EXAM COVER SHEET

Name of the Exam:	Numerical Optimization
Time and Date:	13.02.2025, 10 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	14	12	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 $\min_{x \in \mathbb{R}^n} f(x)$, where f is semismooth. true ◯ false ◯

(b) By the introduction of a slack variable, one can transform a minimization problem with inequality constraints into a minimization problem with only equality constraints. true ◯ false ◯

(c) Let (x^*, λ^*, μ^*) be a saddle point of the Lagrange function $\mathcal{L}(x, \lambda, \mu)$ of the problem $\min_{x \in \mathcal{F} \subset \mathbb{R}^n} f(x)$, with $\mathcal{F}$ being the feasible set. Then, x^* can not be a minimizer. true ◯ false ◯

(d) The dual function $\varphi(\lambda, \mu) := \inf_{x \in \mathcal{F}_P} \mathcal{L}(x, \lambda, \mu)$ is concave on the dual feasible region $\mathcal{F}_D$. true ◯ false ◯

(e) The Zoutendijk condition $\sum_{k=0}^{\infty} \left(\cos(\theta^{(k)})\right)^2 \|\nabla f(x^{(k)})\|^2 < \infty$ ensures the convergence of iterations $(x^{(k)})_{k \in \mathbb{N}}$, because it follows that $\|\nabla f(x^{(k)})\|^2$ has to be a null sequence. 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) Let $A \in \mathbb{R}^{m \times n}$ have full rank and $m > n$. Then, the Gauß-Newton method solves the problem $g(x) := \|Ax - b\|^2 \to \min$ in m iterations. true ◯ false ◯

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

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

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

Lösung:

(a) false

(b) false

(c) false

(d) true

(e) false

(f) false

(g) false

(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 which numerical method you would use and why.
Remark: A maximum of *one point* is awarded for a method that can be applied but does not meet the requirements. However, the applicability must be justified.
Moreover, if a method reduces a minimization problem to a sequence of minimization problems, then the numerical method for the sub-problems also has to be given.

a)
- Situation: Astronomers measured the coordinates $u_i \in \mathbb{R}^2$ at times $t_i > 0$, $i = 1, \dots, 100$ of an object orbiting in the gravitational field of the earth and moon. The orbit of such an object can be described by the differential equation $(x''(t), y''(t))^T = f(t, x(t), x'(t), y(t), y'(t); m, u_0)$, where f is a known function, m is the unknown mass of the object and $u_0 \in \mathbb{R}^4$ is the vector of unknown initial conditions. Given parameters (m, u_0), this equation can be solved with MATLABs `ode45`-function to get a numerical solution $(x_i, y_i)^T \in \mathbb{R}^2$, $i = 1, \dots, 100$. The astronomers want to determine (m, u_0) by fitting the data u_i.
- Problem: Solve $\min\limits_{m \in \mathbb{R}, u_0 \in \mathbb{R}^4} \sum\limits_{i=1}^{100} \|u_i - (x_i, y_i)^T\|_2^2$, where $[x_i, y_i] = \texttt{ode45}(f, m, u_0)$.
- Requirements: Evaluations of `ode45` are slow and should be avoided as much as possible.
- Numerical method:
- Justification:

b)
- 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

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

- 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:
- Justification:

c)
- Situation: Photographs of some objects have been taken at a crime scene. A digital forensics expert tries now to generate 3D models of the objects, given the pixel information from the photographs. The expert wants you to train a neural network for this. Thus, the parameters $\theta \in \mathbb{R}^{10^6}$ of the neural network has to be determined by minimizing the loss function $\mathcal{L} \in C^2(\mathbb{R}^{10^6})$. Derivatives of $\mathcal{L}$ can be computed exactly by automatic differentiation, although it is computational demanding. The loss function $\mathcal{L}$ is expected to have flat areas, i.e. areas where the norm of the gradient is small.
- Problem: Solve $\min\limits_{\theta \in \mathbb{R}^{10^6}} \mathcal{L}(\theta)$
- Requirements: The problem should be solved on a graphics processing unit (GPU) with limited memory, such that there can only 10^8 values be stored at most. Convergence speed is important due to the time restriction in the investigation.
- Numerical method:
- Justification:

d)
- Situation: A merchant sells n products. When placing an order, he must decide what quantity x_i of product i to buy. We summarise the order quantities in a vector $x = (x_1, ..., x_n)^T$. The delivery of the ordered quantities is associated with various costs. As a rule, only a maximum amount c_i is available for the total costs $f_i(x)$ for product i. Therefore, the conditions $f_i(x) \leq c_i$, $i = 1, \ldots, n$ must be fulfilled. The cost functions f_i are convex and derivatives are available analytically. The merchant tries to calculate an optimal order x^*, fulfilling all requirements.
- Problem: Solve

$$\min_{x \in \mathbb{R}^n} \max_{1 \leq i \leq n} (f_i(x) - c_i)$$
$$\text{s.t. } x_i \geq 0, \quad i = 1, \ldots, n$$

