Optimalisering og regulering TTK4135 - Assignment 8

Ola Kristoffer Hoff

21st March, 2023

1 Second-Order Necessary Conditions

a Formulate Theorem 2.3 (page 15) as stated in the textbook.

As stated in the book, this is the formulation of the theorem: "If x^* is a local minimizer of f and $\nabla^2 f$ exists and is continuous in an open neighborhood of x^*, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite." [1] (page. 15, Theorem 2.3 (Second-Order Necessary Conditions)).

b Go through the proof so that you understand it. Briefly explain the proof.

The proof bases itself on contradiction. We assume that we have an optimal point where $\nabla^2 f(x^*)$ is not positive semidefinite. We know from another theorem (2.2) that $\nabla f(x^*) = 0$. Then it shows that we then have to have a direction p such that in all directions from x^*, the neighbourhood, we have to have a decreasing direction. Hence, it's not a minimiser. So under the case where the Hessian is positive semidefinite we get increasing directions in all directions, hence it is a minimiser.

c Compare Theorem 2.3 and Theorem 2.4 (page 16). Why does Theorem 2.3 not give sufficient conditions for a strict local minimum?

The difference between the necessary (2.3) and sufficient (2.4) condition theorems are that the first is positive semidefinite while the second one is definite. Since it is not strictly positive, it includes zero, it could be a flat plane too. A strict local minimiser is a point in which all values in it's neighbourhood are greater then it. The semidefinite property allows for equal values.

2 The Newton Direction

Consider the model function m_k based on the second-order Taylor approximation (see equation (2.14) in the textbook):

$$m_k(p) := f_k + p^T \nabla f_k + \frac{1}{2} p^T \nabla^2 f_k p \approx f(x_k + p) \tag{1}$$

a Derive the Newton direction

$$p_k^N = -(\nabla^2 f_k)^{-1} \nabla f_k \tag{2}$$

using the model function m_k.

"Assuming for the moment that $\nabla^2 f_k$ is positive definite, we obtain the Newton direction by finding the vector p that minimizes $m_k(p)$. By simply setting the derivative of $m_k(p)$ to zero, we obtain the following explicit formula:" [1] (page. 22).

We do as stated in the book, solve it by setting the model functions derivative equal to zero:

$$\begin{aligned} \frac{dp}{dx} m_k(p) &= 0 \\ \nabla f_k + \nabla^2 f_k p_k^N &= 0 \\ p_k^N &= -(\nabla^2 f_k)^{-1} \nabla f_k \end{aligned}$$

b Assume that $\nabla^2 f_k$ is not positive definite. In this case, is the Newton-direction p_k^N a descent direction? Is it even defined? Explain.

If it is not positive definite we loose the property of guaranteed invertibility. Hence, the inverse might not exist, and therefore the direction too might not exist. Let's say it is defined, then it is no guarantee that it is a descent direction, since this too came as a property from the positive definiteness.

The book states the following regarding the direction in this case: "In these situations, line search methods modify the definition of p_k to make it satisfy the descent condition while retaining the benefit of the second-order information contained in $\nabla^2 f_k$."[1].

c Given an unconstrained minimization problem with objective function

$$f(x) = \frac{1}{2}x^T Gx + x^T c \tag{3}$$

with $G = G^T > 0$ and $x \in \mathbb{R}^n$. Show that an iteration algorithm based on the Newton direction (i.e., $x_{k+1} = x_k + p_k^N$) always converges to the optimum in one step.

The properties of G guarantees that it is invertable. It also gives us a strictly convex problem. We can thus derive the following:

$$\begin{aligned}
\nabla f(x) &= Gx_k + c \\
\nabla^2 f(x) &= G \\
p_k^N &= -(\nabla^2 f_k)^{-1} \nabla f_k \\
p_k^N &= -(G)^{-1}(Gx_k + c) \\
p_k^N &= -x_k - G^{-1}c \\
x_{k+1} &= x_k + p_k^N \\
x_{k+1} &= x_k - x_k - G^{-1}c \\
x_{k+1} &= -G^{-1}c
\end{aligned}$$

As we can see the next position x_{k+1} is independent of the previous step x_k. This means that we are going directly to the final place we will go. And since we know that we are going towards a minimum we end up there. We also note that $-G^{-1}c$ is the definition of minimiser of the problem for x^*. This follows from:

$$\begin{aligned}
\nabla f(x^*) &= Gx^* + c = 0 \\
Gx^* &= -c \\
x^* &= -G^{-1}c
\end{aligned}$$

Hence, we have shown that given an unconstrained problem, the Newton direction converges to the solution in one step.

d Show that

$$f(x) = \frac{1}{2}x^T Gx + x^T c, \quad x \in X, \quad X = \{x \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq 1\} \tag{4}$$

where

$$G = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{5}$$

is a convex function.

Here we have that $G = G^T > 0$ which, as noted in the previous subsection, gives us a strictly convex problem. We can show why:

There are three conditions for convexity[1](page. 8):

- the objective function is convex,
- the equality constraint functions $c_i(\cdot),\ \ i \in E$, are linear, and
- the inequality constraint functions $c_i(\cdot),\ \ i \in I$, are concave.

