EXAM COVER SHEET

Name of the Exam:	Numerical Optimization		
Time and Date:	22.04.2025, 8 O 'clock	Institute:	Numerical Mathematics
Duration:	120 Minutes	Examiner:	Prof. Dr. Karsten Urban

To be completed by the exam participant:

First name:		Course:	
Last name:		Degree:	

Student ID / Matriculation no.:

Date and signature of the exam participant

I hereby declare that I am capable of taking the exam.

Should I not be listed on the list of registered students due to lack of registration through the University Portal or through the Student Administration Office, I hereby acknowledge that this exam will not be given any grade.

Authorized Auxiliaries:

- Non-native speakers may use a dictionary.
- One handwritten DIN A4 sheet.

Further information for the exam:

- 50 points equates to 100 % !
- *Good Luck with the Exam!*

Please leave this field blank for the barcode!

To be completed by the examiner:

Exercise	1	2	3	4	5	6	$\sum$
Maximum number of points	10	12	16	10	8	5	61
Achieved points							
Corrector							

Grade:

Signature of the examiner

Last Name, First Name: ______________________________

Task 1 (Numerical Optimization: General Knowledge) (10 Points)

Decide whether the following statements are *true* or *false*. Mark your answers **clearly**. Each correct answer gives one point.

(a) The semi-smooth Newton method solves the problem $f(x) = 0$, where f is semismooth. — true ◯ false ◯

(b) The introduction of a slack variable can be used to avoid nonlinear inequality constraints. — true ◯ false ◯

(c) There is always a saddle point (x^*, λ^*, μ^*) of the Lagrange function $\mathcal{L}(x, \lambda, \mu)$ of the problem $\min\limits_{x \in \mathcal{F} \subset \mathbb{R}^n} f(x)$, with $\mathcal{F}$ being the feasible set, with the property, that x^* is not a minimizer. — true ◯ false ◯

(d) The Newton-SQP method is second order convergent for all initial values, due to the fact that the Newton-SQP method is based upon Newton's method. — true ◯ false ◯

(e) The active set method transforms a quadratic optimization problem with linear equality and inequality constraints into a sequence of linear optimization problems with only equality constraints. — true ◯ false ◯

(f) The curvature condition of Wolfe's conditions ensures that the step size is not too large in order to guarantee global convergence in case of large gradients. — true ◯ false ◯

(g) Using a descent direction in every iteration is a crucial requirement in the Zoutendijk theorem. — true ◯ false ◯

(h) The set $\{x \in \mathbb{R}^2 : 0 \leq x_2 \leq x_1,\ x_1 \geq 0\}$ is a cone. — true ◯ false ◯

(i) The tangential cone of the set $\{x \in \mathbb{R}^2 : 0 \leq x_2 \leq x_1,\ x_1 \geq 1\}$ in $(1, 0)^T \in \mathbb{R}^2$ is given by $[0, \infty) \times [0, \infty)$. — true ◯ false ◯

(j) A minimization problem with equality constraints can always be rewritten as a minimization problem with only inequality constraints. — true ◯ false ◯

Lösung:

(a) true
(b) true
(c) false
(d) false
(e) false
(f) false
(g) true
(h) true
(i) true
(j) true

Last Name, First Name: ____________________

Task 2 (Application of Numerical Optimization Methods) (3+3+3+3 = 12 Points)

In this exercise you are given situations (already mathematically modeled) with different requirements for which you should decide whether the given numerical method is the *right choice* or *not*. Justify your answer!
Remark: Zero points are awarded for the correct decision, with a wrong justification!

