Prof. Dr. Karsten Urban
M. Sc. Lewin Ernst

Winter term 2024/2025
Sheet 12, 28 Points

Numerical Optimization
Solution to exercise sheet

review on 22.01.2024 during the exercise class

1. *(SQP method for nonlinear programs with equality constraints)*
We consider the following nonlinear program with nonlinear equality constraints

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} & f(x), \quad f : \mathbb{R}^n \to \mathbb{R} \\ \text{s.t.} & h(x) = 0, \quad h : \mathbb{R}^n \to \mathbb{R}^m \end{cases} \tag{1}$$

and we derive the SQP method for this problem. But first we look at the so-called Lagrange-Newton method. If x^* is a local solution of the problem above and a CQ holds, the KKT conditions hold, i.e. there exists $\lambda^* \in \mathbb{R}^m$ with

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\ h(x^*) &= 0. \end{aligned}$$

To solve problem (1) it seems to be a good idea to solve the KKT system to determine (x^*, λ^*). We do this by applying Newton's method to the system

$$F(x, \lambda) := \begin{pmatrix} \nabla_x \mathcal{L}(x, \lambda) \\ h(x) \end{pmatrix} = 0$$

Therefore, let $x^{(k)}$ and $\lambda^{(k)}$ be iterates and let f as well as h be two times continuously differentiable. Then the Newton update is

$$\begin{pmatrix} \nabla^2_{xx} \mathcal{L}(x^{(k)}, \lambda^{(k)}) & (\nabla h(x^{(k)}))^T \\ \nabla h(x^{(k)}) & 0 \end{pmatrix} \begin{pmatrix} d_x^{(k)} \\ d_\lambda^{(k)} \end{pmatrix} = \begin{pmatrix} -\nabla_x \mathcal{L}(x^{(k)}, \lambda^{(k)}) \\ -h(x^{(k)}) \end{pmatrix} {}^1 \tag{2}$$

and we can update our iterates $x^{(k)}$ and $\lambda^{(k)}$.

a) Write down a pseudo algorithm which realizes the above procedure.
b) To apply the convergence theorem of Newton's method to your algorithm in a) we have to ensure, that the matrix of the Newton update (2) is regular. Therefore, prove that if $\nabla h(x)$ has full rank and $s^T \nabla^2_{xx} \mathcal{L}(x, \lambda) s > 0$ for all $s \in \mathbb{R}^n \setminus \{0\}$ with $\nabla h(x) s = 0$ holds, then the matrix of the Newton update (2) is regular.
c) Derive a quadratic program (QP) with affine linear equality constraints, such that the KKT system of this QP is (2).
d) Write down the SQP pseudo algorithm with your QP in c).

(2 + 4 + 6 + 2 = 14 Points)

Solution:

[1] We make here the agreement, that the Jacobian of $h : \mathbb{R}^n \to \mathbb{R}^m$ is given by the matrix $(\nabla h(x))_{i,j} = \partial_{x_j} h_i(x)$ for $1 \le i \le m, 1 \le j \le n$, i.e. $\nabla h(x) \in \mathbb{R}^{m \times n}$.

a) The algorithm could look as follows:

Algorithm 1 Lagrange-Newton Method

Require: Functions $\mathcal{L}$, $\nabla_x\mathcal{L}$, $\nabla^2_{xx}\mathcal{L}$, h, ∇h, initial value $x^{(0)} \in \mathbb{R}^n$ and $\lambda^{(0)} \in \mathbb{R}^m$;

$(x^{(k)}, \lambda^{(k)}) \leftarrow (x^{(0)}, \lambda^{(0)})$;

for $k = 0, 1, \dots$ **do**

  if $h(x^{(k)}) = 0$ and $\nabla_x\mathcal{L}(x^{(k)}, \lambda^{(k)}) = 0$ **then**

    Return $(x^{(k)}, \lambda^{(k)})$;

  end if

  Solve $\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x^{(k)}, \lambda^{(k)}) & (\nabla h(x^{(k)}))^T \\ \nabla h(x^{(k)}) & 0 \end{pmatrix} \begin{pmatrix} d_x^{(k)} \\ d_\lambda^{(k)} \end{pmatrix} = \begin{pmatrix} -\nabla_x\mathcal{L}(x^{(k)}, \lambda^{(k)}) \\ -h(x^{(k)}) \end{pmatrix}$;

  $(x^{(k+1)}, \lambda^{(k+1)}) \leftarrow (x^{(k)}, \lambda^{(k)}) + (d_x^{(k)}, d_\lambda^{(k)})$;

end for

return $(x^{(k+1)}, \lambda^{(k+1)})$;

