Continuous Optimization: Algorithms and Complexity
ORIE 6365

Lecture 24: Global Rates for Cubic Newton Method

Nikita Doikov

Cornell University
School of Operations Research and Information Engineering (ORIE)

April 27, 2026

Outline

- Review: Computing Cubic Newton's Step
- Global Rates for Non-Convex Problems
- Convex Optimization

Problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function (possibly non-convex)
- Bounded from below: $f^\star := \inf_{x \in \mathbb{R}^n} f(x) > -\infty$
- The Hessian is Lipschitz continuous:

$$\|\nabla^2 f(y) - \nabla^2 f(x)\| \quad \leq \quad L\|y - x\|, \qquad \forall x, y \in \mathbb{R}^n$$

Lemma. For any $x, y \in \mathbb{R}^n$ it holds:

1. $\|\nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y - x)\|_* \leq \frac{L}{2}\|y - x\|^2$
2. $|f(y) - Q(x; y)| \leq \frac{L}{6}\|y - x\|^3$

where

$$Q(x; y) \quad = \quad f(x) + \langle \nabla f(x), y - x\rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y - x), y - x\rangle$$

is the second-order Taylor polynomial.

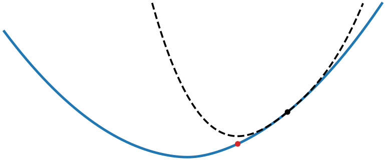
Cubic Newton Step

At a current point $x \in \mathbb{R}^n$, fix a regularization coefficient $H \geq 0$

- Denote the cubically regularized second-order model:

$$\begin{aligned} \Omega_H(x;y) \quad &:= \quad Q(x;y) + \tfrac{H}{6}\|y-x\|^3 \\ &= \quad f(x) + \langle \nabla f(x), y-x \rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y-x), y-x \rangle + \tfrac{H}{6}\|y-x\|^3 \end{aligned}$$

- Cubic Newton step: $\boxed{x^+ := \arg\min_{y \in \mathbb{R}^n} \Omega_H(x;y)}$



H = 0.1

- **From Lemma:** for $H \geq L$ we have $f(x^+) \leq \Omega_H(x;x^+) \Rightarrow$ progress of each step

Solving Regularized Non-Convex Subproblem

Subproblem:

$$\min_{y \in \mathbb{R}^n} \Big[\, P(y) \;\; := \;\; \langle \nabla f(x), y - x \rangle + \tfrac{1}{2} \langle \nabla^2 f(x)(y-x), y - x \rangle + \varphi(\|y - x\|^2) \, \Big]$$

where

- $\|y - x\|^2 := \langle B(y-x), y-x \rangle$ is a quadratic form with $B = B^\top \succ 0$ (e.g., $B = I$)
- $\varphi(s) := \frac{H}{6} s^{3/2}$, for $s > 0$
- Note that for non-convex f, we might have $\nabla^2 f(x) \not\succeq 0 \Rightarrow$ non-convex subproblem

Another popular approach (trust-region methods), for a parameter $r > 0$:

$$\varphi(s) \;\; := \;\; \begin{cases} 0, & s \leq r \\ +\infty, & \text{otherwise} \end{cases}$$

- Subproblem: $\min\limits_{y \in \mathbb{R}^n \, : \, \|y-x\| \leq r} \Big\{ \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x)(y-x), y-x \rangle \Big\}$
- Well-developed literature (see, e.g., Chapter 7 in [Conn, Gould, Toint; 2000]).

Strong Duality

Theorem. It holds:

$$\min_{y \in \mathbb{R}^n} P(y) \quad = \quad \sup_{\tau \in \mathcal{W}} D(\tau)$$

where $D(\tau)$ is a concave univariate function:

$$\begin{aligned} D(\tau) \quad &:= \quad -\tfrac{1}{2}\langle \nabla f(x), (\nabla^2 f(x) + \tau B)^{-1} \nabla f(x) \rangle - \varphi^\star\Big(\tfrac{\tau}{2}\Big) \\ &= \quad -\tfrac{1}{2}\langle \nabla f(x), (\nabla^2 f(x) + \tau B)^{-1} \nabla f(x) \rangle - \tfrac{2}{3H^2}\tau^3 \end{aligned}$$

defined on

$$\begin{aligned} \mathcal{W} \quad &:= \quad \Big\{ \tau \geq 0 \;\; : \;\; \nabla^2 f(x) + \tau B \succ 0 \Big\} \\ &= \quad \Big\{ \tau \geq 0 \;\; : \;\; \tau > -\lambda_{\min}\Big(B^{-1/2} \nabla^2 f(x) B^{-1/2}\Big) \Big\}. \end{aligned}$$

(see, e.g., Theorem 10 in [Nesterov-Polyak, 2006] for the proof)

Practical Computation

Consider, for simplicity, $B = I$ (the standard Euclidean norm $\|\cdot\|_2$)