a)
- Situation: Researchers measured data $(a_i, t_i) \subset \mathbb{R} \times [0,1]$, $1 \leq i \leq 10^3$ during a medical study. For each patient i with age a_i the tolerability t_i of a new vaccine has been recorded. It seems to be a nonlinear relation (with five parameters) between the age a_i and the tolerability t_i. To visualize this relation, we perform a nonlinear regression, which is not sensitive to data outliers. We use the convex residual function $F : \mathbb{R}^5 \to [0,1]^{10^3}$.
- Problem: Solve $\min\limits_{x \in \mathbb{R}^5} \|F(x)\|_1$, where $\|y\|_1 := \sum\limits_{i=1}^{n} |y_i|,\ y \in \mathbb{R}^n$;
- Requirements: None
- Numerical method: Levenberg-Marquardt Method
- Decision:
- Justification:

b)
- Situation: In molecular modeling one tries to find a stable arrangement of the atoms by minimizing the system energy E. Unfortunately, there are usually many arrangements which are local minima for E. Thus, for a carbon chain consisting of 10^5 carbon atoms with positions $x_i \in \mathbb{R}^3$, $i = 1, \ldots, 10^5$ we seek as many stable arrangements as possible. The derivatives of E can be computed exactly.
- Problem: Solve $\min\limits_{x_1,\ldots,x_{10^5} \in \mathbb{R}^3} E(x_1, \ldots, x_{10^5})$ for many initial values $x_1^{(0)}, \ldots, x_{10^5}^{(0)}$;
- Requirements: Efficient method with fast convergence speed; memory consumption of the method should be bounded by a constant times 10^5;
- Numerical method: Newton's method applied to $\nabla E(x_1, \ldots, x_{10^5}) = 0$ with Armijo line search;
- Decision:
- Justification:

c)

- Situation: For the development of a space rocket an on-the-fly stabilization is needed. The rocket should be stabilized with minimal energy. Given measured parameters $h, r, \omega \in \mathbb{R}$ during a flight, we have to solve a minimization problem, where any conditions for existence and uniqueness of a solution can be assumed to hold.
- Problem: For $N = 10^3$ consider

$$\min_{u\in\mathbb{R}^{2N}, z\in\mathbb{R}^{2N}} \frac{r}{2}h\sum_{i=1}^{N} u_i^2 + \frac{1}{2}(z_N^2 + z_{2N}^2)$$
$$\text{s.t. } z_{i+1} - z_i = h(\omega z_{N+i} + u_i), \quad i = 1, \dots, N$$
$$z_{N+i+1} - z_{N+i} = h(-\omega z_i + u_{N+i}), \quad i = 1, \dots, N$$

- Requirements: The numerical method should be efficient, because the problem has to be solved quickly when stabilization problems occur. The Lagrange multiplier with respect to the constraints are needed in the application. The accuracy of a solution is crucial for a precise control system.
- Numerical method: Nullspace method
- Decision:
- Justification:

d)

- Situation: An airline always tries to find the best routes, in terms of cost minimization, for their airplanes. We try to find the optimal parameters $x \in \mathbb{R}^{40}$ of the route, which minimizes the cost function $f(x)$. However, there must be met several constraints exactly $Bx = b \in \mathbb{R}^{15}$, e.g. safety constraints. Other constraints, stemming from weather forecasts are just bounds $g(x) \leq 0 \in \mathbb{R}^{25}$. Fortunately, the functions f and g are smooth and convex, but nonlinear.
- Problem: Solve

$$\min_{x\in\mathbb{R}^{15}} f(x)$$
$$\text{s.t. } Bx = b$$
$$g(x) \leq 0$$

- Requirements: The equality constraints should be fulfilled with high precision.
- Numerical method: Barrier method and the adapted (for nonlinear objective functions f) nullspace method for the subproblems.
- Decision:
- Justification:

Lösung:

a)
- Decision: The method is the wrong choice.
- Justification: The Levenberg-Marquardt method requires smoothness of the objective function, which is not the case here. The absolute value function is not differentiable at the point zero, which can happen if the residual is zero (i.e. the model function lies exactly on a data point, whcih is likely). Although, there are no requirements, the Levenberg-Marquardt method is simply not applicable and thus the wrong choice.

b)
- Decision: The method is the wrong choice.
- Justification: Although, the convergence speed of Newton's method is locally quadratic, it uses the Hessian of the function E which is a matrix of size $10^5 \times 10^5$. Thus, the memory consumption is not in $\mathcal{O}(10^5)$.

c)
- Decision: The method is the right choice.
- Justification: The objective function is quadratic in the variables u and z. The constraints are linear equality constraints. Furthermore, all conditions for existence and uniqueness are ensure. Also, the method should be efficient and the accuracy of the solution is important. Therefore the nullspace method is appropriate, since it calculates the solution by solving linear systems of equations. The Lagrange multiplier can be calculated within the nullspace method as well.

d)
- Decision: The method is the right choice.
- Justification: With the Barrier method the problem is reduced to a sequence of optimization problems with only equality constraints. Those subproblems can be solved using the nullspace mehtod, if we adapt the method for nonlinear objective functions (which is possible). This is the right choice because the method minimizes f only in the subspace of vectors fulfilling $Bx = b$ exactly.