---

b) We have seen this condition before when we introduced the nullspace method for a quadratic problem. So the proof is similiar to the one we did on sheet 9. We have to show

$$F'(x, \lambda) \begin{pmatrix} v \\ w \end{pmatrix} = 0 \qquad \Rightarrow \qquad v = 0 \wedge w = 0.$$

The second equation reads $\nabla h(x^{(k)})v = 0$, which we can use if we multiply the first equation by v^T from the right to receive

$$0 = v^T \nabla^2_{xx}\mathcal{L}(x^{(k)}v + v^T(\nabla h(x^{(k)}))^T w = v^T \nabla^2_{xx}\mathcal{L}(x^{(k)}v.$$

The latter equation would be strictly greater than zero, if we would had $v \neq 0$. Therefore, it must hold $v = 0$. From this we deduce again from the first equation $(\nabla h(x^{(k)}))^T w = 0$, but the matrix $h(x^{(k)}))^T$ has full column rank so that the only possibility to represent the zero is with coefficients $w = 0$.

c) In order to derive a QP we make the abbreviations $A := \nabla^2_{xx}\mathcal{L}(x^{(k)}$, $a := \nabla_x\mathcal{L}(x^{(k)}$, $B = \nabla h(x^{(k)})$ and $b := -h(x^{(k)})$. With that and section 3.6.1 of the lecture notes we can formulation the KKT conditions for the (QP) as

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} d_x^{(k)} \\ d_\lambda^{(k)} \end{pmatrix} = \begin{pmatrix} -a \\ b \end{pmatrix}$$

which is given in (3.6.3). The gradient of the Lagrange function of the (QP) is given by

$$\nabla\mathcal{L}_{QP}(d_x^{(k)}, d_\lambda^{(k)}) = \begin{pmatrix} Ad_x^{(k)} + a + B^T d_\lambda^{(k)} \\ Bd_x^{(k)} - b \end{pmatrix} \tag{3}$$

so that the gradient of our quadratic function is given by

$$\nabla q(d_x^{(k)}) = Ad_x^{(k)} + a.$$

Integrating this function gives

$$q(d_x^{(k)}) = \frac{1}{2}(d_x^{(k)})^T A d_x^{(k)} + a^T d_x^{(k)}$$

and the equality constraints are given in the second equation of (3). Thus, we have derived the following (QP)

$$\begin{cases} \min\limits_{d\in\mathbb{R}^n} & \frac{1}{2}d^T\nabla^2_{xx}\mathcal{L}(x^{(k)}d + (\nabla_x\mathcal{L}(x^{(k)})^T d \\ \text{s.t.} & \nabla h(x^{(k)})d = -h(x^{(k)}) \end{cases} \tag{4}$$

This equation looks similar to the (QP) in (3.7.3). If we define $d_\lambda^{(k)} := \tilde{\lambda}^{(k)} - \lambda^{(k)}$ and plug this into (3), we can write the resulting (QP) as

$$\begin{cases} \min\limits_{d\in\mathbb{R}^n} & \frac{1}{2}d^T\nabla^2_{xx}\mathcal{L}(x^{(k)}d + (\nabla_x f(x^{(k)})^T d \\ \text{s.t.} & \nabla h(x^{(k)})d = -h(x^{(k)}) \end{cases} \tag{5}$$

which is exactly the problem (3.7.3) if h would be affine linear and $x^{(k)} \in \mathcal{F}$. Make that clear to yourself.

d) The pseudo algorithm could now look like

Algorithm 2 SQP Method

Require: Functions $\mathcal{L}$, $\nabla_x\mathcal{L}$, $\nabla^2_{xx}\mathcal{L}$, h, ∇h, initial value $x^{(0)} \in \mathbb{R}^n$ and $\lambda^{(0)} \in \mathbb{R}^m$;

$(x^{(k)}, \lambda^{(k)}) \leftarrow (x^{(0)}, \lambda^{(0)})$;
for $k = 0, 1, \dots$ **do**
    if $h(x^{(k)}) = 0$ and $\nabla_x\mathcal{L}(x^{(k)}, \lambda^{(k)}) = 0$ **then**
        Return $(x^{(k)}, \lambda^{(k)})$;
    end if
    Solve the (QP) (4) or (5);
    If (4) is solved: $(x^{(k+1)}, \lambda^{(k+1)}) \leftarrow (x^{(k)}, \lambda^{(k)}) + (d_x^{(k)}, d_\lambda^{(k)})$;
    If (5) is solved: $x^{(k+1)} \leftarrow x^{(k)} + d_x^{(k)}$, $\lambda^{(k+1)} \leftarrow \tilde{\lambda}^{(k)}$;
end for
return $(x^{(k+1)}, \lambda^{(k+1)})$;

---

2. *(Interior point methods, QPs)*
Consider the quadratic program (QP)

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} & \frac{1}{2}x^T Ax + a^T x, \\ \text{s.t.} & Bx = b, \\ & Cx \leq c, \end{cases}$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric positive semi-definite and $B \in \mathbb{R}^{m\times n}$ has full rank. Derive an analogue of the Newton correction (4.1.3).