- Requirements: None
- Numerical method:
- Justification:

Lösung:

a)
- Numerical method: Nelder-Mead/Hooke-Jeeves method
- Justification: Although, it is a nonlinear regression problem, the non-availability of gradients excludes all gradient-based methods. Also the approximation of derivatives with, for instance, finite differences should be avoided. Thus, the only available method left is the Nelder-Mead and the Hooke-Jeeves method. Which method needs less function evaluations can hardly be estimated.

b)
- Numerical method: Nullspace method
- 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.

c)
- Numerical method: Nonlinear cg-method with Wolfe stepsize control
- Justification: Derivatives can be computed, although the memory is limited. Therefore, all Hessian-based algorithms are excluded. The convergence speed is important for the application and so a gradient-based method seems appropriate. Due to the flat areas in the objective function a stepsize control is needed to speed-up the iterations.

d)
- Numerical method: Projected sub-gradient method
- Justification: The only requirement here is the applicability of the method. The objective function is convex, due to the convexity of the functions f_i and the convexity of the maximum function. Moreover, it is non-differentiable such that we have to apply either derivative-free methods or a non-smooth optimization method. Only the latter can deal with constraints, such that we choose the projected sub-gradient method.

Last Name, First Name: ____

Task 3 (Optimality Conditions) (4 + 4 + 4 + 2 = 14 Points)

In this exercise 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 $j = 1, \ldots, m$.

   Prove that, if a local solution $x^* \in \mathcal{F}$ satisfies the LICQ condition, than the Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ at a KKT point $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ of (1) are unique.

b) Let f be given by $f(x) := (x_1 - 2)^2 + x_2^2$, let the tangential cone be given by $\mathcal{T}(\mathcal{F}, x^*) := \{x \in \mathbb{R}^2 : x_1 \leq 0,\ x_2 = 0\}$ and the linearized tangential cone by $\mathcal{T}_{\text{lin}}(\mathcal{F}, x^*) := \{x \in \mathbb{R}^2 : x_2 = 0\}$ at the minimum $x^* := (1, 0)^T \in \mathcal{F}$. Check the necessary condition for x^* being a local minimum. What do you observe if you replace $\mathcal{T}(\mathcal{F}, x^*)$ by $\mathcal{T}_{\text{lin}}(\mathcal{F}, x^*)$?

c) Let $f(x, y) := (x_1 - 2)^2 + (x_2 + 1)^2 - 2x_1 - x_2$ and let $g(x_1, x_2) := 2x_1 + \alpha x_2 + \beta$ for $\alpha, \beta \in \mathbb{R}$ be the functions in Problem(1). Furthermore, there is no equality constraint h. Determine α and β such that $x^* = (2, 1)^T$ is the *unique* solution of the minimization problem (1).

d) Let Problem (1) be given with $f(x) := \frac{1}{2}x^T Ax + a^T x + g(x)$ and $g(x) := (x - c)^T (x - c) - 9$. Let further $A \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite and let $a \in \mathbb{R}^n$ as well as $c \in \mathbb{R}^n$ be fix. There is no equality constraint. Show that a minimum $x^* \in \mathbb{R}^n$ fulfills the KKT conditions.
Hint: You can use any statement from the lecture notes and the exercise sheets.