Last Name, First Name: ______________________________

Task 3 (Optimality Conditions) **(4 + 4 + 4 + 4 = 16 Points)**

In this task we use the following notation for a constraint optimization problem.
For $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$, we consider

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } &h(x) = 0, \\ &g(x) \leq 0. \end{aligned} \tag{1}$$

The feasible set is defined as $\mathcal{F} := \{x \in \mathbb{R}^n : h(x) = 0,\ g(x) \leq 0\}$.

a) *Definition*: Problem (1) satisfies the *LICQ condition* in a feasible point $\hat{x} \in \mathcal{F} \subset \mathbb{R}^n$, if $\nabla g_i(\hat{x}) \in \mathbb{R}^n$ and $\nabla h_j(\hat{x}) \in \mathbb{R}^n$ are linearly independent for all $i \in \mathcal{A}(\hat{x})$ and for all $j = 1, \ldots, m$.

   Definition: We say that the *MFCQ* holds at $\hat{x} \in \mathcal{F}$, if the gradients $\nabla h_j(\hat{x})$ for $j = 1, \ldots, m$ are linear independent and there exists a vector $d \in \mathbb{R}^n$ such that

$$\nabla g_i(\hat{x})^T d < 0, \quad i \in \mathcal{A}(x), \quad \nabla h(\hat{x})^T d = 0.$$

   Prove: LICQ(x) $\quad \Rightarrow \quad$ MFCQ(x)

b) Let in this subtask $f(x_1, x_2) := x_1^2 + x_2^2$ and $\hat{x} := (0, 1)^T$ be given. Construct constraints, such that LICQ($\hat{x}$) does not hold, but MFCQ($\hat{x}$) holds.
Hint: Keep it as simple as possible, however, adding identical constraints does not count!

c) Let in this subtask $f(x_1, x_2) := 2x_1^2 + x_1x_2 + 3x_2^2 - 4x_1 - x_2$, $h(x_1, x_2) := 3x_1 + \alpha x_2 - \beta$ for $\alpha, \beta \in \mathbb{R}$ and $x^* = (3, -2)^T \in \mathbb{R}$ be given. Furthermore, there is no inequality constraint g.

i) Determine α and β such that x^* is the *unique* solution of the minimization problem (1).

ii) Determine the feasible set $\mathcal{F}$ as well as the tangential cone $\mathcal{T}(\mathcal{F}, x^*)$ and the linearized cone $\mathcal{T}_{\text{lin}}(\mathcal{F}, x^*)$ explicitly.
Hint: Use $\alpha = -5$ and $\beta = 19$ if you didn't solve i). The feasible set is of the form $\mathcal{F} = x + \operatorname{span}\{v\} \subset \mathbb{R}^2$

Lösung:

a) **Proof**: Obviously, the gradients $\nabla h_j(\hat{x})$ for $j = 1, \ldots, m$ are linear independent due to LICQ. It remains to find a suitable vector $d \in \mathbb{R}^n$ such that

$$\nabla g_i(\hat{x})^T d < 0, \quad i \in \mathcal{A}(x), \quad \nabla h(\hat{x})^T d = 0.$$

hold. Lets define the matrix

$$\begin{pmatrix} \nabla g_i(\hat{x})^T & i \in \mathcal{A}(\hat{x}) \\ \nabla h_j(\hat{x})^T & j = 1, \ldots, m \end{pmatrix} \in \mathbb{R}^{(|\mathcal{A}(\hat{x})|+m)\times n}$$

which has full rank by LICQ. Hence we can add rows to obtain a non-singular matrix $A(\hat{x}) \in \mathbb{R}^{n\times n}$. Then, the linear system

$$A(\hat{x})d = \begin{pmatrix} -e \\ 0 \end{pmatrix},$$

where $e = (1, \ldots, 1)^T \in \mathbb{R}^{|\mathcal{A}(\hat{x})|}$ has a solution $\hat{d} \in \mathbb{R}^n$ which fulfills the requirements of MFCQ.

