Numerical Optimization - Assignment 5

1 Theoretical

1.1 Q1

Here we will will show that Newtons Method is invariant to all invertible linear transformations of the objective function.

We will Assume that Newtons Method with backtracking line-search is run on f with starting point x_0 producing a sequence of iterates $x_1, ..., x_T$ with step-candidates $p_1, ..., p_T$. Further we assume we run the algorithm on $g(x) = f(Ax)$ where A is square and invertible and we choose as starting point $\tilde{x}_0 = A^{-1}x_0$ leading to iteratex $\tilde{x}_1, ..., \tilde{x}_T$ with step candidates $\tilde{p}_1, ..., \tilde{p}_T$.

1. Show that the hessian is $\nabla^2 g(x) = A^T[\nabla^2 f(Ax)]A$

$$\begin{aligned} g(x) &= f(Ax) \\ \nabla g(x) &= A^T \nabla f(Ax) \\ \nabla^2 g(x) &= A^T[\nabla^2 f(Ax)]A \end{aligned}$$

2. Assume that $\tilde{x}_k = A^{-1}x_k$. Show that $\tilde{p}_k = A^{-1}p_k$
The search direction for Newton's method is $-\nabla^2 f^{-1} \nabla f$ so we have that

$$\begin{aligned} \tilde{p}_k &= -\nabla^2 f^{-1} \nabla f \\ &= -\nabla^2 g(\tilde{x}_k)^{-1} \nabla g(\tilde{x}_k) \\ &= -(A^T[\nabla^2 f(x_k)]A)^{-1} A^T \nabla f(x_k) \\ &= -A^{-1}[\nabla^2 f(x_k)]^{-1}(A^T)^{-1} A^T \nabla f(x_k) \\ &= A^{-1} \cdot -\nabla^2 f(x_k)^{-1} \nabla f(x_k) \\ &= A^{-1} p_k \end{aligned}$$

3. Show that $g(\tilde{x}_k + a\tilde{p}_k) = f(x_k + ap_k)$ and conclude that step-lengths a_k remain the same.
From the previous step we have that $\tilde{p}_k = A^{-1}p_k$ and $\tilde{x}_k = A^{-1}x_k$.

$$
\begin{aligned}
g(\tilde{x}_k + a\tilde{p}_k) &= f(A\tilde{x}_k + Aa\tilde{p}_k) \\
&= AA^{-1}x_k + AaA^{-1}p_k \\
&= f(x_k + ap_k)
\end{aligned}
$$

Conclude that step-lenghts a_k remain the same (TODO)? We have the Wolfe conditions (expression 3.4) given as:

$$
f(x_k + ap_k) \leq f(x_k) + c_1 a \nabla f_k^T p_k
$$

4. Conclude that assumption $\tilde{x}_k = A^{-1}x_k$ holds by induction, thus $g(\tilde{x}_k) = f(x_k)$

To show that this hold by induction, we first know that it holds for k=0 from the starting point $\tilde{x}_0 = A^{-1}x_0$.

We can show that it holds for any k

$$
\begin{aligned}
g(\tilde{x}_k) &= g(A^{-1}x_k) \\
&= f(AA^{-1}x_k) \\
&= f(x_k)
\end{aligned}
$$

5. Discuss the implications of this result. To what degree does BFGS have this property?

The implications of this result is that Newtons method is invariant to all invertible linear transformations of the objcetive functions, thus invertible linear transformations, such as scaling, will not have an effect on the performance of the method.

The weighted Frobenius norm makes BFGS scale-invariant, the weighted Frobenius norm is given as:

$$
||A||_W \equiv ||W^{1/2}AW^{1/2}||_F
$$