Part II. Smooth Convex Programming.

Lecture 2.

Minimization of Smooth Functions.

– Smooth convex functions.

– Lower complexity bounds.

– Strongly convex functions.

– Lower complexity bounds.

– Gradient method.

Smooth convex functions.

Problem:

$$\min_{x \in R^n} f(x), \quad f \in C^1(R^n).$$

The class of general smooth functions is hopeless for optimization.

What could be a good class?

Desired Properties:

1. If $f \in \mathcal{F}$ then

$$f'(\bar{x}) = 0 \quad \Rightarrow \quad f(x) \geq f(\bar{x}) \quad \forall x \in R^n.$$

2. If $f_1, f_2 \in \mathcal{F}$ and $\alpha, \beta \geq 0$ then

$$\alpha f_1 + \beta f_2 \in \mathcal{F}.$$

3. $\alpha + \langle a, x \rangle \in \mathcal{F}$.

What it should be?

Let $f \in \mathcal{F}$. Let us fix some $x_0 \in R^n$ and consider the function

$$\phi(y) = f(y) - \langle f'(x_0), y \rangle.$$

Then $\phi \in \mathcal{F}$ in view of 2) and 3).

Note that

$$\phi'(y) \mid_{y=x_0} = f'(y) \mid_{y=x_0} - f'(x_0) = 0.$$

Therefore, in view of 1) for any $y \in R^n$ we have:

$$\phi(y) \geq \phi(x_0) = f(x_0) - \langle f'(x_0), x_0 \rangle.$$

That is

$$f(y) \geq f(x_0) + \langle f'(x_0), x - x_0 \rangle.$$

Definition. A continuously differentiable function $f(x)$ is called *convex* on R^n ($f \in \mathcal{F}^1(R^n)$) if for any $x, y \in R^n$ we have:

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle. \tag{2.1}$$

Notation: $f \in \mathcal{F}_L^{k,l}(Q)$. The meaning of the indices is the same as for the class $C_L^{k,l}(Q)$.

If $-f(x)$ is convex, we call it *concave*.

Properties of convex functions

Theorem 2.1 *If* $f \in \mathcal{F}^1(R^n)$ *and* $f'(x^*) = 0$ *then*

$$f(x) \geq f(x^*) \quad \forall x \in R^n.$$

Proof. In view of the definition for any $x \in R^n$ we have

$$f(x) \geq f(x^*) + \langle f'(x^*), x - x^* \rangle = f(x^*).$$

□

Lemma 2.1 *If* $f_1, f_2 \in \mathcal{F}^1(R^n)$ *and* $\alpha, \beta \geq 0$ *then*

$$f = \alpha f_1 + \beta f_2 \in \mathcal{F}^1(R^n).$$

Proof. For any $x, y \in R^n$ we have:

$$\begin{aligned} f_1(y) &\geq f_1(x) + \langle f_1'(x), y - x \rangle, \\ f_2(y) &\geq f_2(x) + \langle f_2'(x), y - x \rangle. \end{aligned}$$

It remains to multiply these equations by α and β and add the results. □

Lemma 2.2 *If* $f \in \mathcal{F}^1(R^m)$, $b \in R^m$ *and* $A : R^n \to R^m$ *then*

$$\phi(x) = f(Ax + b) \in \mathcal{F}^1(R^n).$$

Proof. Indeed, let $x, y \in R^n$. Denote

$$\bar{x} = Ax + b, \quad \bar{y} = Ay + b.$$

Since $\phi'(x) = A^T f'(Ax + b)$, we have:

$$\begin{aligned} \phi(y) &= f(\bar{y}) \geq f(\bar{x}) + \langle f'(\bar{x}), \bar{y} - \bar{x} \rangle \\ &= \phi(x) + \langle f'(\bar{x}), A(y - x) \rangle \\ &= \phi(x) + \langle A^T f'(\bar{x}), y - x \rangle \\ &= \phi(x) + \langle \phi'(x), y - x \rangle. \end{aligned}$$

□

Equivalent Definitions

Theorem 2.2 *Function $f \in \mathcal{F}^1(R^n)$ if and only if it is continuously differentiable and for any $x, y \in R^n$ and $\alpha \in [0,1]$ we have:*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y). \qquad (2.2)$$

Proof. Denote $x_\alpha = \alpha x + (1-\alpha)y$.

1. Let $f \in \mathcal{F}^1(R^n)$. then

$$\begin{aligned} f(x_\alpha) &\leq f(y) + \langle f'(x_\alpha), y - x_\alpha \rangle \\ &= f(y) + \alpha \langle f'(x_\alpha), y - x \rangle, \\ f(x_\alpha) &\leq f(x) + \langle f'(x_\alpha), x - x_\alpha \rangle \\ &= f(x) - (1-\alpha)\langle f'(x_\alpha), y - x \rangle. \end{aligned}$$

Multiplying the first inequality by $(1-\alpha)$, the second one by α and adding the results, we get (2.2).

2. Let (2.2) be true for all $x, y \in R^n$ and $\alpha \in [0,1]$.

Let us choose $\alpha \in [0,1)$. Then

$$\begin{aligned} f(y) &\geq \tfrac{1}{1-\alpha}[f(x_\alpha) - \alpha f(x)] \\ &= f(x) + \tfrac{1}{1-\alpha}[f(x_\alpha) - f(x)] \\ &= f(x) + \tfrac{1}{1-\alpha}[f(x + (1-\alpha)(y-x)) - f(x)]. \end{aligned}$$

Tending α to 1, we get (2.1). □

Theorem 2.3 *Function $f \in \mathcal{F}^1(R^n)$ if and only if it is continuously differentiable and for any $x, y \in R^n$ we have:*

$$\langle f'(x) - f'(y), x - y \rangle \geq 0. \qquad (2.3)$$

Proof. 1. Let f be a convex continuously differentiable function. Then

$$\begin{aligned} f(x) &\geq f(y) + \langle f'(y), x - y \rangle, \\ f(y) &\geq f(x) + \langle f'(x), y - x \rangle, \end{aligned}$$

Adding these inequalities, we get (2.3).

2. Let (2.3) holds for all $x, y \in R^n$. Denote

$$x_\tau = x + \tau(y - x).$$

Then

$$\begin{aligned} f(y) &= f(x) &&+ \int_0^1 \langle f'(x + \tau(y - x)), y - x \rangle d\tau \\ &= f(x) &&+ \langle f'(x), y - x \rangle \\ & &&+ \int_0^1 \langle f'(x_\tau) - f'(x), y - x \rangle d\tau \\ &= f(x) &&+ \langle f'(x), y - x \rangle \\ & &&+ \int_0^1 \frac{1}{\tau} \langle f'(x_\tau) - f'(x), x_\tau - x \rangle d\tau \\ &\geq f(x) &&+ \langle f'(x), y - x \rangle. \end{aligned}$$

□

Theorem 2.4 *Function* $f \in \mathcal{F}^2(R^n)$ *if and only if it is twice continuously differentiable and for any* $x \in R^n$ *we have:*

$$f''(x) \geq 0. \qquad (2.4)$$

Proof. 1. Let $f \in C^2(R^n)$ be convex. Denote

$$x_\tau = x + \tau s, \quad \tau > 0.$$

Then, in view of (2.3), we have:

$$\begin{aligned} 0 &\leq \tfrac{1}{\tau^2}\langle f'(x_\tau) - f'(x), x_\tau - x\rangle = \tfrac{1}{\tau}\langle f'(x_\tau) - f'(x), s\rangle \\ &= \tfrac{1}{\tau}\int_0^\tau \langle f''(x + \lambda s)s, s\rangle d\lambda, \end{aligned}$$

and we get (2.4) by tending $\tau \to 0$.

2. Let (2.4) holds for all $x \in R^n$. Then

$$\begin{aligned} f(y) &= f(x) + \langle f'(x), y - x\rangle \\ &\quad + \int_0^1 \int_0^\tau \langle f''(x + \lambda(y - x))(y - x), y - x\rangle d\lambda d\tau \\ &\geq f(x) + \langle f'(x), y - x\rangle. \end{aligned}$$

$\square$

Examples $(f \in \mathcal{F}^1(R^n))$.

1. Linear function

$$f(x) = \alpha + \langle a, x \rangle$$

is convex.

2. Let the matrix A be symmetric and positive semidefinite. Then the quadratic function

$$f(x) = \alpha + \langle a, x \rangle + \tfrac{1}{2}\langle Ax, x \rangle$$

is convex (since $f''(x) = A$).

3. The following functions of one variable belong to $\mathcal{F}^1(R)$:

$$\begin{aligned} f(x) &= e^x, \\ f(x) &= \mid x \mid^p, \quad p > 1, \\ f(x) &= \tfrac{x^2}{1-|x|}, \\ f(x) &= \mid x \mid - \ln(1+ \mid x \mid), \end{aligned}$$

and many others.

Geometric Programming:

$$f(x) = \sum_{i=1}^{m} e^{\alpha_i + \langle a_i, x \rangle}.$$

L_p-approximation:

$$f(x) = \sum_{i=1}^{m} \mid \langle a_i, x \rangle - b_i \mid^p .$$

Class $\mathcal{F}_L^{1,1}(R^n)$

Theorem 2.5 *For inclusion $f \in \mathcal{F}_L^{1,1}(R^n)$ all of the following conditions are equivalent:*

$$\begin{aligned} 0 &\leq f(y) - f(x) - \langle f'(x), y - x \rangle \\ &\leq \tfrac{L}{2} \| x - y \|^2 . \end{aligned} \tag{2.5}$$

$$\begin{aligned} f(y) \geq f(x) &+ \langle f'(x), y - x \rangle \\ &+ \tfrac{1}{2L} \| f'(x) - f'(y) \|^2 . \end{aligned} \tag{2.6}$$

$$\langle f'(x) - f'(y), x - y \rangle \geq \tfrac{1}{L} \| f'(x) - f'(y) \|^2 . \tag{2.7}$$

Proof.. Indeed, (2.5) follows from Lemma 1.2.3.
Further, let us fix $x_0 \in R^n$. Consider the function

$$\phi(y) = f(y) - \langle f'(x_0), y \rangle.$$

Note that $\phi \in \mathcal{F}_L^{1,1}(R^n)$ and $y^* = x_0$. Therefore, in view of (2.5), we have:

$$\phi(y^*) \leq \phi(y - \frac{1}{L}\phi'(y)) \leq \phi(y) - \frac{1}{2L} \| \phi'(y) \|^2 .$$

And we get (2.6) since $\phi'(y) = f'(y) - f'(x_0)$.

We obtain (2.7) from (2.6) by adding two inequalities with x and y interchanged.

Finally, from (2.7) we get

$$\| f'(x) - f'(y) \| \leq L \| x - y \| . \quad \square$$

Lower complexity bounds for $\mathcal{F}_L^{\infty,1}(R^n)$

Problem formulation:

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

Problem class:

$$f \in \mathcal{F}_L^{1,1}(R^n).$$

Oracle:
First-order black box.

Approximate solution:
Find $\bar{x} \in R^n$ such that

$$f(\bar{x}) - f^* \leq \epsilon.$$

Methods:
Generate a sequence $\{x_k\}$:

$$x_k \in x_0 + \mathrm{Lin}\,\{f'(x_0), \ldots, f'(x_{k-1})\}.$$

Consider the family of functions

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

$$+(x^{(k)})^2] - x^{(1)}\}, \quad k = 1, \ldots, n.$$

Denote $R^{k,n} = \{x \in R^n \mid \ x^{(i)} = 0, \ k+1 \leq i \leq n\}$. Then

$$f_{k+p}(x) = f_k(x), \quad \forall x \in R^{k,n}, \ p \geq 1.$$

Note that for all $h \in R^n$ we have:

$$\langle f_k''(x)h, h\rangle$$

$$= \frac{L}{4}[(h^{(1)})^2 + \sum_{i=1}^{k-1}(h^{(i)} - h^{(i+1)})^2 + (h^{(k)})^2] \geq 0,$$

and

$$\langle f_k''(x)h, h\rangle$$

$$\leq \frac{L}{4}[(h^{(1)})^2 + \sum_{i=1}^{k-1} 2((h^{(i)})^2 + (h^{(i+1)})^2) + (h^{(k)})^2]$$

$$\leq L \sum_{i=1}^{n}(h^{(i)})^2.$$

Thus, $0 \leq f_k''(x) \leq LI_n$. Therefore

$$f_k(x) \in \mathcal{F}_L^{\infty,1}(R^n).$$

Since $f_k''(x) = \frac{L}{4} A_k$, where

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

the equation $f_k'(x) = 0$ (that is $A_k x = e_1$) has the solution

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

Therefore

$$\begin{aligned} f_k^* &= \tfrac{L}{4}[\tfrac{1}{2}\langle A_k \bar{x}_k, \bar{x}_k \rangle - \langle e_1, \bar{x}_k \rangle] \\ &= -\tfrac{L}{8}\langle e_1, \bar{x}_k \rangle = \tfrac{L}{8}\left(-1 + \tfrac{1}{k+1}\right). \end{aligned}$$

Note that

$$\| \bar{x}_k \|^2 = \sum_{i=1}^{n} (\bar{x}_k^{(i)})^2 = \sum_{i=1}^{k} \left(1 - \tfrac{i}{k+1}\right)^2$$

$$= k - \tfrac{2}{k+1} \sum_{i=1}^{k} i + \tfrac{1}{(k+1)^2} \sum_{i=1}^{k} i^2$$

$$\leq k - \tfrac{2}{k+1} \cdot \tfrac{k(k+1)}{2} + \tfrac{1}{(k+1)^2} \cdot \tfrac{(k+1)^3}{3} = \tfrac{1}{3}(k+1).$$

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

Lemma 2.3 *Let $x_0 = 0$. Then for any sequence $\{x_k\}_{k=0}^p$:*

$$x_k \in \mathcal{L}_k = \text{Lin}\,\{f_p'(x_0), \ldots, f_p'(x_{k-1})\},$$

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

Proof. 1. Since $x_0 = 0$ we have

$$f_p'(x_0) = -\frac{L}{4} e_1 \in R^{1,n}.$$

Therefore $\mathcal{L}_1 \equiv R^{1,n}$.

2. Let $\mathcal{L}_k \subseteq R^{k,n}$ for some $k < p$. Since A_p is three-diagonal, we have

$$f_p'(x) \in R^{k+1,n}, \quad \forall x \in R^{k,n}.$$

Therefore $\mathcal{L}_{k+1} \subseteq R^{k+1,n}$. □

Corollary 2.1 *For any sequence $\{x_k\}_{k=0}^p$ such that*

$$x_0 = 0, \quad x_k \in \mathcal{L}_k,$$

we have $f_p(x_k) \geq f_k^*$.

Proof. Indeed, $x_k \in R^{k,n}$ and therefore

$$f_p(x_k) = f_k(x_k) \geq f_k^*.$$

□

Theorem 2.6 *For any k, $1 \leq k \leq \frac{1}{2}(n-1)$, and any $x_0 \in R^n$ there exists a function $f \in \mathcal{F}_L^{\infty,1}(R^n)$ such that for any first order method $\mathcal{M}$ generating a sequence*

$$x_k \in x_0 + \mathrm{Lin}\,\{f'(x_0), \ldots, f'(x_{k-1})\},$$

we have

$$f(x_k) - f^* \geq \frac{3L\|x_0 - x^*\|^2}{32(k+1)^2},$$

$$\| x_k - x^* \|^2 \geq \frac{1}{32} \| x_0 - x^* \|^2 .$$

Proof. 1. Indeed, let us fix k and apply $\mathcal{M}$ to minimizing $f(x) = f_{2k+1}(x)$. Then

$$x^* = \bar{x}_{2k+1}, \quad f^* = f^*_{2k+1}$$

and $f(x_k) = f_{2k+1}(x_k) = f_k(x_k) \geq f^*_k$.

Therefore

$$\frac{f(x_k)-f^*}{\|x_0-x^*\|^2} \geq \frac{\frac{L}{8}\left(-1+\frac{1}{k+1}+1-\frac{1}{2k+2}\right)}{\frac{1}{3}(2k+2)} = \frac{3}{8}L \cdot \frac{1}{4(k+1)^2}.$$

2. Since $x_k \in R^{k,n}$ we have:

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

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

$$\geq k+1 - \frac{1}{k+1} \cdot \frac{(k+1)(3k+2)}{2} + \frac{1}{4(k+1)^2} \cdot \frac{(2k+1)^3 - k^3}{3}$$

$$= \frac{k^2-k+1}{12(k+1)} \geq \frac{k+1}{48} \geq \frac{1}{32} \| x_0 - \bar{x}_{2k+1} \|^2 . \quad \square$$

Note:

1. