Lecture 24

24.1 Main Inequalities for Cubic Newton Step

Let us consider one step $x \mapsto x^+$ of the cubic Newton method as applied to $f : \mathbb{R}^n \to \mathbb{R}$, defined by

$$
\begin{aligned}
x^+ \;&=\; \arg\min_{y \in \mathbb{R}^n} \Omega_H(x; y) \\
&=\; \arg\min_{y \in \mathbb{R}^n} \Big\{ \langle \nabla f(x), y - x \rangle + \tfrac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle + \tfrac{H}{6} \|y - x\|^3 \Big\},
\end{aligned}
$$

and satisfying the first-order optimality condition:

$$
\nabla f(x) + \nabla^2 f(x)(x^+ - x) + \tfrac{Hr}{2} B(x^+ - x) \;=\; 0, \tag{24.1}
$$

where $r := \|x^+ - x\| = \langle B(x^+ - x), x^+ - x \rangle^{1/2}$ and $B = B^\top \succ 0$ is the matrix that defines the generalized Euclidean norm. In the non-degenerate case, the optimality condition (24.1) is equivalent to the step written in the following classic form:

$$
x^+ \;=\; x - \Big(\nabla^2 f(x) + \tfrac{Hr}{2} B \Big)^{-1} \nabla f(x).
$$

We have the following main inequalities that involve r:

1. By Lemma 23.2.1 from the previous lecture, for any $H \geq 0$, it holds:

$$
\|\nabla f(x^+)\|_* \;\leq\; \tfrac{L+H}{2} r^2. \tag{24.2}
$$

2. By properties of the solution to the cubic subproblem, we know that:

$$
\nabla^2 f(x) + \tfrac{Hr}{2} B \;\succeq\; 0. \tag{24.3}
$$

   Using the Lipschitz continuity of the Hessian, we conclude that

$$
\tfrac{Hr}{2} B \;\overset{(24.3)}{\succeq}\; -\nabla^2 f(x) \;\succeq\; -LrB.
$$

   Rearranging the terms, we get, for any $H \geq 0$:

$$
r \;\geq\; \tfrac{2}{H + 2L} \mu(x^+), \tag{24.4}
$$

   where $\mu(x^+) := -\lambda_{\min}(B^{-1/2} \nabla^2 f(x^+) B^{-1/2})$.

Now, let us choose $H \geq L$, and substitute the optimality condition (24.1) into the global upper bound on the objective function (see Lemma 23.1.1 in the previous lecture), which was the main motivation for defining the cubic step:

$$\begin{aligned} f(x^+) \quad &\leq \quad f(x) + \langle \nabla f(x), x^+ - x \rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(x^+ - x), x^+ - x \rangle + \tfrac{H}{6} r^3 \\ &\overset{(24.1)}{=} \quad f(x) - \tfrac{1}{2}\langle \nabla^2 f(x)(x^+ - x), x^+ - x \rangle - \tfrac{H}{2} r^3 + \tfrac{H}{6} r^3 \\ &= \quad f(x) - \tfrac{1}{2}\langle \Big[\nabla^2 f(x) + \tfrac{Hr}{2} B \Big] (x^+ - x), x^+ - x \rangle - \tfrac{H}{12} r^3 \\ &\overset{(24.3)}{\leq} \quad f(x) - \tfrac{H}{12} r^3. \end{aligned}$$

Therefore, we have established the progress in the function value:

Lemma 24.1.1. *For any $H \geq L$, it holds:*

$$f(x) - f(x^+) \quad \geq \quad \tfrac{H}{12} r^3. \tag{24.5}$$

Now, using that $H \geq L$, we have:

$$r \quad \overset{(24.2)}{\geq} \quad \Big(\tfrac{2}{H+L} \| \nabla f(x^+) \|_* \Big)^{1/2} \quad \geq \quad \Big(\tfrac{1}{H} \| \nabla f(x^+) \|_* \Big)^{1/2}$$

and

$$r \quad \overset{(24.4)}{\geq} \quad -\tfrac{2}{H+2L} \mu(x^+) \quad \geq \quad -\tfrac{2}{3H} \mu(x^+).$$

Combining the progress (24.5) with these lower bounds on r, we obtain:

Theorem 24.1.2. *Let $H \geq L$. Then,*

$$f(x) - f(x^+) \quad \geq \quad \max\Big\{ \tfrac{1}{12 H^{1/2}} \| \nabla f(x^+) \|_*^{3/2}, \; \tfrac{2}{3^4 H^2} \mu(x^+)^3 \Big\}. \tag{24.6}$$

24.2 Convergence to Second-Order Stationary Point

Consider iterations of the cubic Newton method, starting from some $x_0 \in \mathbb{R}^n$:

$$x_{k+1} \quad = \quad \operatorname*{arg\,min}_{y \in \mathbb{R}^n} \Omega_L(x_k; y), \qquad k \geq 0. \tag{24.7}$$

where we fix the regularization parameter as $H := L$, for simplicity.

Then, for each iteration we have:

$$f(x_k) - f(x_{k+1}) \quad \overset{(24.6)}{\geq} \quad p_{k+1} \; := \; \max\Big\{ \tfrac{1}{12 L^{1/2}} \| \nabla f(x_{k+1}) \|_*^{3/2}, \; \tfrac{2}{3^4 L^2} \mu(x_{k+1})^3 \Big\}.$$

Telescoping it for the first $k \geq 1$ iterations:

$$f(x_0) - f^\star \quad \geq \quad f(x_0) - f(x_k) \quad \geq \quad k \cdot \tfrac{1}{k} \sum_{i=1}^k p_i \quad \geq \quad k \cdot \min_{1 \leq i \leq k} p_i,$$

we ensure the decrease: $\min_{1 \leq i \leq k} p_i = \mathcal{O}(1/k)$.

Therefore, we obtain the following global convergence rates:

Theorem 24.2.1. *For the iterations of the cubic Newton method* (24.7)*, we have:*

$$\min_{1 \leq i \leq k} \| \nabla f(x_i) \|_* \quad \leq \quad \Big(\tfrac{12 L^{1/2} (f(x_0) - f^\star)}{k} \Big)^{2/3}$$

and

$$\min_{1 \leq i \leq k} \mu(x_i) \quad \leq \quad \Big(\tfrac{3^4 L^2}{2} \cdot \tfrac{f(x_0) - f^\star}{k} \Big)^{1/3}.$$