Lösung:

a) **Proof**: We assume there exist two KKT points (x^*, λ^*, μ^*) and $(x^*, \overline{\lambda}, \overline{\mu})$ of the optimization problem (1) with $(\lambda^*, \mu^*) \neq (\overline{\lambda}, \overline{\mu})$. This means the KKT conditions of Theorem 3.1.17 (first order optimality conditions) hold for these points. Further, let the LICQ condition be fulfilled at x^*. We insert both KKT points in the "multiplier rule". For (x^*, λ^*, μ^*) we get

$$\nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \mu_i^* \nabla g_i(x^*) + \sum_{j=1}^{m} \lambda_j^* \nabla h_j(x^*) = 0. \tag{2}$$

Analogously, for $(x^*, \overline{\lambda}, \overline{\mu})$ it is

$$\nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \overline{\mu}_i \nabla g_i(x^*) + \sum_{j=1}^{m} \overline{\lambda}_j \nabla h_j(x^*) = 0. \tag{3}$$

Next, we subtract (3) from (2) to get

$$\sum_{i \in \mathcal{A}(x^*)} (\mu_i^* - \overline{\mu}_i) \nabla g_i(x^*) + \sum_{j=1}^{m} (\lambda_j^* - \overline{\lambda}_j) \nabla h_j(x^*) = 0. \tag{4}$$

For satisfying the LICQ condition it is required that $\nabla g_i(x^*)$ and $\nabla h_j(x^*)$ are linearly independent for all $i \in \mathcal{A}(x^*)$ and for all $j = 1, ..., m$, which is:

$$\sum_{i \in \mathcal{A}(x^*)} a_i \nabla g_i(x^*) + \sum_{j=1}^{m} b_j \nabla h_j(x^*) = 0 \quad \Rightarrow a_i = 0,\ i \in \mathcal{A}(x^*),\ \text{and } b_j = 0,\ j = 1, ..., m. \tag{5}$$

It follows $(\lambda^*, \mu^*) = (\overline{\lambda}, \overline{\mu})$.

b) The necessary condition for a local minimum is $\nabla f(x^*)^T d \geq 0$ for all $d \in \mathcal{T}(\mathcal{F}, x^*)$. The gradient of f is given by $\nabla f(x) = (2(x_1 - 2), 2x_2)^T$. Therefore, we have $\nabla f(x^*) = (-2, 0)^T$ and

$$(\nabla f(x^*))^T d = -2d_1 \geq 0, \quad \forall d_1 \leq 0.$$

This is fulfilled for all $d \in \mathcal{T}(\mathcal{F}, x^*)$. If we replace $\mathcal{T}(\mathcal{F}, x^*)$ by $\mathcal{T}_{\text{lin}}(\mathcal{F}, x^*)$, the condition is not longer satisfied, because d_1 can also be negative. This is the reason for the ACQ condition.

c) The problem is convex, because the Hessian $\nabla^2 f(x) = I$ is positive definite and the inequality constraints are affine linear for all α, β, so in particular convex. Thus we have uniqueness of a solution. To determine the constants $\alpha, \beta \in \mathbb{R}$ we have to derive the KKT system. This is given by

$$\begin{pmatrix} 2(x_1 - 2) - 2 + 2\mu \\ 2(x_2 + 1) - 1 + \alpha\mu \\ \mu(2x_1 + \alpha x_2 + \beta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We plug in $x^* = (2, 1)^T$ and derive from the first equation $\mu = 1$. The second equation gives $\alpha = -3$ and the last one $\beta = -1$. Now we have to check the complementarity conditions. We have obviously $\mu \geq 0$. Plugging x^* into g gives $g(x^*) = 2 \cdot 2 + -3 \cdot 1 - 1 = 0$ so that also $\mu \cdot g(x^*) = 0$ is fulfilled.

d) As before the problem is convex, because the Hessians of $\nabla^2 f = A + I$ and $\nabla^2 g = I$ are positive definite. The slater condition holds in this case, because

$$\mathring{\mathcal{F}} := \{x \in \mathbb{R}^n : g(x) < 0\} \neq \emptyset.$$

holds, due to $c \in \mathring{\mathcal{F}}$. This proves that the minimum fulfills the KKT conditions.

Last Name, First Name: ______________________________

Task 4 (From Quadratic to Nonlinear Problems) (3 + 4 + 2 + 3 = 12 Points)

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

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) := \frac{1}{2}x^T Ax + a^T x \\ &\text{s.t. } Bx = b. \end{aligned} \tag{6}$$

We further assume throughout the task, that A is positive definite on the kernel of B, i.e. $z^T Az > 0$ for all $z \in \ker(B)$.

a) Show that the matrix

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$$

is regular.

b) Show that the solution x^* of

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -a \\ b \end{pmatrix}$$

is the unique solution of Problem (6).

