variable X with mean μ and variance σ^2? The answer is $p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \ldots$

(a) ☐ $e^{\frac{(x-\mu)^2}{2\sigma}}$	(b) ☐ $e^{-\frac{(x-\mu)^2}{2\sigma}}$	(c) ☐ $e^{\frac{(x-\mu)^2}{2\sigma^2}}$	(d) ☒ $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

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

(a) ☐ $p_Z(x) = (d-c)$	(b) ☐ $p_Z(x) = (c-d)^2$	(c) ☐ $p_Z(x) = \frac{x}{\sqrt{d-c}}$	(d) ☒ $p_Z(x) = \frac{1}{d-c}$

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

(a) ☐ $e^{-\frac{1}{2}x^T\Sigma x}$	(b) ☒ $e^{-\frac{1}{2}x^T\Sigma^{-1}x}$	(c) ☐ $e^{\frac{1}{2}x^T\Sigma x}$	(d) ☐ $e^{\frac{1}{2}x^T\Sigma^{-1}x}$

4. Regard a random variable $X \in \mathbb{R}^n$ with mean $d \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. For a fixed $a \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$, regard another random variable Y defined by $Y = a + AX$. What is the mean μ_Y of Y ?

(a) ☐ $Y = a + AX$	(b) ☒ $a + Ad$	(c) ☐ AXX^TA^T	(d) ☐ $a^TAd + d^T\Sigma d$

5. Above in Question ??, what is the covariance matrix of Y?

(a) ☐ $d^T\Sigma d$	(b) ☒ $A\Sigma A^T$	(c) ☐ $A^T\Sigma^{-1}A$	(d) ☐ $A\Sigma^{-1}A^T$

6. Above in Question ??, which statement is true? Cov(Y) =

(a) ☐ $Y^\top Y - \mu_Y^\top \mu_Y$	(b) ☒ $\mathbb{E}\{YY^\top\} - \mu_Y\mu_Y^\top$ = Cov(Y)
(c) ☐ $YY^\top - \mu_Y\mu_Y^\top$	(d) ☐ $\mathbb{E}\{Y^\top Y\} - \mu_Y^\top\mu_Y$

7. (*) Above in Question ??, what is the mean of the matrix valued random variable $Z = YY^T$?

(a) ☐ $(a + Ad)(a + Ad)^T$	(b) ☐ $aa^T + Add^TA^T + A\Sigma A^T$
(c) ☒ $(a + Ad)(a + Ad)^T + A\Sigma A^T$	(d) ☐ $aa^T + Add^TA^T$

$\mathbb{E}(Y) = \mathbb{E}(a + AX)$
$= a + A\,\mathbb{E}(X)$
$= a + Ad$

$\mathbb{E}\{(Y-\mu_Y)(Y-\mu_Y)^T\}$

$Y = a + A\cdot X$

$Cov(Y) = A \cdot cov(X) \cdot A^T$

$\mathbb{E}\{YY^T\} = Cov(Y) + \mu_Y\mu_Y^T$
$= A\Sigma A^T$
$+ (a+Ad)(a+Ad)^T$

8. A scalar random variable has the standard deviation y. What is its variance?

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

9. A scalar random variable has the variance w. What is its standard deviation?

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

10. Regard a random variable $\beta \in \mathbb{R}$ with zero mean and variance σ^2. What is the mean of the random variable $z = \beta^2$?

(a) ☐ $\beta + \sigma^2$	(b) ☐ σ	(c) ☒ σ^2	(d) ☐ $\beta + \sigma$

11. (*) Regard a random variable $X \in \mathbb{R}^n$ with zero mean and covariance matrix Σ. What is the mean of $Z = X^T X$?

