The University of Texas at Austin
Department of Electrical and Computer Engineering

ECE 381V: Large-Scale Optimization II — Spring 2022

LECTURE 26

Caramanis & Mokhtari Wednesday, April 27, 2022

Goal: In this lecture, we study the convergence properties of the Accelerated version of the Cubic Regularization of Newton's method for constrained convex problems.

1 Setup

Suppose we aim to solve the following unconstrained problem

$$\min_{\mathbf{x}\in\mathcal{Q}} f(\mathbf{x})$$

where $\mathcal{Q}$ is a closed convex set and f is a convex function with Lipschitz continuous Hessian.

Assumption 1. *f is twice differentiable and its Hessian is L_2-Lipschitz continuous on $\mathcal{Q}$, i.e.,*

$$\|\nabla^2 f(\mathbf{x}) - \nabla^2 f(\hat{\mathbf{x}})\|_2 \leq L_2\|\mathbf{x}-\hat{\mathbf{x}}\|, \qquad \text{for all} \quad \mathbf{x},\hat{\mathbf{x}}\in\mathcal{Q}$$

2 The Cubic Regularized Newton main concepts

We define

$$T_M(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{y}\in\mathcal{Q}}\{\phi_M(\mathbf{y};\mathbf{x})\}$$

where

$$\phi_M(\mathbf{y};\mathbf{x}) = \nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^\top\nabla^2 f(\mathbf{x})(\mathbf{y}-\mathbf{x}) + \frac{M}{6}\|\mathbf{y}-\mathbf{x}\|^3$$

and we further define

$$f_M(\mathbf{x}) := \min_{\mathbf{y}\in\mathcal{Q}} \{\phi_M(\mathbf{y};\mathbf{x})\}$$

Lemma 1. *For the iterates of CRN when Assumption 1 holds we have*

$$f_M(\mathbf{x}) \leq \min_{\mathbf{y}\in\mathcal{Q}}\left\{f(\mathbf{y}) + \frac{L_2+M}{6}\|\mathbf{y}-\mathbf{x}\|^3\right\}$$

Proof. Note that for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{Q}$ we have

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^\top\nabla^2 f(\mathbf{x})(\mathbf{y}-\mathbf{x}) \leq f(\mathbf{y}) + \frac{L_2}{6}\|\mathbf{y}-\mathbf{x}\|^3$$

By adding $\frac{M}{6}\|\mathbf{y}-\mathbf{x}\|^3$ to both sides and computing the minimum with respect to $\mathbf{y}$ the claim follows. □

Lemma 2. *If $M \geq 2L_2$ then we have*

$$\nabla f(T_M(\mathbf{x}))^\top(\mathbf{x}-T_M(\mathbf{x})) \geq \sqrt{\frac{2}{L_2+M}}\|\nabla f(T_M(\mathbf{x}))\|^{3/2}$$

Proof. Check Lemma 4.2.5 in Nesterov's book. □

3 Main Idea for Acceleration

We aim to estimate the function using a sequence of cubic functions defined as

$$\psi_k(\mathbf{x}) = \ell_k(\mathbf{x}) + \frac{C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3$$

where $\ell_k(\mathbf{x})$ is a linear function of $\mathbf{x}$ and C is a positive parameter.

Now consider the sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$

$$A_{k+1} := A_k + a_k$$

We aim to show that our estimating sequences satisfy the following conditions for any $k \geq 1$

- $\mathcal{H}_k^1: \quad A_k f(\mathbf{x}_k) \leq \psi_k^* := \min_{\mathbf{x}\in\mathcal{Q}} \psi_k(\mathbf{x})$
- $\mathcal{H}_k^2: \quad \psi_k(\mathbf{x}) \leq A_k f(\mathbf{x}) + \frac{2L_2+C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3, \qquad$ for all $\mathbf{x} \in \mathcal{Q}$

Initial Step: First we show that for $\mathbf{x}_1 = T_{L_2}(\mathbf{x}_0)$, $\ell_1(\mathbf{x}) = f(\mathbf{x}_1)$, and $A_1 = 1$ the above conditions hold for $k = 1$.
In this case we have $\psi_1(\mathbf{x}) = \ell_1(\mathbf{x}) + \frac{C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3 = f(\mathbf{x}_1) + \frac{C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3$. Hence, $\psi_1^* = f(\mathbf{x}_1)$. Therefore, $\mathcal{H}_1^1$ holds, since $A_1 = 1$.

We further know that for any $\mathbf{x} \in \mathcal{Q}$ we have

$$\begin{aligned}
\psi_1(\mathbf{x}) &= f(\mathbf{x}_1) + \frac{C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3 \\
&\leq \min_{\mathbf{y}\in\mathcal{Q}} \left[f(\mathbf{y}) + \frac{2L_2}{6}\|\mathbf{y} - \mathbf{x}_0\|^3 \right] + \frac{C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3 \\
&\leq f(\mathbf{x}) + \frac{2L_2}{6}\|\mathbf{x} - \mathbf{x}_0\|^3 + \frac{C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3
\end{aligned}$$

where the first inequality follows from Lemma 1 for $\mathbf{x} = \mathbf{x}_0$ and $M = L_2$. Therefore, $\mathcal{H}_1^2$ holds.

General Step: Consider the case that $M \geq 2L_2$. Let us first define the following scalars and vectors:

$$\alpha_k = \frac{a_k}{a_k + A_k}, \qquad \mathbf{v}_k = \operatorname*{argmin}_{\mathbf{y}\in\mathcal{Q}} \psi_k(\mathbf{y})$$

$$\mathbf{y}_k = (1 - \alpha_k)\mathbf{x}_k + \alpha_k \mathbf{v}_k, \qquad \mathbf{x}_{k+1} = T_M(\mathbf{x}_k)$$

$$\psi_{k+1}(\mathbf{x}) = \psi_k(\mathbf{x}) + a_k[f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x} - \mathbf{x}_{k+1})]$$

Now considering these sequences one can show that for any $\mathbf{x} \in \mathcal{Q}$ we have

$$\begin{aligned}
\psi_{k+1}(\mathbf{x}) &= \psi_k(\mathbf{x}) + a_k[f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x} - \mathbf{x}_{k+1})] \\
&\leq A_k f(\mathbf{x}) + \frac{2L_2 + C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3 + a_k[f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x} - \mathbf{x}_{k+1})] \\
&\leq A_k f(\mathbf{x}) + \frac{2L_2 + C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3 + a_k f(\mathbf{x}) \\
&= A_{k+1} f(\mathbf{x}) + \frac{2L_2 + C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3
\end{aligned}$$

where the last inequality follows from convexity of f. Hence, if $\mathcal{H}_k^2$ holds, then $\mathcal{H}_{k+1}^2$ also holds.

Now note that based on the definition of $\mathbf{v}_k$ we have

$$\psi_k(\mathbf{x}) = \ell_k(\mathbf{x}) + \frac{C}{6}\|\mathbf{x}-\mathbf{x}_0\|^3, \qquad \mathbf{v}_k = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{Q}} \psi_k(\mathbf{x})$$

Hence, we have

$$\psi_k(\mathbf{x}) \geq \psi_k^* + \frac{C}{12}\|\mathbf{x}-\mathbf{v}_k\|^3$$

which implies that

$$\psi_k(\mathbf{x}) \geq A_k f(\mathbf{x}_k) + \frac{C}{12}\|\mathbf{x}-\mathbf{v}_k\|^3$$

Therefore,

$$\begin{aligned}
\psi_{k+1}^* &= \min_{\mathbf{x}\in\mathcal{Q}} \left\{\psi_k(\mathbf{x}) + a_k[f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}-\mathbf{x}_{k+1})]\right\} \\
&= \min_{\mathbf{x}\in\mathcal{Q}} \left\{A_k f(\mathbf{x}_k) + \frac{C}{12}\|\mathbf{x}-\mathbf{v}_k\|^3 + a_k[f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}-\mathbf{x}_{k+1})]\right\} \\
&\geq \min_{\mathbf{x}\in\mathcal{Q}} \left\{A_k f(\mathbf{x}_{k+1}) + A_k\nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}_k-\mathbf{x}_{k+1}) + \frac{C}{12}\|\mathbf{x}-\mathbf{v}_k\|^3 + a_k[f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}-\mathbf{x}_{k+1})]\right\} \\
&= \min_{\mathbf{x}\in\mathcal{Q}} \left\{(A_k+a_k) f(\mathbf{x}_{k+1}) + A_k\nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}_k-\mathbf{x}_{k+1}) + a_k\nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}-\mathbf{x}_{k+1}) + \frac{C}{12}\|\mathbf{x}-\mathbf{v}_k\|^3\right\} \\
&= \min_{\mathbf{x}\in\mathcal{Q}} \left\{A_{k+1} f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top(A_{k+1}\mathbf{y}_k - a_k\mathbf{v}_k - A_k\mathbf{x}_{k+1}) + a_k\nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}-\mathbf{x}_{k+1}) + \frac{C}{12}\|\mathbf{x}-\mathbf{v}_k\|^3\right\} \\
&= \min_{\mathbf{x}\in\mathcal{Q}} \left\{A_{k+1} f(\mathbf{x}_{k+1}) + A_{k+1}\nabla f(\mathbf{x}_{k+1})^\top(\mathbf{y}_k-\mathbf{x}_{k+1}) + a_k\nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}-\mathbf{v}_k) + \frac{C}{12}\|\mathbf{x}-\mathbf{v}_k\|^3\right\}
\end{aligned}$$

Now using Lemma 2 and considering the fact that $M \geq 2L_2$ we have

$$\nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}-\mathbf{x}_{k+1}) \geq \sqrt{\frac{2}{L_2+M}}\|\nabla f(\mathbf{x}_{k+1})\|^{3/2}$$

Therefore, we have

$$\psi_{k+1}^* \geq A_{k+1}f(\mathbf{x}_{k+1}) + \min_{\mathbf{x}\in\mathcal{Q}} \left\{A_{k+1}\sqrt{\frac{2}{L_2+M}}\|\nabla f(\mathbf{x}_{k+1})\|^{3/2} + a_k\nabla f(\mathbf{x}_{k+1})^\top(\mathbf{x}-\mathbf{v}_k) + \frac{C}{12}\|\mathbf{x}-\mathbf{v}_k\|^3\right\}$$

The above expression shows that $\mathcal{H}_{k+1}^1$ holds if we show that the solution of the minimization is non-negative.
It can be verified that this condition holds if we select our parameters such that

$$A_{k+1}\sqrt{\frac{2}{L_2+M}} \geq \frac{4}{3\sqrt{C}}\ a_k^{3/2}$$

Now if we select the parameters as

$$A_k = \frac{k(k+1)(k+2)}{6}, \qquad a_k = \frac{(k+1)(k+2)}{2}$$

It can be verified that

$$a_k^{-3/2}A_{k+1} \geq \frac{2}{3}$$

Hence, the above inequality holds if

$$\frac{1}{L_2 + M} \geq \frac{2}{C}$$

Hence, we can choose

$$M = 2L_2, \qquad C = 2(L_2 + M) = 6L_2$$

4 Accelerated CRN

At step 0:

- Select $\mathbf{x}_0 \in \mathcal{Q}$. Set $M = 2L_2$ and $C = 6L_2$
- Compute $\mathbf{x}_1 = T_{L_2}(\mathbf{x}_0)$ and define $\psi_1(\mathbf{x}) = f(\mathbf{x}_1) + \frac{C}{6}\|\mathbf{x} - \mathbf{x}_0\|^3$

At step $k \geq 1$:

- Compute $\mathbf{v}_k = \operatorname{argmin}_{\mathbf{x} \in \mathcal{Q}} \psi_k(\mathbf{x})$
- Set $\mathbf{y}_k = \frac{k}{k+3}\mathbf{x}_k + \frac{3}{k+3}\mathbf{v}_k$
- Compute $\mathbf{x}_{k+1} = T_M(\mathbf{y}_k)$
- Update $\psi_{k+1}(\mathbf{x}) = \psi_k(\mathbf{x}) + \frac{(k+1)(k+2)}{2}[f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top (\mathbf{x} - \mathbf{x}_{k+1})]$

5 Main convergence result

Now we are ready to state our main theorem.

Theorem 1. *Suppose the objective function is convex and its Hessian is L_2-Lipschitz. Then, we have*

$$f(\mathbf{x}_k) - f^* \leq \frac{8L_2\|\mathbf{x}_0 - \mathbf{x}^*\|^3}{k(k+1)(k+2)}$$

Proof. We know that

$$A_k f(\mathbf{x}_k) \leq \psi_k^* \leq A_k f(\mathbf{x}_k) + \frac{2L_2 + C}{6}\|\mathbf{x}_0 - \mathbf{x}^*\|^3$$

Hence,

$$f(\mathbf{x}_k) - f^* \leq \frac{2L_2 + C}{6A_k}\|\mathbf{x}_0 - \mathbf{x}^*\|^3 = \frac{8L_2}{k(k+1)(k+2)}\|\mathbf{x}_0 - \mathbf{x}^*\|^3.$$

□