"The term "convex" can be applied both to sets and to functions. A set $S \in \mathbb{R}^n$ is a convex set if the straight line segment connecting any two points in S lies entirely inside S. Formally, for any two points $x \in S$ and $y \in S$, we have $\alpha x + (1-\alpha)y \in S$ for all $\alpha \in [0,1]$. The function f is a convex function if its domain S is a convex set and if for any two points x and y in S, the following property is satisfied:"[1](page. 8).

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y),\ \ \forall \alpha \in [0,1] \tag{6}$$

The last two conditions are obviously fulfilled since we don't have any constraints. Showing that the function is convex:

$$\begin{aligned}
f(\alpha x + (1-\alpha)y) &\leq \alpha f(x) + (1-\alpha)f(y) \\
&\leq \alpha(\frac{1}{2}x^TGx + x^Tc) + (1-\alpha)(\frac{1}{2}y^TGy + y^Tc) \\
&\leq \alpha\frac{1}{2}x^TGx + \alpha x^Tc + (1-\alpha)\frac{1}{2}y^TGy + (1-\alpha)y^Tc \\
&\frac{1}{2}(\alpha x + (1-\alpha)y)^TG(\alpha x + (1-\alpha)y) + (\alpha x + (1-\alpha)y)^Tc \leq \\
&\frac{1}{2}\alpha^2x^TGx + \frac{1}{2}\alpha(1-\alpha)x^TGy + \frac{1}{2}(1-\alpha)\alpha y^TGx + \frac{1}{2}(1-\alpha)^2y^TGy + \alpha x^Tc + (1-\alpha)y^Tc \leq \\
&\frac{1}{2}\alpha^2x^TGx + (1-\alpha)\alpha x^TGy + \frac{1}{2}(1-\alpha)^2y^TGy + \alpha x^Tc + (1-\alpha)y^Tc \leq \\
&\frac{1}{2}\alpha^2x^TGx + (1-\alpha)\alpha x^TGy + \frac{1}{2}(1-\alpha)^2y^TGy - \alpha\frac{1}{2}x^TGx - (1-\alpha)\frac{1}{2}y^TGy \leq 0 \\
&\qquad \vdots \\
&-\frac{1}{2}\alpha(1-\alpha)(x-y)^TG(x-y) \leq 0
\end{aligned}$$

Now to prove that the last equation holds. Since $\alpha \in [0,1] \Rightarrow \alpha \geq 0 \wedge (1-\alpha) \geq 0$. As stated G is positive definite giving us $(x-y)^TG(x-y) > 0$ too as long as $x - y \neq 0$. However, when they are zero the equation zeros out and still holds. Since all terms in the equation is positive we get a negative value due to the minus at the start, hence it is always less then or equal to zero. Thus, the function is proven convex.

3 The Rosenbrock Function

Solve problem 2.1 in the textbook.

Compute the gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ of the Rosenbrock function

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \tag{7}$$

Show that $x^* = (1,1)^T$ is the only local minimizer of this function, and that the Hessian matrix at that point is positive definite.

Deriving the gradient and the Hessian:

$$\begin{aligned}
\nabla f(x) &= \begin{bmatrix} -200(x_2 - x_1^2)2x_1 - 2(1-x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} \\
\nabla^2 f(x) &= \begin{bmatrix} -400(x_2 - 3x_1^2) + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}
\end{aligned}$$

Finding the minimiser is done by setting the gradient equal to zero:

$$
\begin{aligned}
\nabla f(x) = \begin{bmatrix} -200(x_2 - x_1^2)2x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} &= 0 \\
\begin{bmatrix} -200(x_2 - x_1^2)2x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} &= 0 \\
\Rightarrow x_1^2 &= x_2 \\
\begin{bmatrix} 0 - 2(1 - x_1) \\ 0 \end{bmatrix} &= 0 \\
\Rightarrow x_1 &= 1 \\
\Rightarrow x_1 = 1 \wedge x_2 &= 1 \\
\Rightarrow x^* &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\end{aligned}
$$

We observed above that the second equation in the gradient constrained the relation between x_1 and x_2. Applying this constraint in we see that x_1 must be equal to 1 and from there we derive that the only valid minimiser is x^*.

Now to check that the Hessian is positive definite in this point.

$$
\begin{aligned}
\nabla^2 f(x^*) &= \begin{bmatrix} -400(x_2 - 3x_1^2) + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix} \\
\nabla^2 f(x^*) &= \begin{bmatrix} -400(1 - 3) + 2 & -400 \\ -400 & 200 \end{bmatrix} \\
\nabla^2 f(x^*) &= \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} \\
\lambda(\nabla^2 f(x^*)) &= \begin{bmatrix} 0.4 \\ 1001.6 \end{bmatrix}
\end{aligned}
$$

As we can see both eigenvalues of the Hessian is positive and greater than zero, hence the Hessian is positive definite in x^*.

References

[1] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Science+Business Media, LLC, 2006.