(4 Points)

Solution: Let us first derive the system (3.3.3) from the lecture notes for Example (3.3.1) to see what we need to change when equality constraints are added. We rewrite the minimization problem (3.3.2), due to the fact that we introduced a slack variable s and now minimizing w.r.t.

$(x^T, s^T)^T$. Thus we blow up the dimensions and write

$$\begin{cases} \min\limits_{(x^T,s^T)\in\mathbb{R}^{n+p}} & \frac{1}{2}(x^T, s^T)\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} + \begin{pmatrix} a^T & 0 \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix}, \\ \text{s.t.} & \begin{pmatrix} C & I \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} - c = 0, \\ & \begin{pmatrix} 0 & -I \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} \leq 0, \end{cases}$$

We derive the KKT system for that problem

$$\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} + \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} C^T \\ I^T \end{pmatrix}\lambda + \begin{pmatrix} 0 \\ -I \end{pmatrix}\mu = 0 \tag{6}$$

$$\begin{pmatrix} C & I \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} - c = 0 \tag{7}$$

$$\mu \geq 0 \tag{8}$$

$$\begin{pmatrix} 0 & -I \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} \leq 0 \tag{9}$$

$$\mu^T \begin{pmatrix} 0 & -I \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} = 0 \tag{10}$$

This is a bit overblown but exactly the KKT conditions for the (QP) above. We can rewrite the equations of the system and we get

$$\begin{aligned} Ax + a + C^T\lambda &= 0 \\ \lambda &= \mu \\ Cx + s - c &= 0 \\ \mu &\geq 0 \\ -s &\leq 0 \\ \mu_i s_i &= 0, \quad \forall i = 1, \dots, p \end{aligned}$$

We can use the equation $\lambda = \mu$ to receive exactly (3.3.3)

$$\begin{aligned} Ax + a + C^T\lambda &= 0 \\ Cx + s - c &= 0 \\ \lambda &\geq 0 \\ s &\geq 0 \\ \lambda_i s_i &= 0, \quad \forall i = 1, \dots, p \end{aligned}$$

From this system we get the function (4.1.2). Now the only thing which is affected by the introduction of equality constraints $Bx = b$ is (6) and (7). We get for those equations

$$\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} + \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} B^T & C^T \\ 0 & I^T \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} + \begin{pmatrix} 0 \\ -I \end{pmatrix}\mu = 0 \tag{11}$$

$$\begin{pmatrix} B & 0 \\ C & I \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} - \begin{pmatrix} b \\ c \end{pmatrix} = 0 \tag{12}$$

The second equation in (11) states $\lambda_2 = \mu$. Thus we get the system

$$\begin{aligned} Ax + a + B^T\lambda_1 + C^T\lambda_2 &= 0 \\ Bx - b &= 0 \\ Cx + s - c &= 0 \\ \lambda_2 &\geq 0 \\ s &\geq 0 \\ (\lambda_2)_i s_i &= 0, \quad \forall i = 1, \dots, p \end{aligned}$$

The analogous function to (4.1.2) is with the complementarity measure $\mu = (\lambda_2)^T s/p$ defined as

$$F_\mu(x, \lambda_1, \lambda_2, s; \sigma) := \begin{pmatrix} Ax + a + B^T\lambda_1 + C^T\lambda_2 \\ Bx - b \\ Cx + s - c \\ S\Lambda e - \sigma\mu e \end{pmatrix} = 0$$

Applying Newton's methods leads us to the Newton correction (note that μ is kept fix, when we calculate the gradient)

$$\nabla F_\mu(x, \lambda_1, \lambda_2, s; \sigma) \begin{pmatrix} \Delta x \\ \Delta\lambda_1 \\ \Delta\lambda_2 \\ \Delta s \end{pmatrix} = \begin{pmatrix} A & B^T & C^T & 0 \\ B & 0 & 0 & 0 \\ C & 0 & 0 & I \\ 0 & 0 & S & \Lambda \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda_1 \\ \Delta\lambda_2 \\ \Delta s \end{pmatrix} = -F_\mu(x, \lambda_1, \lambda_2, s; \sigma)$$

