Online Lecture Notes

Prof. Boris Houska

May 12, 2022

1 Line-search for Newton's method in the context of unconstrained optimization

Recall that our goal is to solve the minimization problem

$$\min_x F(x)\,,$$

where $F : \mathbb{R}^n \to \mathbb{R}$ is twice Lipschitz continuously differentiable and radially unbounded. This means that

$$\lim_{\|x\|\to\infty} F(x) \ = \ \infty\ .$$

Notice that second assumption on the radial unboundedness is need to examples like $F(x) = e^{-x}$, which would be smooth, but the local minimizer is at "$x = \infty$". Notice that the above conditions ensure that F is bounded from below and, consequently, has at least one local minimizer. We are interested in computing search directions by using a Newton type method of the form

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

but we only allow positive definite Hessian approximation $M(x_k) \succ 0$. In order to formally prove convergence we make the assumption that

$$M(x_k) = M(x_k)^\intercal \quad \text{and} \quad \underline{\sigma} I \ \preceq \ M(x_k) \ \preceq \ \overline{\sigma} I$$

for some (uniform) constants $0 < \underline{\sigma} \leq \overline{\sigma} < \infty$. Our actual line-search based update is given by

$$x_{k+1} = x_k + \alpha_k \Delta x_k\ ,$$

where $\alpha_k \in (0,1]$ is called a line search parameter. There are various heuristics for choosing α, for example, we could do exact line search by minimizing

$$\alpha_k = \operatorname*{argmin}_{\alpha_k \in [0,1]} F(x_k + \alpha_k \Delta x_k)$$

since this ensures $F(x_{k+1}) < F(x_k)$ whenever $\nabla F(x_k) \neq 0$. In order to prove the latter claim, we check that

$$F(x_k + \alpha_k \Delta x_k) \leq F(x_k) + \alpha_k F'(x_k)\Delta x_k + \frac{\omega}{2}\|\Delta x_k\|_2^2 \alpha_k^2$$

where ω is the Lipschitz bound on the second derivative of F. If we substitute the above step direction, we find

$$\begin{aligned} F(x_k + \alpha_k \Delta x_k) &\leq F(x_k) - \alpha_k F'(x_k) M(x_k)^{-1} \nabla F(x_k) + \frac{\omega}{2} \|\Delta x_k\|_2^2 \alpha_k^2 \\ &= F(x_k) - \alpha_k \nabla F(x_k)^\intercal M(x_k)^{-1} \nabla F(x_k) + \frac{\omega}{2} \|\Delta x_k\|_2^2 \alpha_k^2 \\ &\leq F(x_k) - \alpha_k \frac{\|\nabla F(x_k)\|_2^2}{\overline{\sigma}} + \frac{\omega}{2} \|\Delta x_k\|_2^2 \alpha_k^2 \, . \end{aligned} \tag{1}$$

If we run an exact line search, we are minimizing $F(x_k + \alpha_k \Delta x_k)$ over α_k, which means that we can take the minimum over α_k on both sides of the above inequality. This yields

$$\begin{aligned} F(x_{k+1}) &= \min_{\alpha_k} F(x_k + \alpha_k \Delta x_k) \\ &\leq \min_{\alpha_k} \left[F(x_k) - \alpha_k \frac{\|\nabla F(x_k)\|_2^2}{\overline{\sigma}} + \frac{\omega}{2} \|\Delta x_k\|_2^2 \alpha_k^2 \right] \end{aligned}$$

In order to find the minimum of the right hand expression, we have to solve the stationarity equation

$$0 = -\frac{\|\nabla F(x_k)\|_2^2}{\overline{\sigma}} + \omega \|\Delta x_k\|_2^2 \alpha_k^\star \qquad \Longrightarrow \qquad \alpha_k = \frac{\|\nabla F(x_k)\|_2^2}{\overline{\sigma} \omega \|\Delta x_k\|_2^2}$$

By substituting $\alpha^\star$, we find that

$$\begin{aligned} F(x_{k+1}) &\leq \min_{\alpha_k} \left[F(x_k) - \alpha_k \frac{\|\nabla F(x_k)\|_2^2}{\overline{\sigma}} + \frac{\omega}{2} \|\Delta x_k\|_2^2 \alpha_k^2 \right] \\ &= F(x_k) - \frac{1}{2} \frac{\|\nabla F(x_k)\|_2^4}{\overline{\sigma}^2 \omega \|\Delta x_k\|_2^2} \end{aligned} \tag{2}$$

Additionally, we can use that

$$\|\Delta x_k\|_2^2 = \|M(x_k)^{-1} \nabla F(x_k)\|_2^2 \geq \frac{1}{\overline{\sigma}^2} \|\nabla F(x_k)\|_2^2 \, .$$

By substituting this inequality, we find that

$$\begin{aligned} F(x_{k+1}) &\leq F(x_k) - \frac{1}{2} \frac{\|\nabla F(x_k)\|_2^4}{\overline{\sigma}^2 \omega \|\Delta x_k\|_2^2} \\ &\leq F(x_k) - \frac{1}{2\omega} \|\nabla F(x_k)\|_2^2 \end{aligned} \tag{3}$$