b) We keep the constraints as simple as possible and drop the equality constraints. We use

$$g(x) := \begin{pmatrix} -x_1 \\ -2x_1 \end{pmatrix} \quad \text{and so} \quad \nabla g(x) = \begin{pmatrix} -1 & 0 \\ -2 & 0 \end{pmatrix}.$$

It holds $g(\hat{x}) = (0,\ 0)^T$ and the gradients $\nabla g_i(\hat{x})$ can not be linear independent, however, there is the vector $d := (1,\ 0)^T$ with $(\nabla g_i(\hat{x}))^T d < 0$ for $i \in \mathcal{A}(\hat{x})$.

c) i) The necessary condition for $(3, -2)^T \in \mathbb{R}^2$ to be a minimum is with respect to the constraint $h(x, y) = 0$ the existence of a Lagrange multiplier $\lambda^* \in \mathbb{R}$, s.t. $\nabla L(3, -2, \lambda^*) = 0$. The optimization problem is convex with linear equality constraints, therefore this condition is even sufficient (see Theorem 3.1.20).
The Lagrange function is given by

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda g(x, y) = 2x^2 + xy + 3y^2 - 4x - y + \lambda(3x + \alpha y - \beta).$$

and for the gradient it holds:

$$\nabla\mathcal{L}(x, y, \lambda) = \begin{pmatrix} 4x + y - 4 + 3\lambda \\ x + 6y - 1 + \alpha\lambda \\ 3x + \alpha y - \beta \end{pmatrix}.$$

We plug in the point $(3, -2)^T \in \mathbb{R}^2$ for which we get:

$$0 \stackrel{!}{=} \nabla\mathcal{L}(3, -2, \lambda) = \begin{pmatrix} 6 + 3\lambda \\ -10 + \alpha\lambda \\ 9 + \alpha y - \beta \end{pmatrix},$$

This leads us to the solution $\lambda = -2$, $\alpha = -5$ and $\beta = 19$.

ii) The feasible set is defined by $\mathcal{F} = \{x \in \mathbb{R}^2 : h(x_1, x_2) = 3x_1 - 5x_2 - 19 = 0\}$. Therefore, the set can be formulated as $\mathcal{F} = x^* + \operatorname{span}\{(5/3,\ 1)^T\}$, because $h(x^*) = 0$ and $(5/3,\ 1)\begin{pmatrix} 3 \\ 5 \end{pmatrix} = 0$.
The linearized cone is then given by

$$\mathcal{T}_{\text{lin}}(\mathcal{F}, x^*) = \{d \in \mathbb{R}^2 : (3,\ -5)^T d = 0\} = \operatorname{span}\{(5/3,\ 1)^T\}$$

The tangential cone is also given by $\operatorname{span}\{(5/3,\ 1)^T\}$, due to the fact that $\mathcal{T}(\mathcal{F}, x^*) \subseteq \mathcal{T}_{\text{lin}}(\mathcal{F}, x^*)$ and for $d \in \mathcal{T}_{\text{lin}}(\mathcal{F}, x^*)$ we take $\eta^{(\ell)} := \ell$ and $x^{(\ell)} := x^* + \frac{1}{\ell}d$. With that we have

$$\eta^{(l)}(x^{(l)} - x^*) \stackrel{x^{(l)}=x^*+d/\ell}{=} d.$$

Therefore $d = \lim\limits_{l\to\infty} \eta^{(l)}(x^{(l)} - x^*) \in \mathcal{T}(\mathcal{F}, x^*)$.