3. *(Central Path)*
Consider the convex quadratic problem

$$\begin{cases} \min\limits_{x\in\mathbb{R}^2} & f(x) := \frac{1}{2}x_1^2 + x_2 \\ \text{s.t.} & g_1(x) := -x_1 \leq 0, \; g_2(x) := -x_2 \leq 0. \end{cases} \tag{13}$$

a) Write down the *barrier problem* (4.2.2). Moreover, calculate the gradient and the Hessian matrix of the objective function for this problem.
b) Calculate the solution x_μ of the barrier problem and sketch the central path.
c) Prove, that $x^* = \lim_{\mu\to 0^+} x_\mu$ exists and solves (13).
d) In how far is the central path changing, if the constraint $-x_2 \leq 0$ of problem (13) occures M-times.

(4 + 2 + 2 + 2 = 10 Points)

Solution:

a) The barrier problem is given by:

$$(B) \left\{ \varphi(x; \mu) := \tfrac{1}{2}x_1^2 + x_2 - \tfrac{1}{\mu}\left(\log(x_1) + \log(x_2)\right) \to \min \right. .$$

For $x_1, x_2 > 0$ the gradient $\nabla\varphi(x;\mu)$ and the Hessian $\nabla^2\varphi(x;\mu)$ is given by:

$$\nabla\varphi(x;\mu) = \begin{pmatrix} x_1 - \frac{1}{\mu x_1} \\ 1 - \frac{1}{\mu x_2} \end{pmatrix}, \qquad \nabla^2\varphi(x;\mu) = \begin{pmatrix} 1 + \frac{1}{\mu x_1^2} & 0 \\ 0 & \frac{1}{\mu x_2^2} \end{pmatrix}.$$

b) The Hessian $\nabla^2\varphi(x;\mu)$ is positive definite on $I := (0,\infty)\times(0,\infty)$. Thus $\varphi(x;\mu)$ is strictly convex on I and therefore there is only one global minimum x_μ^*. The minimum x_μ^* fulfills

$$\nabla\varphi(x_\mu^*;\mu) = \begin{pmatrix} x_1^* - \frac{1}{\mu x_1^*} \\ 1 - \frac{1}{\mu x_2^*} \end{pmatrix} \stackrel{!}{=} 0,$$

woraus sich für $\mu > 0$ der zentrale Pfad $x_\mu^* = (1/\sqrt{\mu}, 1/\mu)^T$ ergibt. Mit der Substitution $\tau := 1/\sqrt{\mu}$ erhält man $x_\mu^* = (\tau, \tau^2)^T$. Also ist der zentrale Pfad der rechte Ast einer Normalparabel.

c) Es gilt $\lim_{\mu\to\infty} x_\mu^* = 0$. Offensichtlich ist $x^* = (0,0)^T$ Lösung von (13), da die Zielfunktion *auf dem zulässigen Bereich* nach unten durch Null beschränkt ist und x^* den minimalen Zielfunktionswert 0 liefert.

d) Das Barriereproblem ist nun gegeben durch:

$$(B)\left\{\varphi(x;\mu) := \tfrac{1}{2}x_1^2 + x_2 - \tfrac{1}{\mu}\left(\log(x_1) + M\log(x_2)\right) \to \min\right. .$$

Für $x_1, x_2 > 0$ sind $\nabla\varphi(x;\mu)$ bzw. $\nabla^2\varphi(x;\mu)$ nun gegeben durch:

$$\nabla\varphi(x;\mu) = \begin{pmatrix} x_1 - \frac{1}{\mu x_1} \\ 1 - \frac{M}{\mu x_2} \end{pmatrix}, \qquad \nabla^2\varphi(x;\mu) = \begin{pmatrix} 1 + \frac{1}{\mu x_1^2} & 0 \\ 0 & \frac{M}{\mu x_2^2} \end{pmatrix}.$$

Analog zu c) erhlät man die Lösung $x_\mu^* = (1/\sqrt{\mu}, M/\mu)^T$, bzw. mit Hilfe der Substitution $\tau := 1/\sqrt{\mu}$ die Lösung $x_\mu^* = (\tau, M\tau^2)^T$. Damit ist der zentrale Pfad wieder der rechte Ast einer Parabel, welche durch M gestaucht wird. Für sehr große M verläuft der zentrale Pfad nun entlang der x_2-Achse, wodurch für Pfad-Verfolgungsstrategien nur kleine Umgebungen um den zentralen Pfad verwendet werden können.