(a) ☐ $\|\Sigma\|_F^2$	(b) ☐ $\det(\Sigma)$	(c) ☐ $\|\Sigma\|_2^2$	(d) ☒ $\text{trace}(\Sigma)$

points on page: 11	

$$\mathbb{E}(\beta^2) - 0.0 = \text{cov}(\beta)$$

$$Z = \sum_i X_i^2$$

$$\mathbb{E}(Z) = \sum_i \mathbb{E}(X_i^2) = \sum_{i=1}^{n} \sigma_i^2$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & * \\ & \ddots & \\ * & & \sigma_n^2 \end{pmatrix}$$

12. What is the minimizer x^* of the convex function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x - 2x$?

(a) ☐ $x^* = -1$	(b) ☐ $x^* = 1$	(c) ☒ $x^* = \log_e(2)$	(d) ☐ $x^* = 0$

13. What is the minimizer x^* of the convex function $f : \mathbb{R} \to \mathbb{R}$, $\boxed{f(x) = \alpha + \beta x + \frac{1}{2}\gamma x^2}$ with $\gamma > 0$?

(a) ☐ $x^* = \frac{2\beta}{\alpha}$	(b) ☒ $x^* = -\frac{\beta}{\gamma}$	(c) ☐ $x^* = -\frac{\beta}{2\gamma}$	(d) ☐ $x^* = -\frac{\beta}{\alpha}$

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

(a) ☐ $-(\Phi\Phi^T)^{-1}\Phi^T y$	(b) ☐ $-(\Phi^T\Phi)^{-1}\Phi^T y$	(c) ☒ $(\Phi^T\Phi)^{-1}\Phi^T y$	(d) ☐ $(\Phi\Phi^T)^{-1}\Phi^T y$

15. What is the minimizer of the function $f : \mathbb{R}^n \to \mathbb{R}$, $\boxed{f(x) = \|b + B^T x\|_2^2}$ (with B^T of rank n)? The answer is $x^* = \ldots$

(a) ☐ $(BB^T)^{-1}B^T b$	(b) ☒ $-(BB^T)^{-1}Bb$	(c) ☐ $(B^T B)^{-1}B^T b$	(d) ☐ $-(B^T B)^{-1}B^T b$

16. For a matrix $\Phi \in \mathbb{R}^{N \times d}$ with rank d (and $N \geq d$), what is its pseudo-inverse Φ^+ ?

(a) ☐ $(\Phi\Phi^T)^{-1}\Phi^T$	(b) ☐ $(\Phi\Phi^T)^{-1}\Phi$	(c) ☒ $(\Phi^T\Phi)^{-1}\Phi^T$	(d) ☐ $(\Phi^T\Phi)^{-1}\Phi$

$\frac{\partial f}{\partial x} = e^x - 2 \overset{!}{=} 0$

$e^x = 2 \iff \log_e 2 = x$

$\frac{\partial f}{\partial x} = \beta + \gamma \cdot x \overset{!}{=} 0 \iff x = -\frac{\beta}{\gamma}$

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

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

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

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

19. Given a sequence of i.i.d. scalar random variables $X(1), \dots, X(N)$, each with mean μ and variance σ^2, what is the expected value of their arithmetic mean, i.e. of the random variable Y defined by $Y = \frac{1}{N}\sum_{k=1}^{N} X(k)$?

(a) ☒ μ	(b) ☐ $\frac{\mu}{N}$	(c) ☐ $\frac{\mu}{\sigma^2}$	(d) ☐ $\frac{\mu}{\sqrt{\sigma^2}}$

20. In Question ??, what is the variance of the variable Y ?

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

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

(a) ☐ $\begin{bmatrix} x(1)^2 & 1 \\ \vdots & \vdots \\ x(1)^2 & 1 \end{bmatrix}$	(b) ☒ $\begin{bmatrix} 1 & x(1)^2 \\ \vdots & \vdots \\ 1 & x(N)^2 \end{bmatrix}$	(c) ☐ $\begin{bmatrix} 1 & x(1) \\ \vdots & \vdots \\ 1 & x(N) \end{bmatrix}$	(d) ☐ $\begin{bmatrix} 1 & -x(1) \\ \vdots & \vdots \\ 1 & -x(N) \end{bmatrix}$

$Y = A \cdot X$

$A = (\frac{1}{N} | \dots | \frac{1}{N})$

$\mathbb{E}(Y) = \frac{1}{N} \sum_k E X(k) = \mu$

$\mathrm{cov}(Y) = A \cdot \mathrm{cov}(X) \cdot A^T = \frac{1}{N^2}\left(\sum_{k=1}^{N} \sigma^2\right) = \frac{\sigma^2 \cdot N}{N^2} = \frac{\sigma^2}{N}$