Last Name, First Name: ______________________________

Task 4 (Nonlinear Least Squares Problems) (3 + 1 + 4 + 2 = 10 Points)

In this task we consider the following optimization problem. For $F \in C^2(\mathbb{R}^n;\mathbb{R}^m)$ with $m > n$ and $g(x) := \frac{1}{2}\|F(x)\|_2^2$ we consider

$$\min_{x\in\mathbb{R}^n} g(x). \tag{2}$$

For your convenience:

$$\nabla g(x) = (F'(x))^T F(x), \quad \nabla^2 g(x) \approx (F'(x))^T F'(x) := H(x)$$

a) Show that, if the gradient $\nabla g(x^{(k)})$ is nonzero and the Jacobian $F'(x^{(k)})$ has full rank, then the direction $s^{(k)}$ of the Gauß-Newton algorithm determined by

$$F'(x^{(k)})^T F'(x^{(k)}) s^{(k)} = -F'(x^{(k)})^T F(x^{(k)})$$

is a descent direction, i.e. $(s^{(k)})^T \nabla g(x^{(k)}) < 0$.

b) In the Levenberg-Marquardt method a parameter μ is introduced. *Explain* briefly what this parameter does to the step $s^{(k)}$.

c) Consider the following constraint optimization problem:

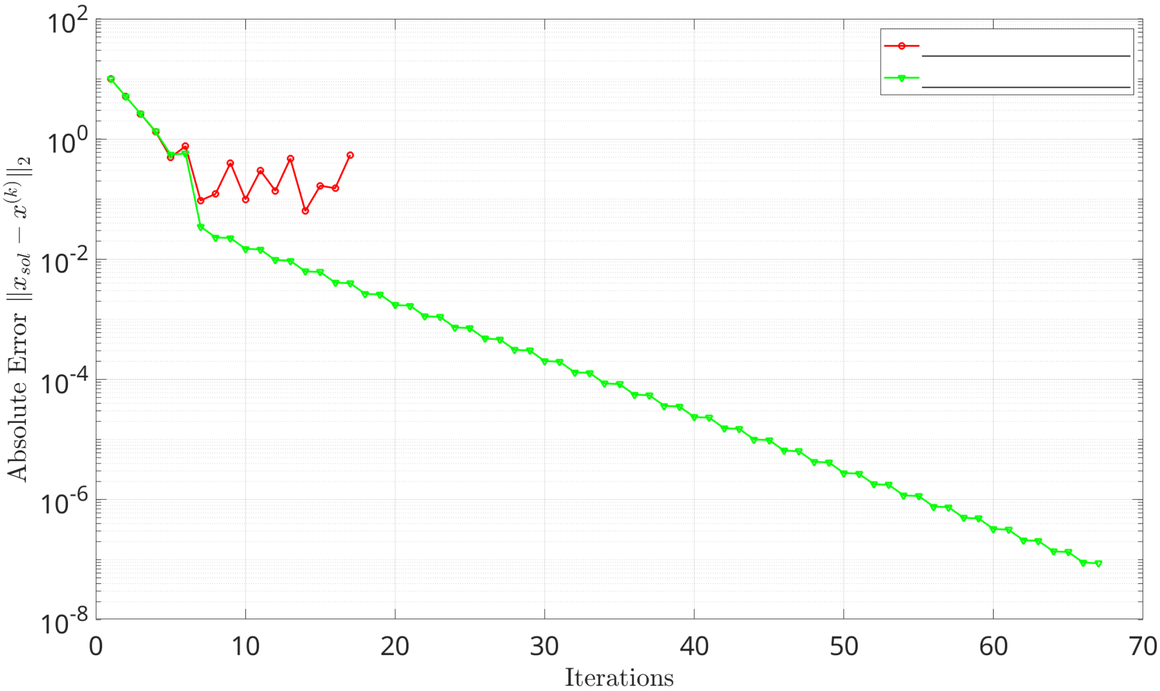
$$\begin{aligned} &\min_{s\in\mathbb{R}^n} \frac{1}{2}s^T H(x)s + \nabla g(x)s + g(x) \\ &\text{s.t. } \frac{1}{2}(\|s\|_2^2 - \Delta^2) \leq 0. \end{aligned} \tag{3}$$

In which optimization method does this subproblem appear? Derive the KKT conditions for (3). In which way is the multiplier rule used within the Levenberg-Marquardt method?

d) The following plot shows the convergence of the Levenberg-Marquardt and the Gauß-Netwon method. However, one algorithm breaks with the message:

```
Warning: A is rank deficient to within machine precision.
```

Assign the methods to the curves and explain your decision.



