MAT3220: Operation Research
Homework 5

李肖鹏 (116010114)

Due date: April 9, 12 p.m., 2019

Problem 1. Suppose that the gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuity, that is, $\exists\, L > 0$ such that $\|\nabla f(\vec{x}) - \nabla f(\vec{y})\| \le L\|\vec{x} - \vec{y}\|$, $\forall\, \vec{x}, \vec{y} \in \mathbb{R}^n$. Prove that for any $\vec{u}, \vec{v} \in \mathbb{R}^n$ we have

$$f(\vec{u}) \le f(\vec{v}) + \nabla f(\vec{v})^{\mathrm{T}}(\vec{u} - \vec{v}) + \frac{L}{2}\|\vec{u} - \vec{v}\|^2$$

Consider function $h(t) = f(\vec{v} + t(\vec{u} - \vec{v}))$ and Newton-Leibniz formula, we have

$$h(1) - h(0) = \int_0^1 h'(t)\, dt \Longrightarrow f(\vec{u}) - f(\vec{v}) = \int_0^1 \nabla f(\vec{v} + t(\vec{u} - \vec{v}))^{\mathrm{T}}(\vec{u} - \vec{v})\, dt$$

Therefore, we have

$$\begin{aligned}
|f(\vec{u}) - f(\vec{v}) - \nabla f(\vec{v})^{\mathrm{T}}(\vec{u} - \vec{v})| &= \left| \int_0^1 \left[\nabla f(\vec{v} + t(\vec{u} - \vec{v})) - \nabla f(\vec{v})\right]^{\mathrm{T}} (\vec{u} - \vec{v})\, dt \right| \\
&\le \int_0^1 \|\nabla f(\vec{v} + t(\vec{u} - \vec{v})) - \nabla f(\vec{v})\| \cdot \|\vec{u} - \vec{v}\|\, dt \\
&\le \int_0^1 Lt\|\vec{u} - \vec{v}\|^2\, dt \\
&= \frac{L}{2}\|\vec{u} - \vec{v}\|^2
\end{aligned}$$

In conclusion, we have

$$f(\vec{u}) \le f(\vec{v}) + \nabla f(\vec{v})^{\mathrm{T}}(\vec{u} - \vec{v}) + \frac{L}{2}\|\vec{u} - \vec{v}\|^2$$

Problem 2. The gradient descent method with Armijo's line-search rule is as follows,

> Set parameters $s > 0$, $\beta \in (0, 1)$ and $\sigma \in (0, 1)$. Initially, set $k = 1$.
> **For** iterate k, let ℓ be the smallest integer satisfying
>
> $$f(\vec{x}^k) - f(\vec{x}^k + \beta^\ell s \vec{d}^k) \ge -\sigma \beta^\ell s \nabla f(\vec{x}^k)^{\mathrm{T}} \vec{d}^k$$
>
> where $\vec{d}^k = -\nabla f(\vec{x}^k)$. Let $\alpha_k = s\beta^\ell$, $\vec{x}^{k+1} = \vec{x}^k + \alpha_k \vec{d}^k$. Let $k = k + 1$, and return to **For**.

Prove that if f is convex and twice differentiable with a bounded Hessian matrix, then there exists a constant $c > 0$, such that

$$f(\vec{x}^k) - f(\vec{x}^*) \le \frac{c}{k}, \quad \text{for all } k \ge 1$$

We first make the assumption that the level set of the initial point is bounded, which means $\|\vec{x}^0 - \vec{x}^*\| \le C$. Also, suppose $\nabla^2 f(\vec{x}) \preceq MI$ for $M > 0$. Since we use backtrack line search, the following Armijo's condition must be satisfied,

$$f(\vec{x}^k) - f(\vec{x}^k + s\beta^\ell \vec{d}^k) \ge -\sigma\beta^\ell s \nabla f(\vec{x}^k)^{\mathrm{T}} \vec{d}^k \tag{2.1}$$

$$f(\vec{x}^k) - f(\vec{x}^k + s\beta^{\ell-1} \vec{d}^k) < -\sigma\beta^{\ell-1} s \nabla f(\vec{x}^k)^{\mathrm{T}} \vec{d}^k \tag{2.2}$$

Substitute $\vec{d}^k = -\nabla f(\vec{x}^k)$ and for simplicity, let $\alpha_k = s\beta^\ell$, then by (2.1), we have

$$f(\vec{x}^{k+1}) \le f(\vec{x}^k) - \sigma\alpha_k \|\nabla f(\vec{x}^k)\|^2 \tag{2.3}$$

This shows that the function value is strictly decreasing, so the level set of each iteration will be contained in the level set of the initial point, hence we have $\|\vec{x}^k - \vec{x}^*\| \le C$, for all k. By Taylor expansion of f at $\vec{x}^k$ and (2.2), we obtain

$$\begin{aligned} f(\vec{x}^k + (\alpha_k/\beta)\vec{d}^k) &= f(\vec{x}^k) + (\alpha_k/\beta)\nabla f(\vec{x}^k)^{\mathrm{T}}\vec{d}^k + \frac{1}{2}(\alpha_k/\beta)^2 \vec{d}^{k\mathrm{T}} \nabla^2 f(\vec{x}')\vec{d}^k \\ &> f(\vec{x}^k) + \sigma(\alpha_k/\beta)\nabla f(\vec{x}^k)^{\mathrm{T}}\vec{d}^k \end{aligned}$$

Since the hessian is bounded

$$\frac{1}{2}(\alpha_k/\beta)^2 M\|f(\vec{x})\|^2 > (1-\sigma)\alpha_k/\beta\|f(\vec{x})\|^2 \Longrightarrow \alpha_k > \frac{2(1-\sigma)\beta}{M} \tag{2.4}$$

Also, by the convexity of f, we have

$$f(\vec{x}^k) - f(\vec{x}^*) \le \|\nabla f(\vec{x}^k)\|\|\vec{x}^k - \vec{x}^*\| \tag{2.5}$$

Use (2.5) to eliminate the gradient term in (2.3), we have

$$f(\vec{x}^{k+1}) - f(\vec{x}^k) \le -\sigma\alpha_k \frac{\|f(\vec{x}^k) - f(\vec{x}^*)\|^2}{\|\vec{x}^k - \vec{x}^*\|^2} \le -\frac{\sigma\alpha_k}{C^2}\|f(\vec{x}^k) - f(\vec{x}^*)\|^2$$

Use (2.4), we finally have

$$f(\vec{x}^{k+1}) - f(\vec{x}^k) \le -\frac{2\sigma(1-\sigma)\beta}{MC^2}\|f(\vec{x}^k) - f(\vec{x}^*)\|^2 \tag{2.6}$$

Let $e_k = f(\vec{x}^k) - f(\vec{x}^*)$ and denote $K = 2\sigma(1-\sigma)\beta/(MC^2)$, then we have the recurrence relation as follows (for all k, $e_k > 0$, and this implies $1 - Ke_k > 0$),

$$e_{k+1} \le e_k - Ke_k^2 \Longrightarrow \frac{1}{e_{k+1}} \ge \frac{1}{e_k} + \frac{K}{1 - Ke_k} \ge \frac{1}{e_k} + K \tag{2.7}$$

It is easy to obtain

$$e_{k+1} \le \frac{e_1}{1 + Kke_1} \le \frac{1}{Kk} = \frac{c}{k}, \qquad \text{where } \ c = \frac{1}{K} = \frac{MC^2}{2\beta\sigma(1-\sigma)} \tag{$*$}$$

so $f(\vec{x}^k) - f(\vec{x}^*) = O(1/k)$, which is equivalent to what we need to prove.

Problem 3. Suppose that the objective function f is uniformly convex, i.e., there exist $0 < m \leq M < \infty$,

$$mI \preceq \nabla^2 f(\vec{x}) \preceq MI, \quad \text{for all } \vec{x}$$

Consider Newton's method with Armijo's line search rule as follows,

Set parameters $s > 0$, $\beta \in (0,1)$ and $\sigma \in (0,1)$. Initially, set $k = 1$.
For iterate k, let ℓ be the smallest integer satisfying

$$f(\vec{x}^k) - f(\vec{x}^k + \beta^\ell s \vec{d}^k) \geq -\sigma \beta^\ell s \nabla f(\vec{x}^k)^{\mathrm{T}} \vec{d}^k$$

where $\vec{d}^k = -(\nabla^2 f(\vec{x}^k))^{-1} \nabla f(\vec{x}^k)$.
Compare $f(\vec{x}^k + s\beta^\ell \vec{d}^k)$ with $f(\vec{x}^k + \vec{d}^k)$ and let α_k be either $s\beta^\ell$ or 1, whichever is lesser in terms of the $f(\vec{x}^k + \alpha_k \vec{d}^k)$ value.
Let $\vec{x}^{k+1} = \vec{x}^k + \alpha_k \vec{d}^k$. Let $k = k+1$, and return to **For**.

Prove that the above algorithm has a global linear rate of convergence.

If $\nabla^2 f(\vec{x})$ is not guaranteed to be continuous, then use (2.1) and (2.2), and substitute $\alpha_k' = s\beta^\ell$, $\vec{d}^k = -(\nabla^2 f(\vec{x}^k))^{-1} \nabla f(\vec{x}^k)$, we have

$$f(\vec{x}^k + \alpha_k' \vec{d}^k) \leq f(\vec{x}^k) - \sigma \alpha_k' \nabla f(\vec{x}^k)^{\mathrm{T}} (\nabla^2 f(\vec{x}^k))^{-1} \nabla f(\vec{x}^k) \tag{3.1}$$

$$f(\vec{x}^k + (\alpha_k'/\beta) \vec{d}^k) > f(\vec{x}^k) - \sigma (\alpha_k'/\beta) \nabla f(\vec{x}^k)^{\mathrm{T}} (\nabla^2 f(\vec{x}^k))^{-1} \nabla f(\vec{x}^k) \tag{3.2}$$

Use the same argument by which we obtained (2.4) and since $M^{-1} I \preceq (\nabla^2 f(\vec{x}^k))^{-1} \preceq m^{-1} I$,

$$\frac{\alpha_k'}{2\beta} (\vec{d}^k)^{\mathrm{T}} \nabla^2 f(\vec{x}^k) \vec{d}^k > (\sigma - 1) \nabla f(\vec{x}^k)^{\mathrm{T}} \vec{d}^k \implies \alpha_k' > \frac{2\beta(1-\sigma)m}{M} \tag{3.3}$$

Denote that

$$\alpha_k = \underset{\alpha_k' \in \{1, s\beta^\ell\}}{\arg\min}\ f(\vec{x}^k + \alpha_k' \vec{d}^k), \qquad f(\vec{x}^k + \alpha_k \vec{d}^k) = \min_{\alpha_k' \in \{1, s\beta^\ell\}} f(\vec{x}^k + \alpha_k' \vec{d}^k)$$

With (3.1), the upper bound of $f(\vec{x}^k + \alpha_k \vec{d}^k)$ is given by

$$\begin{aligned} f(\vec{x}^k + \alpha_k \vec{d}^k) &\leq f(\vec{x}^k + \alpha_k' \vec{d}^k) \\ &\leq f(\vec{x}^k) - \sigma \alpha_k' \nabla f(\vec{x}^k)^{\mathrm{T}} (\nabla^2 f(\vec{x}^k))^{-1} \nabla f(\vec{x}^k) \\ &\leq f(\vec{x}^k) - \frac{\sigma \alpha_k'}{M} \|\nabla f(\vec{x}^k)\|^2 \end{aligned}$$

By substituting (3.3), we obtain

$$f(\vec{x}^k + \alpha_k \vec{d}^k) \leq f(\vec{x}^k) - \frac{2\beta\sigma(1-\sigma)m}{M^2} \|\nabla f(\vec{x}^k)\|^2 \tag{3.4}$$

To derive the bound of gradient, consider Taylor expansion

$$f(\vec{y} + \lambda \vec{e}) = f(\vec{y}) + \lambda \nabla f(\vec{y})^{\mathrm{T}} \vec{e} + \frac{\lambda^2}{2} \vec{e}^{\mathrm{T}} \nabla^2 f(\vec{y}') \vec{e}$$

Let $\overrightarrow{y} = \overrightarrow{x}^*$, where $\overrightarrow{x}^*$ is the optimal solution, $\lambda = 1$, and $\overrightarrow{e} = \overrightarrow{x}^k - \overrightarrow{x}^*$, combining the lower and upper bound of the hessian, we have

$$\frac{m}{2}\|\overrightarrow{x}^k - \overrightarrow{x}^*\|^2 \leq f(\overrightarrow{x}^k) - f(\overrightarrow{x}^*) \leq \frac{M}{2}\|\overrightarrow{x}^k - \overrightarrow{x}^*\|^2 \tag{3.5}$$

Then, let $\overrightarrow{y} = \overrightarrow{x}$, $\lambda = 1$, and $\overrightarrow{e} = \overrightarrow{x}^* - \overrightarrow{x}^k$,

$$f(\overrightarrow{x}^*) \geq f(\overrightarrow{x}^k) - \|\nabla f(\overrightarrow{x}^k)\|\|\overrightarrow{x}^k - \overrightarrow{x}^*\| + \frac{m}{2}\|\overrightarrow{x}^k - \overrightarrow{x}^*\|^2$$

Use (3.5) to eliminate $f(\overrightarrow{x}^*) - f(\overrightarrow{x}^k)$, we have

$$m\|\overrightarrow{x}^k - \overrightarrow{x}^*\| \leq \|\nabla f(\overrightarrow{x}^k)\|$$

Consider mean value theorem, we have

$$\|\nabla f(\overrightarrow{x}^k) - \nabla f(\overrightarrow{x}^*)\| = \|\nabla f(\overrightarrow{x}^k) - 0\| = \|\nabla^2 f(\overrightarrow{x}')\|\|\overrightarrow{x}^k - \overrightarrow{x}^*\| \leq M\|\overrightarrow{x}^k - \overrightarrow{x}^*\|$$

Therefore, combining the lower and upper bound of gradient together, we have

$$m\|\overrightarrow{x}^k - \overrightarrow{x}^*\| \leq \|\nabla f(\overrightarrow{x}^k)\| \leq M\|\overrightarrow{x}^k - \overrightarrow{x}^*\| \tag{3.6}$$

Notice that (3.5) and (3.6) does not depend on algorithm but they are only related to the uniform convexity of function f. Thus, we can apply them to (3.4). Use (3.6) to eliminate gradient and use (3.5) to eliminate the iterates difference so that only the function difference remains,

$$f(\overrightarrow{x}^k + \alpha_k \overrightarrow{d}^k) \leq f(\overrightarrow{x}^k) - \frac{4\beta\sigma(1-\sigma)m^3}{M^3}(f(\overrightarrow{x}^k) - f(\overrightarrow{x}^*)) \tag{3.7}$$

Rearrange the term, we finally have

$$f(\overrightarrow{x}^k + \alpha_k \overrightarrow{d}^k) - f(\overrightarrow{x}^*) \leq \left[1 - \frac{4\beta\sigma(1-\sigma)m^3}{M^3}\right](f(\overrightarrow{x}^k) - f(\overrightarrow{x}^*)) \tag{3.8}$$

Notice that $0 < \sigma(1-\sigma) \leq \frac{1}{4}$, $\beta \in (0,1)$ and $m \leq M$, the coefficient

$$C = 1 - \frac{4\beta\sigma(1-\sigma)m^3}{M^3} \in (0,1)$$

so a global linear convergence is ensured.

If the hessian is continuous, then the algorithm will first have linear and then super-linear convergence, so globally linear convergence still holds. If in addtion, the hessian is Lipschitz continuous, then the algorithm will first have linear and then quadratic convergence, so globally linear convergence also holds. The details are shown in Appendix.

Appendix: Additional Proof of Problem 3

In the first phase, where $\|\nabla f(\vec{x}^k)\| \geq \gamma$, from (3.4), we still have

$$f(\vec{x}^k + \alpha_k \vec{d}^k) \leq f(\vec{x}^k) - \frac{2\beta\sigma(1-\sigma)m}{M^2}\|\nabla f(\vec{x}^k)\|^2 \leq f(\vec{x}^k) - \frac{2\beta\sigma(1-\sigma)m}{M^2}\gamma^2 \tag{3.9}$$

This shows that at each iterations the objective value decreases by at least a positive constant. Since this function is uniformly convex, hence its global minimum is finite, this phase can only last for finite steps, so we don't determine its convergence rate in the context of limit, but it is obvious that the rate of convergence is at least linear.

In the second phase, where $\|\nabla f(\vec{x}^k)\| < \gamma$, by (3.6) and bound of hessian, we have

$$\begin{aligned}
\|\nabla f(\vec{x}^{k+1})\| &= \frac{1}{m}\|\nabla f(\vec{x}^{k+1}) - \nabla f(\vec{x}^k) - (-\nabla f(\vec{x}^k))\| \\
&= \left\|\nabla f(\vec{x}^k + \vec{d}^k) - \nabla f(\vec{x}^k) - \nabla^2 f(\vec{x}^k)\vec{d}^k\right\| \\
&= \left\|\int_0^1 \nabla^2 f(\vec{x} + t\vec{d}^k)\vec{d}^k \, dt - \nabla^2 f(\vec{x}^k)\vec{d}^k\right\| \\
&= \left\|\int_0^1 \left[\nabla^2 f(\vec{x} + t\vec{d}^k) - \nabla^2 f(\vec{x}^k)\right]\vec{d}^k \, dt\right\| \\
&\leq \frac{\epsilon}{m}\|\nabla f(\vec{x}^k)\|
\end{aligned} \tag{3.10}$$

Therefore, we have the recurrence relation

$$\|\nabla f(\vec{x}^{k+1})\| \leq \frac{\epsilon}{m}\|\nabla f(\vec{x}^k)\| \tag{3.11}$$

Since $\nabla^2 f(\vec{x})$ is continuous, there exists δ_0, such that for all $\vec{x}$ such that $\|\vec{x}^{k+1} - \vec{x}^k\| < \delta_0$, the inequality (3.11) holds for $\epsilon = m^3/(2M^2)$, and thus (3.11) reduces to

$$\|\nabla f(\vec{x}^{k+1})\| \leq \frac{1}{2}\|\nabla f(\vec{x}^k)\| \tag{3.12}$$

Since the first phase only continue for finite steps, if we take $\gamma = m\delta_0$, there must exists a k_0 such that $\|\nabla f(\vec{x}^{k_0})\| < m\delta_0$. From (3.6), $\|\vec{x}^{k_0} - \vec{x}^*\| < \delta_0$, thus we can apply (3.12), and thus $\|\nabla f(\vec{x}^{k_0+1})\| < \gamma$. By induction, we can see that for all $k \geq k_0$, $\|\nabla f(\vec{x}^k)\| < \gamma$. This means (3.12) holds all iterations after k_0, and also shows that $\|\nabla f(\vec{x}^k)\|$ converges to zero at least linearly. By definition of $\vec{x}^{k+1}$, we have

$$\frac{1}{m}\|\nabla f(\vec{x}^k)\| \leq \|\vec{x}^{k+1} - \vec{x}^k\| \leq \frac{1}{M}\|\nabla f(\vec{x}^k)\| \tag{3.13}$$

Since gradient converges to zero as $k \to \infty$, by (3.13), $\|\vec{x}^{k+1} - \vec{x}^k\| \to 0$, but this just means that $\delta \to 0$ as $k \to \infty$. By continuity, $\epsilon \to 0$ as $k \to \infty$. However, if we combine (3.11), (3.5), and (3.6), we have

$$f(\vec{x}^{k+1}) - f(\vec{x}^*) \leq \frac{M^3\epsilon^2}{m^5}(f(\vec{x}^k) - f(\vec{x}^*)) \tag{3.14}$$

Hence, when $k \to \infty$, the coefficient $M^3\epsilon^2/m^5 \to 0$, which ensures super-linear convergence.

If further more, the hessian is Lipschitz continuous with Lipschitz constant L, then (3.10) will yield

$$\|\nabla f(\vec{x}^{k+1})\| \leq \frac{L}{2m^2}\|\nabla f(\vec{x}^k)\|^2$$

If we take $\gamma = m^2/L$, then there exists k_0, such that $\|\nabla f(\vec{x}^{k_0})\| < m^2/L$, and since we have

$$a_{k+1} = \frac{L}{2m^2}\|\nabla f(\vec{x}^{k+1})\| \leq \frac{L^2}{4m^4}\|\nabla f(\vec{x}^{k})\|^2 = \left(\frac{L}{2m^2}\|\nabla f(\vec{x}^{k})\|\right)^2 = a_k^2$$

by induction, for all $k \geq k_0$, we havewe have

$$a_k \leq a_{k_0}^{2^{k-k_0}} \leq \left(\frac{1}{2}\right)^{2^{k-k_0}}$$

Therefore, we have

$$f(\vec{x}^k) - f(\vec{x}^*) \leq \frac{1}{2m}\|f(\vec{x}^k)\|^2 \leq \frac{2m^3}{L^2}\left(\frac{1}{2}\right)^{2^{k-k_0+1}}$$

which is exactly quadratic convergence.