Introductory Lectures on Optimization

Acceleration Methods (1)

Hui Qian
qianhui@zju.edu.cn

College of Computer Science, Zhejiang University

October 30, 2024

Outline

Part I
Lower Complexity Bounds for $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

Problem Class

Consider an optimization problem where objective function comes from $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$ (and $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$). The problem class is as follows.

Model	$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}), \quad f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$。
Oracle	First-order local Black Box.
Approximate solution	$\bar{\boldsymbol{x}} \in \mathbb{R}^n : f(\bar{\boldsymbol{x}}) - f^* \leq \epsilon$.

$\mathcal{F}_L^{\infty,1}(\mathbb{R}^n) \subset \mathcal{F}_L^{1,1}(\mathbb{R}^n)$, Therefore, the lower bound obtained in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$ must theoretically be looser than that in $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$.

Assumption

For simplicity, let us introduce the following assumptions for the iterative process.

Assumption 1 (Assumption 2.1.4)

An iterative method $\mathcal{M}$ generates a sequence of test points $\{\boldsymbol{x}_k\}$ such that

$$\boldsymbol{x}_k \in \boldsymbol{x}_0 + \mathrm{Lin}\{\nabla f(\boldsymbol{x}_0), \ldots, \nabla f(\boldsymbol{x}_{k-1})\}, \qquad k \geq 1.$$

This assumption is not absolutely necessary and it can be avoided using more sophisticated reasoning. However, it holds for the majority of practical methods.

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

We can prove the lower complexity bounds for our problem class without developing a resisting oracle. Instead, we just point out the “worst function in the world” belonging to the class $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$.

Let us fix some constant $L > 0$. Consider the following family of quadratic functions

$$f_k(\boldsymbol{x}) = \frac{L}{4}\left\{\frac{1}{2}\left[\left(\boldsymbol{x}^{(1)}\right)^2 + \sum_{i=1}^{k-1}\left(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i+1)}\right)^2 + \left(\boldsymbol{x}^{(k)}\right)^2\right] - \boldsymbol{x}^{(1)}\right\}, \qquad (1)$$

for $k = 1 \ldots n$.

Remark. We have $f_k(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top(\frac{L}{4}A_k)\boldsymbol{x} - \frac{L}{4}\boldsymbol{e}_1^\top\boldsymbol{x}$, and $\nabla f_k(\boldsymbol{x}) = A_k\boldsymbol{x} - \boldsymbol{e}_1$.

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

About the function class: Notes that for all $s \in \mathbb{R}^n$, we have

$$\langle \nabla^2 f_k(\boldsymbol{x})\boldsymbol{s},\ \boldsymbol{s}\rangle = \frac{L}{4}\left[\left(\boldsymbol{s}^{(1)}\right)^2 + \sum_{i=1}^{k-1}\left(\boldsymbol{s}^{(i)} - \boldsymbol{s}^{(i+1)}\right)^2 + \left(\boldsymbol{s}^{(k)}\right)^2\right] \geq 0$$

and (Since $-2ab \leq a^2 + b^2$)

$$\begin{aligned}\langle \nabla^2 f_k(\boldsymbol{x})\boldsymbol{s},\ \boldsymbol{s}\rangle &\leq \frac{L}{4}\left[\left(\boldsymbol{s}^{(1)}\right)^2 + \sum_{i=1}^{k-1} 2\left(\left(\boldsymbol{s}^{(i)}\right)^2 + \left(\boldsymbol{s}^{(i+1)}\right)^2\right) + \left(\boldsymbol{s}^{(k)}\right)^2\right] \\ &\leq L\sum_{i=1}^{n}\left(\boldsymbol{s}^{(i)}\right)^2 = \langle LI_n\boldsymbol{s},\ \boldsymbol{s}\rangle.\end{aligned}$$

Therefore, $0 \preceq \nabla^2 f_k(\boldsymbol{x}) \preceq LI_n$. Thus, $f_k(\boldsymbol{x}) \in \mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$, $1 \leq k \leq n$.

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

About the Hessian: In order to calculate the minimum point of the function f_k, we first discuss the Hessian. We get $\nabla^2 f_k(\boldsymbol{x}) = \frac{L}{4} A_k$, where

$$A_k = \left(\begin{array}{c|c} \begin{array}{c} k \\ \text{lines} \end{array} \left\{ \begin{array}{cccccc} 2 & -1 & 0 & & & \\ -1 & 2 & -1 & & 0 & \\ 0 & -1 & 2 & & & \\ & \ldots & & & \ldots & \\ & & & -1 & 2 & -1 \\ & 0 & & 0 & -1 & 2 \end{array} \right. & O_{n-k,k} \\ \hline O_{n-k,k} & O_{n-k,n-k} \end{array} \right),$$

where $O_{k,p}$ is a $(k \times p)$ zero matrix. $(A_k \preceq 4I$ for $1 \le k \le n.)$

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

About the optimal value: We have

$$\boxed{\nabla f_k(\boldsymbol{x}) = 0 \Rightarrow A_k \boldsymbol{x} - \boldsymbol{e}_1 = 0.}$$

We can construct a special solution as follows (Explained in the next frames) :

$$\bar{\boldsymbol{x}}_k^{(i)} = \begin{cases} 1 - \frac{i}{k+1}, & i = 1 \dots k, \\ 0, & k+1 \le i \le n. \end{cases}$$

Therefore, $A_k \bar{\boldsymbol{x}}_k - \boldsymbol{e}_1 = 0$.

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

About the optimal value: (continued.)

From the function $f_k(\boldsymbol{x})$, we have the following conclusions.

$$2\boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)} = 1 \Leftrightarrow \boldsymbol{x}^{(0)} - \boldsymbol{x}^{(1)} = \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)}, \qquad \text{Here, we set } \boldsymbol{x}^{(0)} = 1,$$
$$\cdots$$
$$\boldsymbol{x}^{(i-1)} - \boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i+1)}, \qquad \text{for } 1 < i < k$$
$$\cdots$$
$$\boldsymbol{x}^{(k-1)} - \boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k+1)}, \qquad \text{Here, we set } \boldsymbol{x}^{(k+1)} = 0,$$

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

About the optimal value: (continued.)

Thus, we have

$$\bar{\boldsymbol{x}}^{(i)} = \bar{\boldsymbol{x}}^{(0)} + id, \text{ for } 1 \leq i \leq k$$

因为 $0 = \bar{\boldsymbol{x}}^{(k+1)} = \bar{\boldsymbol{x}}^{(0)} + (k+1)d$，我们有，对于 $1 \leq i \leq k$，$d = -\frac{1}{k+1}$。因此有

$$\bar{\boldsymbol{x}}^{(i)} = 1 - \frac{i}{k+1}, \text{ for } 1 \leq i \leq k$$

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

About the optimal value: (continued.)

Therefore, the optimal value of f_k is

$$\boxed{\begin{aligned} f_k^* &= \frac{L}{4}\left[\frac{1}{2}\langle A_k\bar{\boldsymbol{x}}_k,\ \bar{\boldsymbol{x}}_k\rangle - \langle \boldsymbol{e}_1,\ \bar{\boldsymbol{x}}_k\rangle\right] \\ &= -\frac{L}{8}\langle \boldsymbol{e}_1,\ \bar{\boldsymbol{x}}_k\rangle \\ &= \frac{L}{8}\left(-1+\frac{1}{k+1}\right). \qquad (2) \end{aligned}}$$

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

Bound of $\bar{\boldsymbol{x}}$:

$$\begin{aligned}\|\bar{\boldsymbol{x}}_k\|^2 &= \sum_{i=1}^{n}\left(\bar{\boldsymbol{x}}_k^{(i)}\right)^2 = \sum_{i=1}^{k}\left(1-\frac{i}{k+1}\right)^2 = k - \frac{2}{k+1}\sum_{i=1}^{k} i + \frac{1}{(k+1)^2}\sum_{i=1}^{k} i^2 \\ &\leq k - \frac{2}{k+1}\cdot\frac{k(k+1)}{2} + \frac{1}{(k+1)^2}\cdot\frac{(k+1)^3}{3} = \boxed{\frac{1}{3}(k+1)}.\end{aligned} \tag{3}$$

The above inequality is from

$$\sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6} \leq \frac{(k+1)^3}{3}. \tag{4}$$

Remark. where the first equation can be proved by mathematical induction.

Worst Function in $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

$\mathbb{R}^{k,n}$ sub-space: Let $\mathbb{R}^{k,n} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{x}^{(i)} = 0, k+1 \leq i \leq n\}$. This is the subspace of $\mathbb{R}^n$ in which only the first k components of the point can differ from zero. From the analytical form of the functions $\{f_k\}$, it is easy to see that for all $\boldsymbol{x} \in \mathbb{R}^{k,n}$, we have

$$f_p(\boldsymbol{x}) = f_k(\boldsymbol{x}), \quad p = k \dots n. \tag{5}$$

For example, $k = 4, p = 5, n = 7$,

$$\langle A_p \boldsymbol{x}, \, \boldsymbol{x} \rangle = (* \quad * \quad * \quad * \quad 0 \quad 0 \quad 0) \begin{pmatrix} * & * & 0 & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 & 0 \\ 0 & * & * & * & 0 & 0 & 0 \\ 0 & 0 & * & * & + & 0 & 0 \\ 0 & 0 & 0 & + & + & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} * \\ * \\ * \\ * \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Important Results

Let us fix some p, $1 \leq p \leq n$.

Lemma 2 (Lemma 2.1.3)

Let $\boldsymbol{x}_0 = 0$. Then for any sequence $\{\boldsymbol{x}_k\}_{k=0}^p$ satisfying the condition

$$\boldsymbol{x}_k \in \mathcal{L}_k = \text{Lin}\{\nabla f_p(\boldsymbol{x}_0), \ldots, \nabla f_p(\boldsymbol{x}_{k-1})\},$$

We have $\mathcal{L}_k \subseteq \mathbb{R}^{k,n}$.

Proof. Consider $\mathcal{L}_1 = \text{Lin}\{\nabla f_p(\boldsymbol{x}_0)\}$. Since $\boldsymbol{x}_0 = 0$, we have $\nabla f_p(\boldsymbol{x}_0) = -\frac{L}{4}\boldsymbol{e}_1 \in \mathbb{R}^{1,n}$. Thus, $\mathcal{L}_1 \equiv \mathbb{R}^{1,n}$. Note that,

$$\boxed{\nabla f_p(\boldsymbol{x}) = \frac{L}{4} A_p \boldsymbol{x} - \frac{L}{4}\boldsymbol{e}_1.}$$

Important Results

Proof. (Continued.)
Suppose that for some $k(k < p)$, the conclusion holds. That is, $\boldsymbol{x}_k \in \mathbb{R}^{k,n}$ ($\boldsymbol{x}_k \in \mathcal{L}_k = \mathrm{Lin}$ $\{ \nabla f_p(\boldsymbol{x}_0), \ldots, \nabla f_p(\boldsymbol{x}_{k-1}) \} \subseteq \mathbb{R}^{k,n}$) .

For the case of $k+1$, since A_p is three-diagonal, for any $\boldsymbol{x}_k \in \mathbb{R}^{k,n}$, we have

$$\nabla f_p(\boldsymbol{x}_k) = \frac{L}{4}(A_p \boldsymbol{x}_k - \boldsymbol{e}_1) \in \mathbb{R}^{k+1,n}$$

Thus $\mathcal{L}_{k+1} \subseteq \mathbb{R}^{k+1,n}$. We can complete the proof by induction. □

Important Results

Corollary 3

For any sequence $\boldsymbol{x}_k \in \mathcal{L}_k$ with $\boldsymbol{x}_0 = 0$ and $\{\boldsymbol{x}_k\}_{k=0}^p$, we have

$$f_p(\boldsymbol{x}_k) \geq f_k^*.$$

Proof. Indeed, $\boldsymbol{x}_k \in \mathcal{L}_k \subseteq \mathbb{R}^{k,n}$ and therefore $f_p(\boldsymbol{x}_k) = f_k(\boldsymbol{x}_k) \geq f_k^*$.

Remark. See the previous (5).

Main Theorem

Now we are ready to prove the main result of this section.

Theorem 4 (Theorem 2.1.7)

For any k, $1 \le k \le \frac{1}{2}(n-1)$, and any $x_0 \in \mathbb{R}^n$ there exists a function $f \in \mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$ such that for any first-order method $\mathcal{M}$ satisying Assumption 1 we have

$$f(\boldsymbol{x}_k) - f^* \ge \frac{3L\left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|^2}{32(k+1)^2},$$

$$\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 \ge \frac{1}{8}\left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|^2,$$

where $\boldsymbol{x}^*$ is the minimum of the function f and $f^* = f(\boldsymbol{x}^*)$.

Main Theorem

Proof. It is clear that the methods of this type are invariant with respect to a simultaneous shift of all objects in the space of variables. Thus, the sequence of iterates, which is generated by such a method for the function $f(\cdot)$ starting from $\boldsymbol{x}_0$, is just a shift of the sequence generated for $\bar{f}(\boldsymbol{x}) = f(\boldsymbol{x} + \boldsymbol{x}_0)$. Therefore, we can assume that $\boldsymbol{x}_0 = 0$.

Let us prove the first inequality. For that, let us fix k and apply $\mathcal{M}$ to minimize $f(\boldsymbol{x}) = f_{2k+1}(\boldsymbol{x})$. Then $\boldsymbol{x}^* = \bar{\boldsymbol{x}}_{2k+1}$ and $f^* = f^*_{2k+1}$. Using Corollary 3, we conclude that

$$f(\boldsymbol{x}_k) \triangleq f_{2k+1}(\boldsymbol{x}_k) = f_k(\boldsymbol{x}_k) \geq f_k^*.$$

Main Theorem

$$\boxed{f(\boldsymbol{x}_k) - f^* \geq \frac{3L\left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|^2}{32(k+1)^2}.}$$

Proof. (Continued.) Hence, since $\boldsymbol{x}_0 = 0$, in view of (2) and (3) we get the following estimate

$$\frac{f(\boldsymbol{x}_k) - f^*}{\left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|^2} = \frac{f(\boldsymbol{x}_k) - f^*}{\left\|\bar{\boldsymbol{x}}_{2k+1}\right\|^2} \geq \frac{f_k^* - f_{2k+1}^*}{\left\|\bar{\boldsymbol{x}}_{2k+1}\right\|^2} \geq \frac{\frac{L}{8}\left(-1 + \frac{1}{k+1} + 1 - \frac{1}{2k+2}\right)}{\frac{1}{3}(2k+2)}$$
$$= \frac{3}{8}L \cdot \frac{1}{4(k+1)^2}.$$

Main Theorem

$$\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 \geq \frac{1}{8}\left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|^2.$$

Proof. (Continued.) Let us prove the second inequality. Since $\boldsymbol{x}_k \in \mathbb{R}^{k,n}$ and $\boldsymbol{x}_0 = 0$, we have

$$\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 \geq \sum_{i=k+1}^{2k+1}\left(\bar{\boldsymbol{x}}_{2k+1}^{(i)}\right)^2 = \sum_{i=k+1}^{2k+1}\left(1 - \frac{i}{2k+2}\right)^2$$

$$= k + 1 - \frac{1}{k+1}\sum_{i=k+1}^{2k+1} i + \frac{1}{4(k+1)^2}\sum_{i=k+1}^{2k+1} i^2.$$

Main Theorem

Proof. (Continued.) In view of (4), we have

$$\begin{aligned}
\sum_{i=k+1}^{2k+1} i^2 &= \sum_{i=0}^{2k+1} i^2 - \sum_{i=0}^{k} i^2 \\
&= \frac{1}{6}\left[(2k+1)(2k+2)(4k+3) - k(k+1)(2k+1)\right] \\
&= \frac{1}{6}(k+1)(2k+1)(7k+6).
\end{aligned}$$

Main Theorem

Proof. (Continued.) Therefore, using (3) we finally obtain

$$\begin{aligned}
\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 &\geq k+1 - \frac{1}{k+1} \cdot \frac{(3k+2)(k+1)}{2} + \frac{(2k+1)(7k+6)}{24(k+1)} \\
&= \frac{(2k+1)(7k+6)}{24(k+1)} - \frac{k}{2} = \frac{2k^2+7k+6}{24(k+1)} \\
&= \frac{2k^2+7k+6}{16(k+1)^2}\left\{\frac{2}{3}(k+1)\right\} \geq \frac{2k^2+7k+6}{16(k+1)^2}\|\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_{2k+1}\|^2 \geq \frac{1}{8}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2 .
\end{aligned}$$

$$\boxed{\|\bar{\boldsymbol{x}}_k\|^2 \leq \frac{1}{3}(k+1)}. \qquad (3)$$

Main Theorem

The above theorem is valid only under the assumption that

> the number of steps of the iterative scheme is not too large as compared with the dimension of the space of variables $(k \leq \frac{1}{2}(n-1))$.

Complexity bounds of this type are called uniform in the dimension. Clearly, they are valid for very large problems, in which we cannot even wait for n iterates of the method. However, even for problems with a moderate dimension, these bounds also provide us with some information.

- Firstly, they describe the potential performance of numerical methods at the initial stage of the minimization process.
- Secondly, they warn us that without a direct use of finite-dimensional arguments we cannot justify a better complexity of the corresponding numerical scheme.

Main Theorem

Let us note that the obtained lower bound for the value of the objective function is rather optimistic. Indeed, after 100 iterations we could decrease the initial residual by 10^4 times. However, the result on the behavior of the minimizing sequence is quite disappointing.

> The convergence to the optimal point can be arbitrarily slow.

The only thing we can do is to try to find problem classes in which the situation could be better.

Part II
Lower Complexity Bounds for $\mathcal{S}_{\mu,L}^{\infty,1}(\mathbb{R}^n)$

Problem Class

Let us obtain the lower complexity bounds for unconstrained minimization of functions from the class $\mathcal{S}^{\infty,1}_{\mu,L}(\mathbb{R}^n) \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$. Consider the following problem class.

Model	$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}), \quad f \in \mathcal{S}^{\infty,1}_{\mu,L}(\mathbb{R}^n),\ \mu > 0.$。
Oracle	First-order local Black Box.
Approximate solution	$\bar{\boldsymbol{x}} : f(\bar{\boldsymbol{x}}) - f^* \leq \epsilon, \|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|^2 \leq \epsilon$。

As in the previous section, we consider methods satisfying Assumption 1. We are going to find the lower complexity bounds for our problems in terms of the condition number $Q_f = \frac{L}{\mu}$.

Worst Function in $\mathcal{S}_{\mu,L}^{\infty,1}(\mathbb{R}^n)$

Note that in the description of our problem class, we do not fix the dimension of the space of variables. Therefore, formally this class also includes an infinite-dimensional problem.

- We are going to give an example of a bad function defined in an infinite-dimensional space.
- It is also possible to do this in finite dimensions, but the corresponding reasoning is more complicated.

Consider $\mathbb{R}^\infty \equiv l_2$, the space of all sequences $x = \{\boldsymbol{x}^{(i)}\}_{i=1}^{\infty}$ with finite standard Euclidean norm

$$\|\boldsymbol{x}\|^2 = \sum_{i=1}^{\infty} \left(\boldsymbol{x}^{(i)}\right)^2 < \infty.$$

Worst Function in $\mathcal{S}^{\infty,1}_{\mu,L}(\mathbb{R}^n)$

Let us choose two parameters, $\mu > 0$ and $Q_f > 1$, which define the following function

$$f_{\mu,Q_f}(\boldsymbol{x}) = \frac{\mu(Q_f - 1)}{8} \left\{ \left(\boldsymbol{x}^{(1)}\right)^2 + \sum_{i=1}^{\infty} \left(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i+1)}\right)^2 - 2\boldsymbol{x}^{(1)} \right\} + \frac{\mu}{2} \|\boldsymbol{x}\|^2 .$$

Compare:

$$f_k(\boldsymbol{x}) = \frac{L}{4} \left\{ \frac{1}{2} \left[\left(\boldsymbol{x}^{(1)}\right)^2 + \sum_{i=1}^{k-1} \left(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i+1)}\right)^2 + \left(\boldsymbol{x}^{(k)}\right)^2 \right] - \boldsymbol{x}^{(1)} \right\},$$

Worst Function in $\mathcal{S}_{\mu,L}^{\infty,1}(\mathbb{R}^n)$

Let $L = \mu Q_f$ and

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & \ddots \\ 0 & 0 & \ddots & \ddots \end{pmatrix}.$$

Then $\nabla^2 f(\boldsymbol{x}) = \frac{\mu(Q_f - 1)}{4} A + \mu I$, where I is the unit operator in $\mathbb{R}^\infty$. As in previous section, we can see that $0 \preceq A \preceq 4I$.

Worst Function in $\mathcal{S}^{\infty,1}_{\mu,L}(\mathbb{R}^n)$

Therefore,

$$\mu I \preceq \nabla^2 f(\boldsymbol{x}) \preceq (\mu(Q_f - 1) + \mu)I = \mu Q_f I.$$

This mean that $f_{\mu,Q_f} \in \mathcal{S}^{\infty,1}_{\mu,\mu Q_f}(\mathbb{R}^n)$. Note taht the condition number of the function f_{μ,Q_f} is

$$Q_{f_{\mu,Q_f}} = \frac{\mu Q_f}{\mu} = Q_f.$$

Worst Function in $\mathcal{S}_{\mu,L}^{\infty,1}(\mathbb{R}^n)$

Let us find the minimum of the function $f_{\mu,\mu Q_f}$. The first-order optimality condition

$$\nabla f_{\mu,\mu Q_f}(\boldsymbol{x}) \equiv \left(\frac{\mu(Q_f-1)}{4} A + \mu I \right) \boldsymbol{x} - \frac{\mu(Q_f-1)}{4} \boldsymbol{e}_1 = 0$$

can be written as

$$\left(A + \frac{4}{Q_f-1} I \right) \boldsymbol{x} = \boldsymbol{e}_1.$$

The coordinate form of this equation is as follows:

$$2\frac{Q_f+1}{Q_f-1}\boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)} = 1,$$

$$\boldsymbol{x}^{(k+1)} - 2\frac{Q_f+1}{Q_f-1}\boldsymbol{x}^{(k)} + \boldsymbol{x}^{(k-1)} = 0, \quad k = 2, \ldots \tag{6}$$

Worst Function in $\mathcal{S}_{\mu,L}^{\infty,1}(\mathbb{R}^n)$

Let's guess the solution to the problem:

$$\begin{aligned} \boldsymbol{x}^{(2)} - 2\frac{Q_f+1}{Q_f-1}\boldsymbol{x}^{(1)} + 1 &= 0, \\ \boldsymbol{x}^{(k+1)} - 2\frac{Q_f+1}{Q_f-1}\boldsymbol{x}^{(k)} + \boldsymbol{x}^{(k-1)} &= 0. \end{aligned}$$

Set $\boldsymbol{x}^{(0)} = 1$, assume that $\boldsymbol{x}^{(1)} = q$, we have

$$\begin{aligned} \boldsymbol{x}^{(2)} - 2\frac{Q_f+1}{Q_f-1}\boldsymbol{x}^{(1)} + \boldsymbol{x}^{(0)} &= 0, \\ \boldsymbol{x}^{(3)} - 2\frac{Q_f+1}{Q_f-1}\boldsymbol{x}^{(2)} + q &= 0. \\ &\cdots \end{aligned}$$

Worst Function in $\mathcal{S}_{\mu,L}^{\infty,1}(\mathbb{R}^n)$

Thus, compare the following two equation:

$$\begin{aligned} \boldsymbol{x}^{(2)} q - 2\frac{Q_f+1}{Q_f-1}\boldsymbol{x}^{(1)} q + \boldsymbol{x}^{(0)} q &= 0, \\ \boldsymbol{x}^{(3)} - 2\frac{Q_f+1}{Q_f-1}\boldsymbol{x}^{(2)} + \boldsymbol{x}^{(1)} &= 0. \\ &\dots \end{aligned}$$

It implies that $\boldsymbol{x}^{(k)} q = \boldsymbol{x}^{(k+1)}$, thus $\boldsymbol{x}^{(k)} = q^k$. Let q be the smallest root of the equation

$$q^2 - 2\frac{Q_f+1}{Q_f-1}q + 1 = 0.$$

We have $q = \left(\sqrt{Q_f} - 1\right) / \left(\sqrt{Q_f} + 1\right)$.

Main Theorem

Theorem 5

For any $\boldsymbol{x}_0 \in \mathbb{R}^\infty$ and any constant $\mu > 0$, $Q_f > 1$, there exists a function $f \in \mathcal{S}^{\infty,1}_{\mu,\mu Q_f}(\mathbb{R}^n)$ such that for any first-order method$\mathcal{M}$ satisfying assumption 1, we have

$$\|\boldsymbol{x}_k - x^*\|^2 \geq \left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k} \|\boldsymbol{x}_0 - x^*\|^2,$$

$$f(\boldsymbol{x}_k) - f^* \geq \frac{\mu}{2}\left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k} \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2,$$

where $\boldsymbol{x}^*$ is the unique unconstrained minimum of function f.

Main Theorem

Proof. Indeed, we can assume that $\boldsymbol{x}_0 = 0$. Let us choose $f(\boldsymbol{x}) = f_{\mu,\mu Q_f}(\boldsymbol{x})$. Then

$$\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2 = \sum_{i=1}^{\infty} \left[(\boldsymbol{x}^*)^{(i)}\right]^2 = \sum_{i=1}^{\infty} q^{2i} = \frac{q^2}{1-q^2}. \tag{7}$$

Since $\nabla^2 f_{\mu,\mu Q_f}(\boldsymbol{x})$ is a tri-diagonal operator and $\nabla f_{\mu,\mu Q_f}(0) = -\frac{\mu(Q_f-1)}{4}\boldsymbol{e}_1$, we conclude that $\boldsymbol{x}_k \in \mathbb{R}^{k,\infty}$. Therefore

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 \geq \sum_{i=k+1}^{\infty} \left[(\boldsymbol{x}^*)^{(i)}\right]^2 = \sum_{i=k+1}^{\infty} q^{2i} = \frac{q^{2(k+1)}}{1-q^2} = q^{2k} \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2 .$$

The second bound of this theorem follows form (7) and the definition of strongly convex. $\square$

Main Theorem

Remark. (1):

$\mathcal{L}_1 = \text{Lin}\{\nabla f(\boldsymbol{x}_0)\} \subseteq \mathbb{R}^{1,\infty}$, since $\nabla f(\boldsymbol{x}_0) = \nabla f(0) = -\frac{\mu(Q_f-1)}{4} e_1$。
Suppose that $\mathcal{L}_k = \text{Lin}\{\nabla f(\boldsymbol{x}_0), \cdots, \nabla f(\boldsymbol{x}_{k-1})\} \subseteq \mathbb{R}^{k,\infty}$ holds。
Consider $\nabla f(\boldsymbol{x}_k)$. Since $\boldsymbol{x}_k \in \mathbb{R}^{k,\infty}$ and A is a tri-diagonal operator, we have $\nabla f(\boldsymbol{x}_k) \in \mathbb{R}^{k+1,\infty}$. Thus, we have

$$\mathcal{L}_{k+1} = \text{Lin}\{\nabla f(\boldsymbol{x}_0), \cdots, \nabla f(\boldsymbol{x}_k)\} \subseteq \mathbb{R}^{k+1,\infty}.$$

Remark. (2):

The second inequality comes from

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) - \langle \nabla f^*(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \geq \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{x}^*\|^2.$$

References I

Lloyd N Trefethen and David Bau III. *Numerical linear algebra*, volume 50. Siam, 1997.

Stephen Wright, Jorge Nocedal, et al. Numerical optimization. *Springer Science*, 35(67-68): 7, 1999.

Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards*, 49:409–435, 1952.

Stephen J Wright and Benjamin Recht. *Optimization for data analysis*. Cambridge University Press, 2022.

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87. Springer Science & Business Media, 2003.

Yurii Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018.

Thank You!

Email:qianhui@zju.edu.cn