Exercise 9

1) a) If a function is convex, then every minimum of that function x^* is a global minimum. If we want to minimize a function, we need a convex function.

b) A smooth function has continous derivatives up to some desired order for the domain. This is important when finding the minimums.

2) $g: \mathbb{R} \to [0, \infty)$
$g(x) = |x|$

A function is convex if and only if

$\alpha f(x^a) + (1-\alpha) f(x^b) \geq f(\alpha x^a + (1-\alpha) x^b)$ for all $\alpha \in [0,1]$

Since we have the interval $[0, \infty)$, our function will be both smooth and convex, as the secant lies on top of the function.

b) will not be convex as it oscillates (sinus-properties). The function is smooth, though.

3) $f: \mathbb{R}^n \to \mathbb{R}$
$f(x) = x^T A x + b^T x + c$, $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$

In order for the function to be convex, we need the $\nabla f = 0$ and $\nabla^2 f$ positive definite.

We have that $\nabla^2 f = 2A$, and therefore we need A to be positive definite.

4) $f: \mathbb{R}^n \to \mathbb{R}$
twice continously derivative

$\Rightarrow$ Dimensions of ∇f will therefore be $1 \times n$ and of $\nabla^2 f$ $n \times n$.

5) $f: \mathbb{R}^n \to \mathbb{R}$
local minimum x^*

$\Rightarrow \underline{\nabla f(x^*) = 0}$

We also know that $\underline{\nabla^2 f \geq 0}$ (positive definite)

6) $f: \mathbb{R}^n \to \mathbb{R}$
$\nabla f(x^*) = 0$

This does not necessarily mean that x^* is a local minimum. It can also be a saddle point or a local maxima. We need the Hessian to be positive definite.

2) $x_{k+1} = x_k - \gamma \nabla f(x)$

$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} (\nabla f(x_k))$
$p_k = -(\nabla^2 f(x_k))^{-1} (\nabla f(x_k))$
$(\nabla^2 f(x_k)) p_k = -\nabla f(x)$

1) The steepest descent gives a step in the opposite direction of the gradient, which means that in each point it will go in the direction that decreases the most. This results in finding the minimum.

2) The steepest descent method only needs to calculate the gradient of the function, which results in a lot less computation per step. However, this results in needing more steps. The Newton method calculates the curvature of the function which gives more information about the "direction" towards the minima. Therefore, this method requires alot of computations per step, but as a consequence will require much less steps.

3) The gradient close to a saddle point will be very small. This means that the algorithm might be "stuck" around this point and it might therefore oscillate around this point.

4) It is not guaranteed that the Hessian is positive definite. The function might be concave, which will result in a negative definite. There might also be a problem in the local area of a saddle point

This might implicate that the solver will use a lot more steps as the next steps are not necessarily closer to the minima. Another problem might be that we won't know what type of point we will end up with.