Lösung:

a) We start with

$$(s^{(k)})^T \nabla g(x^{(k)}) = (s^{(k)})^T (F'(x^{(k)}))^T F(x^{(k)}) = -(s^{(k)})^T (F'(x^{(k)}))^T F'(x^{(k)}) s^{(k)} \leq 0$$

The inequality holds, because if $F'(x^{(k)})$ has full rank, then the symmetric matrix $(F'(x^{(k)}))^T F'(x^{(k)})$ is positive definite. Therefore, if the gradient $\nabla g(x^{(k)})$ isn't zero, $s^{(k)}$ isn't zero and so the inequality above holds strictly.

b) This parameter is a damping for the step $s^{(k)}$, i.e. it reduces the norm $\|s^{(k)}\|$ for large parameters μ.

c) Because of the fact that for $\|s\|_2 > 0$ the constraint is equivalent to $\|s\|_2 \leq \Delta$, the subproblem is the quadratic model of the trust region method applied to $\min_{x\in\mathbb{R}^n} g(x)$.
The KKT conditions for (3) are given by

$$\begin{aligned} (F'(x))^T F'(x)s + (F'(x))^T F(x) + \mu s &= 0 \\ \mu \geq 0, \quad \frac{1}{2}(\|s\|_2^2 - \Delta^2) &\leq 0 \\ \mu \frac{1}{2}(\|s\|_2^2 - \Delta^2) &= 0 \end{aligned}$$

Therefore, the multiplier rule is, except for μ instead of μ^2 the update rule for the step $s^{(k)}$.

d) The red curve is the Gauß-Newton method. The green one is the Levenberg-Marquardt. This is due to the fact, that in the Levenberg-Marquardt the matrix $(F'(x))^T F'(x) + \mu^2 I$ is always invertible. The matrix $(F'(x))^T F'(x)$ in the Gauß-Newton method does not have to be invertible for all x.

Last Name, First Name: ____________________

Task 5 (Interior Point Method for Linear Programs) (3 + 2 + 3 = 8 Points)

In this task we consider the following constrained optimization problem.
For $f \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and the full rank matrix $B \in \mathbb{R}^{m\times n}$ we consider

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \; & f^T x \\ \text{s.t. } Bx &= b, \\ x &\geq 0 \end{aligned} \tag{4}$$

a) Derive the dual problem of problem (4) and name the Lagrange multiplier for the inequality constraints s.

b) Derive the KKT conditions for problem (4).

c) The interior point method for the linear program(4) is for $x \geq 0$, $s \geq 0$ based upon the perturbed system $F_\mu(x, \lambda, s) = 0$ with

$$F_\mu(x, \lambda, s) := \begin{pmatrix} B^T\lambda + s - f \\ Bx - b \\ XSe - \mu e \end{pmatrix}, \quad \nabla F_0 = \begin{pmatrix} 0 & B^T & I \\ B & 0 & 0 \\ S & 0 & X \end{pmatrix}$$

as well as $e := (1, \dots, 1)^T \in \mathbb{R}^n$, $X := (\mathrm{diag}(x_i))_{i=1,\dots,n} \in \mathbb{R}^{n \times n}$ and $S := (\mathrm{diag}(s_i))_{i=1,\dots,n} \in \mathbb{R}^{n \times n}$. Bob asks you to complete his code:

```
function [x, lambda, s] = InteriorPointMethod(f, B, b, x, lambda, s, sigma, tol)
% Input:
%   f        : vector f in R^n
%   B        : constraints matrix in R^(mxn)
%   b        : constraints vector in R^m
%   x        : initial value for the solution vector x in R^n
%   lambda   : initial vector for the Lagrange multiplier equality constr. in R^m
%   s        : initial vector for the Lagrange multiplier inequality constr. in R^n
%   tol      : termination tolerance

% Output:
%   x        : solution vector x in R^n
%   lambda   : solution Lagrange multiplier
%   s        : solution Lagrange multiplier

% initial values
eta     = 0.99;
[m,n]   = size(B);
tau     = (x'*s)/n;
I       = eye(n);
zerOnn = zeros(n);
zerOmm = zeros(m);
zerOmn = zeros(m,n);
e       = ones(n,1);

while tau > tol

    % set up DF_0 * (Delta x, Delta lambda, Delta s) = (0, 0, XSe-sigma*tau*e)
    X     = diag(x);
    S     = diag(s);
    % TODO
    DF_0 =

    rhs   = [zeros(n,1); zeros(m,1); X*S*e - sigma*tau*e];

    % solve the system and save out the solutions
    delta_sol    = DF_0\rhs;
    Deltax       = delta_sol(1:n);
    Deltalambda = delta_sol(n+1:n+m);
    Deltas       = delta_sol(n+m+1:end);

    % calculate stepsize alpha:
    alpha =  eta / max([eta; Deltax./x; Deltas./s]);

    % Update iterates: % TODO





    % calculate tau:
    tau = (x'*s)/n;
end
```

Lösung:

a) The Lagrange function is given by $\mathcal{L}(x, \lambda, s) = f(x) + \lambda^T(b - Bx) - s^T x$. Thus, the dual function reads

$$\varphi(\lambda, s) = \inf_{x\in\mathbb{R}^n} \mathcal{L}(x, \lambda, s) = \inf_{x\in\mathbb{R}^n} \{f^T x - \lambda^T Bx - s^T x\} + \lambda^T b$$
$$= \begin{cases} \lambda^T b, & \text{if } s = f - B^T\lambda \\ -\infty, & \text{else} \end{cases}.$$

With the constraint $s \geq 0$ we get the dual problem

$$\begin{aligned} \max_{\lambda\in\mathbb{R}^m, s\in\mathbb{R}^n} \quad & b^T\lambda \\ \text{s.t. } & s = f - B^T\lambda, \\ & s \geq 0 \end{aligned}$$

b) The KKT conditions are given by

$$\begin{aligned} f + B^T\lambda - s &= 0 \\ Bx - b &= 0 \\ x \geq 0,\ s &\geq 0 \\ s^T x &= 0 \end{aligned}$$

c) The code looks like:

```
DF_0 = [zerOnn, B', I; B, zerOmm, zerOmn; S, zerOmn', X];

% Update:
x         = x      - alpha * Deltax;
lambda    = lambda - alpha * Deltalambda;
s         = s      - alpha * Deltas;
```

Last Name, First Name: ____

Task 6 (Projected Subgradient Method) (3 + 2 = 5 Points)

In this task we consider the $\|\cdot\|_1$-regression problem for a linear function $y(x) = c_1 \cdot x + c_2$. We have given data $A \in \mathbb{R}^{p\times 2}$ as well as $b \in \mathbb{R}^p$ and try to determine the parameters c_1 and c_2 of the function y. This results in the following non-smooth optimization problem

$$\min_{c\in\mathbb{R}^2} \left\| A \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} - b \right\|_1 \tag{5}$$

where $\|v\|_1 := \sum_{i=1}^{p} |v_i|$ for $v \in \mathbb{R}^p$.

a) Calculate the convex sub-differential $\partial f(x)$ for all $x \in \mathbb{R}^2$.

Hint: You can find the calculus rules below and you can use $\partial|x| = \begin{cases} \{-1\}, & x < 0 \\ [-1, 1], & x = 0 \\ \{1\}, & x > 0 \end{cases}.$

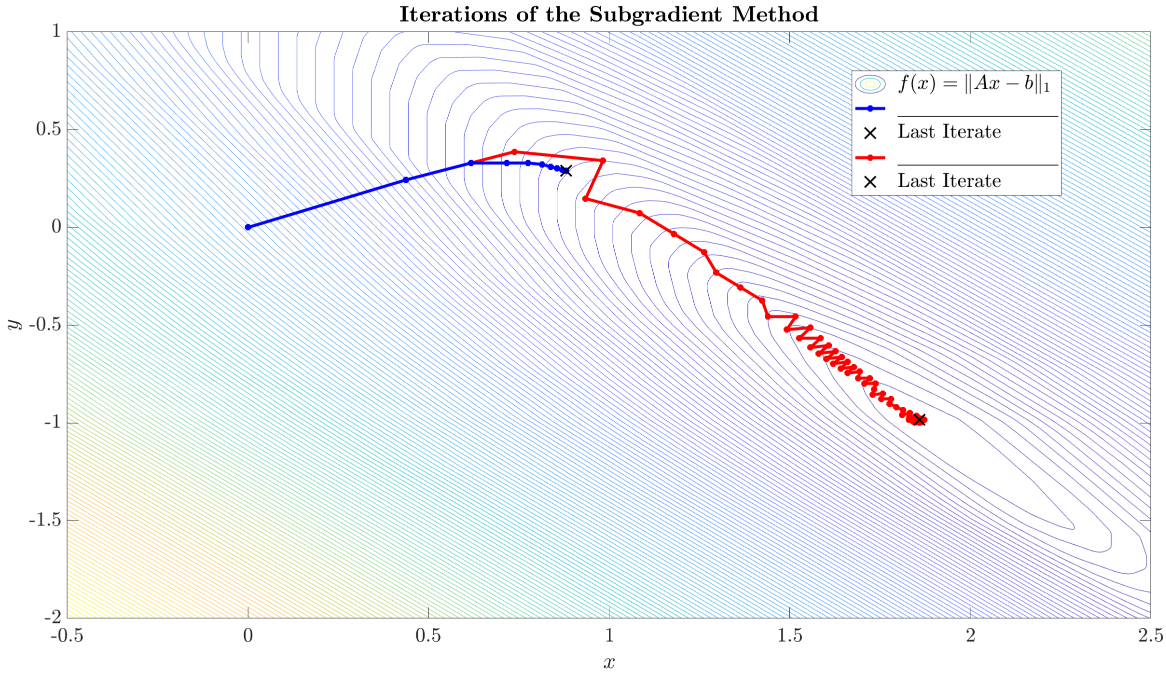
Subdifferential calculus: For convex functions $f_1, \dots, f_p : \mathbb{R}^n \to \mathbb{R}$, there holds

$$\partial(f_1 + \ldots + f_p)(x) = \partial f_1(x) + \ldots \partial f_p(x) = \left\{ \sum_{i=1}^{p} w_i : w_i \in \partial f_i(x) \right\}.$$

Moreover, for convex $f : \mathbb{R}^m \to \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ there holds

$$\partial f(Ax + b) = A^T \partial f(Ax + b).$$

b) Bob and Alice applied the subgradient method to the problem at hand. The following plot shows the respective path of iterations together with the contour plot of the function in (5). Bob has chosen the step size $\sigma_k = \frac{1}{k+1}$, whereas Alice chose $\sigma_k = \frac{1}{k^2+1}$. Assign the paths to the right persons and explain your choice!



Lösung:

a) We define the functions $f_i : \mathbb{R}^2 \to \mathbb{R}$, $c \mapsto |a_i^T c - b_i|$, where a_i^T is a row of A. Therefore, we have by the second rule and the hint

$$\partial f_i(x) = a_i \partial |a_i^T c - b_i| = \begin{cases} \{-a_i\}, & a_i^T c - b < 0 \\ \{\alpha a_i \in \mathbb{R}^2 : \alpha \in [-1,1]\}, & a_i^T c - b = 0 \\ \{a_i\}, & a_i^T c - b > 0 \end{cases}$$

With that and the first subdifferential calculus rule it follows

$$\partial \sum_{i=1}^{p} f_i(x) = \left\{ \sum_{i=1}^{p} w_i : w_i \in \partial f_i(x) \right\}.$$

b) The red curve is the plot from Bob and the blue one is the curve from Alice calculations. The reason is that if we choose a step size such that $\sum\limits_{k=0}^{\infty} \sigma_k < 0$ as Alice did, we get $\|x^{(k)} - x^{(0)}\|_2 \leq \sum\limits_{k=0}^{\infty} \sigma_k =: M$. If the initial vector $x^{(0)}$ is chosen in a way, such that it is far away from the solution, i.e. $\|x^* - x^{(0)}\|_2 > M$, then we can never reach the solution x^*.

Name, Vorname: ____________________

Good luck !!!