1. Compute eigenvalue (or tridiagonal) decomposition of $\nabla^2 f(x)$:

$$\nabla^2 f(x) \quad = \quad U \Lambda U^\top \qquad O(n^3) \text{ arith. operations}$$

2. Solve the univariate dual problem by computing the root of

$$\varphi(r^\star) \quad = \quad \left\| \left(\nabla^2 f(x) + \tfrac{H}{2} r^\star I \right)^{-1} \nabla f(x) \right\|_2 - r^\star \quad = \quad 0$$

   (e.g., using the binary search of univariate Newton's method: $\tilde{O}(n^2)$ arith. operations)

3. Compute the Cubic Newton step as

$$x^+ \quad = \quad x - \left(\nabla^2 f(x) + \tfrac{H}{2} r^\star I \right)^{-1} \nabla f(x)$$

Outline

- Review: Computing Cubic Newton's Step
- Global Rates for Non-Convex Problems
- Convex Optimization

Main Lemmas

Previous time, we have proved the following:

Lemma. For any $H \geq 0$:

$$\|\nabla f(x^+)\|_* \;\leq\; \frac{L+H}{2} r^2, \qquad \text{where} \qquad r \;:=\; \|x^+ - x\|.$$

Another important result (follows from solving the dual problem): for any $H > 0$

$$\boxed{\nabla^2 f(x) + \frac{Hr}{2} B \;\succeq\; 0.}$$

Therefore,

$$\frac{Hr}{2} B \;\succeq\; -\nabla^2 f(x) \;\succeq\; -\nabla^2 f(x^+) - LrB.$$

Lemma. For any $H > 0$:

$$r \;\geq\; -\frac{2}{H+2L} \lambda_{\min}\Big(B^{-1/2} \nabla^2 f(x^+) B^{-1/2} \Big).$$

Functional Progress

Let $H \geq L$. Then, we had the global upper bound:

$$f(x^+) \quad \leq \quad f(x) + \langle \nabla f(x), x^+ - x \rangle + \tfrac{1}{2} \langle \nabla^2 f(x)(x^+ - x), x^+ - x \rangle + \tfrac{H}{6} r^3.$$

- Optimality condition for x^+:

$$\nabla f(x) + \nabla^2 f(x)(x^+ - x) + \tfrac{Hr}{2} B(x^+ - x) \quad = \quad 0.$$

- Multiplyting it by $x^+ - x$, we get:

$$\langle \nabla f(x), x^+ - x \rangle \quad = \quad -\langle \nabla^2 f(x)(x^+ - x), x^+ - x \rangle - \tfrac{H}{2} r^3.$$

Therefore,

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

Cubic Step: Summary

For $H \geq L$, we obtained:

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

At the same time, before we had:

1. $r \geq \left(\frac{2}{H+L}\|\nabla f(x^+)\|_*\right)^{1/2} \geq \left(\frac{1}{H}\|\nabla f(x^+)\|_*\right)^{1/2}$
2. $r \geq \frac{2}{H+2L}\mu(x^+) \geq \frac{2}{3H}\mu(x^+)$, where $\mu(x^+) := -\lambda_{\min}(B^{-1/2}\nabla^2 f(x^+) B^{-1/2})$.

Combining all together, we get the following progress of one Cubic Newton step:

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

$$f(x) - f(x^+) \quad \geq \quad \max\left\{ \frac{1}{12H^{1/2}}\|\nabla f(x^+)\|_*^{3/2}, \ \frac{2}{3^4 H^2}\mu(x^+)^3 \right\}.$$

- NB: for the gradient step, we have $f(x) - f(x^+) \geq \frac{1}{2L_1}\|\nabla f(x)\|_*^2$.

Algorithm: Cubic Newton Method

Initialization: $x_0 \in \mathbb{R}^n$

Iterate, $k \geq 0$:

1. Choose $H_k > 0$
2. Compute the cubic step:

$$
\begin{aligned}
x_{k+1} &= \arg\min_{y \in \mathbb{R}^n} \Big[\, Q(x_k; y) + \tfrac{H_k}{6} \|y - x_k\|^3 \, \Big] \\
&= \arg\min_{y \in \mathbb{R}^n} \Big[\langle \nabla f(x_k), y - x_k \rangle + \tfrac{1}{2} \langle \nabla^2 f(x_k)(y - x_k), y - x_k \rangle + \tfrac{H_k}{6} \|y - x_k\|^3 \Big]
\end{aligned}
$$

How to choose H_k?

- Constant choice: $H_k \equiv L$, where L is the Lipschitz constant of the Hessian
- Adaptive search!

Global Convergence of Cubic Newton

Let $H_k \equiv L$. Then, for each iteration we have

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

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

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

Theorem.