The above inequality implies that

$$F(x_k) \leq F(x_0) - \sum_{i=0}^{k-1} \frac{1}{2\omega} \|\nabla F(x_i)\|_2^2 \tag{4}$$

We know that F is bounded from below. Thus the limit

$$F^\star = \lim_{k \to \infty} F(x_k) > -\infty$$

exists and we have

$$F^\star \leq F(x_0) - \frac{1}{2\omega}\sum_{i=0}^{\infty} \|\nabla F(x_k)\|_2^2 \qquad \Longrightarrow \qquad \sum_{i=0}^{\infty} \|\nabla F(x_k)\|_2^2 \leq 2\omega(F(x_0) - F(x^\star)) < \infty$$

The infinite sum on the left of this inequality must converge to 0 (otherwise the inequality is violated). Thus, we have

$$\lim_{k\to\infty} \nabla F(x_k) \; = \; 0 \; .$$

Consequently, the iterates of the exact line-search based Newton type optimization converge to stationary point. This result is independent of x_0 !!!

Notice that the above prove of global convergence to stationary points can be generalized for the case that the line search is not implemented exactly, but such that the following Armijo line search condition holds:

$$F(x_k + \alpha_k \Delta x_k) \leq F(x_k) - c\alpha_k F'(x_k)\Delta x_k$$

for a small $c > 0$. The above can be generalized by starting with this inequality, but the corresponding argument is analogously.

2 Gauss Newton Methods

The goal of this section to develop a numerical algorithm for finding minimizers of the least-squares optimization problem

$$\min_x \; \|f(x)\|_2^2$$

with $f : \mathbb{R}^n \to \mathbb{R}^m$ being a twice Lipschitz-continuously differentiable function.

2.1 Linear Least-Squares Minimization

For the special case that f is a linear function of the form

$$f(x) = Ax - b$$

we can solve the least-squares optimization explicitly. Namely,

$$\min_x \|Ax - b\|_2^2 \; = \; \min_x \; \{x^\intercal A^\intercal Ax - 2b^\intercal Ax + b^\intercal b\}$$

is equivalent to solving the linear stationarity equation

$$0 \; = \; \nabla \{x^\intercal A^\intercal Ax - 2b^\intercal Ax + b^\intercal b\} \; = \; 2A^\intercal Ax - 2A^\intercal b$$

If A has full row rank, then $A^\intercal A$ is symmetric positive definite and invertible and the solution is given by

$$x^\star = -(A^\intercal A)^{-1}A^\intercal b \; .$$

The matrix

$$A^\dagger \; = \; (A^\intercal A)^{-1}A^\intercal$$

is called the pseudo inverse of the matrix A. It has the following properties

1. If A is an invertible square-matrix, $n = m$, then $A^\dagger = A^{-1}$.
2. We have $A^\dagger A = (A^\intercal A)^{-1} A^\intercal A = I$. Thus, $A^\dagger$ is a left inverse of A.
3. But we need to be careful, since $A^\dagger$ is in general not a right inverse of A, $AA^\dagger \neq I$. Instead, it is sometimes useful to note that

$$A^\dagger A A^\dagger = A^\dagger \ .$$

2.2 Gauss Newton Methods

The main idea of Gauss Newton methods is to solve a sequence of linear least-squares optimization in order to generate step directions. This means that we solve the minimization problem

$$\min_{\Delta x_k} \frac{1}{2} \| f(x_k) + f'(x_k) \Delta x_k \|_2^2$$

in order to find our step direction Δx_k. By using the above definition of the pseodo-inverse matrix we find that

$$\Delta x_k = -f'(x_k)^\dagger f(x_k)$$

Notice that this looks very similar to Newton's method, but we repace the inverse of the Jacobian by the left inverse (or pseudo-inverse).

2.3 Interpretation as Newton-type method

In order to compare Newton methods with Gauss-Newton methods, we work out the Newton type iteration for the minimization problem

$$\min_x F(x) \quad \text{with} \quad F(x) = \frac{1}{2} \| f(x) \|_2^2 ,$$

which yields

$$\Delta x_k = -M(x_k)^{-1} \nabla F(x_k) = -M(x_k)^{-1} \left[f'(x_k)^\intercal f(x_k) \right]$$

This means that if we set

$$M(x_k) = f'(x_k)^\intercal f'(x_k)$$

we obtain the Gauss-Newton method. This means that the term $f'(x_k)^\intercal f'(x_k)$ can be interpreted as a Hessian approximation. In order to analyze this, we work out the exact Hessian of F, which is given by

$$\nabla^2 F(x_k) \ = \ \nabla^\intercal \left[f'(x_k)^\intercal f(x_k) \right] \ = \ f''(x_k) f(x_k) + f'(x_k)^\intercal f'(x_k) \ .$$

Thus, the Hessian approximation error is bounded by

$$\left\| \nabla^2 F(x_k) - M(x_k) \right\| = \left\| \nabla^2 f''(x_k) f(x_k) \right\| \leq \| f''(x_k) \| \| f(x_k) \| \ .$$