Online Lecture Notes

Prof. Boris Houska

May 11, 2022

1 Newton Type Methods For Unconstrained Optimization

Our goal is to solve the optimization problem

$$\min_x \; F(x)$$

for a twice Lipschitz-continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}$. Local minimizers satisfy the stationarity condition

$$\nabla F(x) = 0 \; .$$

Here, ∇F denotes the gradient of F. The corresponding second order sufficient condition for optimality is given by

$$\nabla^2 F(x) \succ 0,$$

where "$\succ$" denotes that the Hessian of F is (symmetric and) positive definite. If we apply a Newton type method to the equation

$$\nabla F(x) = 0$$

we obtain an iteration of the form

$$x_{k+1} = x_k - M(x_k)^{-1} \nabla F(x_k)$$

where $M(x_k) \approx \nabla^2 F(x_k)$ is called a Hessian approximation. Since the Hessian of F is symmetric, we also choose a symmetric M for implementing this iteration. In particular, since we expect that the Hessian is positive definite in a neighborhood of the minimizer, it makes sense to choose symmetric positive definite Hessian approximations $M(x_k)$, such that

$$M(x_k) = M(x_k)^\intercal \quad \text{and} \quad M(x_k) \succ 0 \; .$$

This ensures in particular that $M(x_k)$ is invertible, such that the Newton type iterations remain well-defined.

1.1 Relation to Sequential Quadratic Programming Methods

In this section, we will show that the Newton type iteration of the form

$$x_{k+1} = x_k - M(x_k)^{-1} \nabla F(x_k)$$

can also be interpreted as a sequential quadratic programming (SQP) step of the form

$$x_{k+1} = x_k + \Delta x_k \qquad \text{where} \qquad \Delta x_k = \underset{\Delta x_k}{\operatorname{argmin}} \, \frac{1}{2} \Delta x_k^\intercal M(x_k) \Delta x_k + \nabla F(x_k)^\intercal \Delta x_k$$

This is interesting the sense that the quadratic objective

$$\frac{1}{2} \Delta x_k^\intercal M(x_k) \Delta x_k + \nabla F(x_k)^\intercal \Delta x_k$$

can be interpreted as an approximate second order Taylor expansion of F in the sense that

$$F(x_k + \Delta x_k) \approx F(x_k) + \left[\frac{1}{2} \Delta x_k^\intercal M(x_k) \Delta x_k + \nabla F(x_k)^\intercal \Delta x_k \right] .$$

Notice that we need to assume that $M(x_k)$ is positive definite such that the above quadratic objective is strictly convex. In order to verify this, notice that the optimality condition for the SQP step is given by

$$\begin{aligned} 0 &= \nabla_{\Delta x_k} \left[\frac{1}{2} \Delta x_k^\intercal M(x_k) \Delta x_k + \nabla F(x_k)^\intercal \Delta x_k \right] && (1) \\ &= M(x_k) \Delta x_k + \nabla F(x_k) && (2) \end{aligned}$$

If we solve this equation with respect to Δx_k, we find

$$\Delta x_k = -M(x_k)^{-1} \nabla F(x_k)$$

which implies that

$$x_{k+1} = x_k + \Delta x_k = x_k - M(x_k)^{-1} \nabla F(x_k),$$

which corresponds to the above Newton type iteration. In summary, we have shown that for positive definite Hessian approximation $\nabla^2 F(x_k) \approx M(x_k) \succ 0$, the SQP method is equivalent to a Newton type iteration (for unconstrained optimization).