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

Also,

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

- Convergence to an approximate second-order stationary point $\approx$ local minimum

Comparison with Gradient Method

Complexity to obtain $\|\nabla f(\bar{x})\|_* \leq \varepsilon$?

$$\min_{1\leq i\leq k} \|\nabla f(x_i)\|_* \;\leq\; \left(\frac{12L^{1/2}(f(x_0)-f^\star)}{k}\right)^{2/3} \;\overset{(?)}{\leq}\; \varepsilon$$

- Cubic Newton's method:

$$k \;=\; O\left(\frac{\sqrt{L}(f(x_0)-f^\star)}{\varepsilon^{3/2}}\right)$$

  second-order oracle calls (computing $\nabla f(x)$ and $\nabla^2 f(x)$) and cubic subroblems

- Gradient method:

$$k \;=\; O\left(\frac{L_1(f(x_0)-f^\star)}{\varepsilon^2}\right)$$

  first-order oracle calls

Adaptive Cubic Newton

Initialization: $x_0 \in \mathbb{R}^n$, $\varepsilon > 0$, $H_0 > 0$

Iterate, $k \geq 0$:

1. If $\|\nabla f(x_k)\|_* \leq \varepsilon$ then **return** x_k
2. **For** $t \geq 0$ **iterate**:
   - Set $H_k^+ := H_k \cdot 2^t$
   - Compute the cubic Newton step:

$$x_k^+ \quad := \quad x_k^+(H_k^+) \quad = \quad \operatorname*{arg\,min}_{y \in \mathbb{R}^n} \Big\{ Q(x_k; y) + \tfrac{H_k^+}{6}\|y - x_k\|^3 \Big\}$$

   - If $f(x_k) - f(x_k^+) \geq \frac{1}{12 H_k^+}\|\nabla f(x_k^+)\|_*^{3/2}$ then **break**
3. Set $x_{k+1} = x_k^+$ and $H_{k+1} = H_k^+$

Outline

- Review: Computing Cubic Newton's Step
- Global Rates for Non-Convex Problems
- Convex Optimization

Convex Problem

$$\boxed{\min_{x \in \mathbb{R}^n} f(x)}$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex

- Cubic subproblem — easier to solve:

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

  has unique stationarity point (global minimum). Cubic step:

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

  where $r = \|x^+ - x\|$ and $\nabla^2 f(x) + \frac{H}{2} r B \succ 0$ (unless $\nabla f(x) = 0$)

- Faster rates to the global optimum!

Global Convergence for Convex Functions

Progress of one step. Denote $F_k := f(x_k) - f^\star$.

- $F_k - F_{k+1} \geq \frac{1}{12L^{1/2}} \|\nabla f(x_{k+1})\|_*^{3/2}$
- Method is monotone. We have

$$x_{k+1} \in \mathcal{F}_0 := \Big\{ x : f(x) \leq f(x_0) \Big\}. \qquad \text{Denote} \quad D := \mathrm{diam}(\mathcal{F}_0).$$

- By convexity,

$$F_{k+1} = f(x_{k+1}) - f^\star \leq \langle \nabla f(x_{k+1}), x_{k+1} - x^\star \rangle \leq \|\nabla f(x_{k+1})\|_* \cdot D.$$

Therefore,

$$F_k - F_{k+1} \geq \frac{1}{12L^{1/2}D^{3/2}} F_{k+1}^{3/2} \equiv c \cdot F_{k+1}^{3/2}$$

Convergence Rate

$$\boxed{F_k - F_{k+1} \quad \geq \quad c \cdot F_{k+1}^{3/2}}$$

Continuous time:

$$-\dot{F}_t \quad \geq \quad c \cdot F_t^{3/2}$$

or

$$\frac{d}{dt}\left[\frac{2}{F_t^{1/2}}\right] \quad = \quad -\frac{\dot{F}_t}{F_t^{3/2}} \quad \geq \quad c$$

Integrating, wet get

$$\frac{2}{F_t^{1/2}} - \frac{2}{F_0^{1/2}} \quad = \quad \int\limits_0^t \frac{d}{d\tau}\left[\frac{2}{F_\tau^{1/2}}\right] \quad \geq \quad ct \qquad \Rightarrow \qquad F_t = O(1/t^2).$$

Theorem. On convex functions, the rate of the Cubic Newton is

$$f(x_k) - f^\star \quad \leq \quad O\left(\frac{LD^3}{k^2}\right), \qquad k \geq 1.$$

Comparison with Gradient Methods

- Cubic Newton:

$$f(x_k) - f^\star \;\le\; O\Big(\tfrac{LD^3}{k^2}\Big) \;\overset{(?)}{\le}\; \varepsilon \qquad \Rightarrow \qquad k \;=\; O\Big(\sqrt{\tfrac{LD^3}{\varepsilon}}\Big).$$

- Gradient method:

$$f(x_k) - f^\star \;\le\; O\Big(\tfrac{L_1\|x_0 - x^\star\|^2}{k}\Big) \;\overset{(?)}{\le}\; \varepsilon \qquad \Rightarrow \qquad k \;=\; O\Big(\tfrac{L_1\|x_0 - x^\star\|^2}{\varepsilon}\Big).$$

- These rates can also be accelerated.