Hint: The KKT-Theorem for convex problems can *not* be used here. Let instead $x \in \mathcal{F}$ be any other feasible point and define $p = x^* - x$. Consider then $f(x) = f(x^* - p)$ and try to prove $f(x) > f(x^*)$.

c) Let in this sub-task $f(x)$ be convex but not necessarily quadratic.

i) Let $x^* \in \mathbb{R}^n$ be the unique solution of the minimization problem. Prove, that the Lagrange multiplier $\lambda^* \in \mathbb{R}^m$, which is given by

$$\lambda^* = -(BB^T)^{-1}B\nabla f(x^*)$$

is unique.
Hint: You only need to prove the uniqueness!

ii) Complete the following code-snippet of the nullspace method, such that convex nonlinear functions f can be optimized. Return also the Lagrange multiplier λ^*.
Hint: You are allowed to use MATLABs internal **fminunc** function.

```
function [x, lambda] = convex_nullspace_method(f, gradf, z0, B, b)
% Input:
%   f       : function handle realizing f(x), being convex
%   gradf   : function handle realizing gradf(x)
%   z0      : initial value for the reduced minimization problem
%   B,b     : constraints Bx = b
% Output:
%   x       : solution vector x
%   lambda  : Lagrange multiplier

% dimensions
n=size(B,2);
m=size(B,1);

%-- (1) QR-decomposition and splitting
[Q, R]=qr(B');
R = R(1:m,1:m);
Y = Q(1:n,1:m);
Z = Q(1:n,m+1:n);

%-- (2) determine x_Y
x_Y = (R')\b;

%-- (3) determine x_Z by minimizing f(Y x_y + Z z)
% TODO:


% determine the full solution x
% TODO:


% determine the lagrange multiplier
% TODO:


end
```

Lösung:

a) We show that the matrix
$$K := \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$$
is regular by showing that the kernel $\ker(K)$ is trivial. Therefore, let $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$ be given such that $(v^T, w^T)^T \in \ker(K)$. By multiplying we receive the system
$$\begin{aligned} Av + B^T w &= 0 \\ Bv &= 0. \end{aligned}$$
It follows $v \in \ker(B)$ and by multiplying the first equation with v^T from the left we receive
$$v^T Av + (Bv)^T w = v^T Av = 0.$$
This can only hold for $v = 0$. Again from the first equation we have now $B^T w = 0$, from which $w = 0$ follows, due to the linear independence of the columns of B^T.

b) We follow the hint an derive
$$\begin{aligned} f(x) = f(x^* - p) &= \frac{1}{2}(x^* - p)^T A(x^* - p) + a^T(x^* - p) \\ &= \frac{1}{2}(x^*)^T Ax^* + a^T x^* - (x^*)^T Ap + \frac{1}{2}p^T Ap - a^T p \\ &= f(x^*) - (x^*)^T Ap + \frac{1}{2}p^T Ap - a^T p. \end{aligned}$$
The solution x^*, λ^* of the linear system of equations fulfills $Ax^* + B^T\lambda^* = -a$ and we have $p \in \ker(B)$, due to $Bx^* = Bx = b$. By multiplying the first equation by p^T from left, we get
$$p^T Ax^* = -p^T a$$
and so
$$f(x) = f(x^*) - (x^*)^T Ap + \frac{1}{2}p^T Ap - a^T p = f(x^*) + \frac{1}{2}p^T Ap > f(x^*).$$

c) i) The Lagrange multiplier is unique, because by the LICQ condition it follows that the Lagrange multiplier λ^* is unique for every solution x^*, if $\nabla h(x^*) = B$ has full rank. This is true by assumption.

ii) The code can look like:

```
%-- (3) determine x_Z by minimizing f(Y x_y + Z z)
w = Y*x_Y;
z = fminunc(@(z)f(w + Z*z), z0, optimset('Display','off'));

% determine the full solution x
x = w + Z*z;

% determine the lagrange multiplier
lambda = -(B')\gradf(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

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \; & f^T x \\ \text{s.t. } & Bx = b, \\ & x \geq 0 \end{aligned} \tag{7}$$

a) Derive the dual problem of problem (7) and name the Lagrange multiplier for the inequality constraints s.

b) Derive the KKT conditions for problem (7).

c) The interior point method for the linear program(7) is for $x \geq 0$, $s \geq 0$ based upon the perturbed system $F_\mu(x, \lambda, s) = 0$ with

$$F_\mu(x, \lambda, s) := \begin{pmatrix} B^T\lambda + s - f \\ Bx - b \\ XSe - \mu e \end{pmatrix},$$

as well as $e := (1, \ldots, 1)^T \in \mathbb{R}^n$, $X := (\mathrm{diag}(x_i))_{i=1,\ldots,n} \in \mathbb{R}^{n\times n}$ and $S := (\mathrm{diag}(s_i))_{i=1,\ldots,n} \in \mathbb{R}^{n\times n}$. Bob asked **ChatGPT** about an implementation and he got the following code as an answer:

```
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);
    DF_0  = [I, B', zerOnn; B, zerOmm, X; zerOmn', S, zerOmn];
    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+n);
    Deltas       = delta_sol(n+n+1:end);

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

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

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

A test script for the code above revealed the following error messages:

```
Error using horzcat: Dimensions of arrays being concatenated are not consistent.
Error in InteriorPointMethod (line 31)
DF_0 = [I, B', zer0nn; B, zer0mm, X; zer0mn', S, zer0mn]
```

and also the following

```
Arrays have incompatible sizes for this operation.
Error in InteriorPointMethod (line 41)
alpha = eta / max([eta; Deltax./x; Deltas./s]);
```

Find the errors in the code and fix them:
Hint: There are errors in three lines of the code.

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

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

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

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

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];
Deltalambda = delta_sol(n+1:n+m);
Deltas      = delta_sol(n+m+1:end);
```

Last Name, First Name: ______________________________

Task 6 (Projected Subgradient Method) **(2+1+1+1 = 5 Points)**

In this task we consider the following non-smooth constraint optimization problem

$$\min_{x \in \mathcal{F} \subset \mathbb{R}^2} f(x), \tag{8}$$

where $f(x) := \max_{i \in \{1,2\}} x_i^2$ and $\mathcal{F}$ is convex.

a) Calculate the convex sub-differential $\partial f(x)$ for all $x \in \mathbb{R}^2$.
Hint: You can find the calculus rule below.

b) Why does a step of the projected subgradient method not always decrease the function value?

c) How could a projection $P_{\mathcal{F}}$ onto the set $\mathcal{F}$ look like?

d) In practice, one has to use a suboptimal step size σ_k. A heuristic is to require

$$(i)\ \sigma_k \to 0 \quad \text{and} \quad (ii)\ \sum_{k=0}^{\infty} \sigma_k = \infty.$$

Explain briefly why the second condition (ii) is posed.

Subdifferential calculus: Let f_i, $i = 1, \dots, m$ be convex functions and

$$f(x) := \max_{1 \le i \le m} f_i(x).$$

With the index set

$$I(x) := \{1 \le i \le m : f_i(x) = f(x)\}$$

it holds that

$$\partial f(x) = \operatorname{conv} \left\{ \bigcup_{i \in I(x)} \partial f_i(x) \right\},$$

where the convex hull is given by

$$\operatorname{conv}(X) := \left\{ \sum_{i=1}^{k} \lambda_i x_i : k \in \mathbb{N}, x_i \in X, \lambda_i \ge 0, \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$

Lösung:

a) The convex subdifferential of the functions $f_i(x) := x_i^2$, $i = 1, 2$ is given by $\partial f_i(x) = \{\nabla f_i(x) = 2x_i e_i\}$, where $e_i \in \mathbb{R}^2$ denotes the i-th canonical unit vector. Therefore, for the convex subdifferential of f it follows

$$\partial f(x) = \begin{cases} \{2x_i e_i\}, & \text{if } I(x) = \{i\}, \\ \{2\lambda_1 x_1 e_1 + 2\lambda_2 x_2 e_2 : \ \lambda_i \geq 0, \ \lambda_1 + \lambda_2 = 1\}, & \text{if } I(x) = \{1, 2\} \end{cases}.$$

b) The reason is that within the algorithm one chooses some subgradient $s^{(k)} \in \partial f(x^{(k)})$, which does not have to be a descent direction.

c) One can define for $x \in \mathbb{R}^2$

$$P_{\mathcal{F}} x := \arg\min_{z \in \mathcal{F}} \|x - z\|_2^2.$$

d) If the second condition were not posed, we would get $\|x^{(k)} - x^{(0)}\|_2 \leq \sum_{k=0}^{\infty} \sigma_k =: M$. If the initial vector $x^{(0)}$ is chosen in a way 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 !!!