Continuous Optimization: Assignment 6

Due on June 4, 2024

Honglu Ma **Hiroyasu Akada** **Mathivathana Ayyappan**

Exercise 1

The strong Wolfe condition states that for some $\eta \in (\gamma, 1), \gamma \in (0, 1)$, the following holds:

$$\left|\langle \nabla f(x^{(k)} + \tau_k d^{(k)}), d^{(k)} \rangle\right| \leq \eta \left|\langle \nabla f(x^{(k)}), d^{(k)} \rangle\right|$$

We know the iterative update step for $x^{(k+1)}$ is defined as: $x^{(k+1)} = x^{(k)} + \tau_k d^{(k)}$. The strong curvature condition can be rewritten as such:

$$\left|\langle \nabla f(x^{(k+1)}), d^{(k)} \rangle\right| \leq \eta \left|\langle \nabla f(x^{(k)}), d^{(k)} \rangle\right|$$

By the definition of descent direction, $\langle \nabla f(x^{(k)}), d^{(k)} \rangle < 0$ and $\eta > 0$, we get

$$\begin{aligned} \langle \nabla f(x^{(k+1)}), d^{(k)} \rangle &\geq \eta \langle \nabla f(x^{(k)}), d^{(k)} \rangle \\ \langle \nabla f(x^{(k+1)}), d^{(k)} \rangle - \langle \nabla f(x^{(k)}), d^{(k)} \rangle &\geq \eta \langle \nabla f(x^{(k)}), d^{(k)} \rangle - \langle \nabla f(x^{(k)}), d^{(k)} \rangle \\ \langle \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), d^{(k)} \rangle &\geq (\eta - 1) \langle \nabla f(x^{(k)}), d^{(k)} \rangle > 0 \end{aligned}$$

We know $\tau_k > 0$

$$\begin{aligned} \langle \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), \tau_k d^{(k)} \rangle &> 0 \\ \langle \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), x^{(k+1)} - x^{(k)} \rangle &> 0 \\ \langle y^{(k)}, s^{(k)} \rangle &> 0 \end{aligned}$$

Exercise 2

The secant equation is given by $B_{k+1} s^{(k)} = y^{(k)}$ which is a system of n linear equations (assume the dimension is n). The choice of B_{k+1} is constrained by these n equations which results in a degree of freedom of n. On the other hand, the curvature condition:

$$\begin{aligned} \langle s^{(k)}, B_{k+1} s^{(k)} \rangle &= \langle s^{(k)}, y^{(k)} \rangle \\ \langle s^{(k)}, B_{k+1} s^{(k)} \rangle - \langle s^{(k)}, y^{(k)} \rangle &= 0 \\ \langle s^{(k)}, B_{k+1} s^{(k)} - y^{(k)} \rangle &= 0 \end{aligned}$$

It can be satisfied not only by setting $B_{k+1} s^{(k)} - y^{(k)} = 0$ which is the same as the secant equation, but also by setting $B_{k+1} s^{(k)} - y^{(k)}$ to be orthogonal to $s^{(k)}$. This gives more degree of freedom of choosing B_{k+1}.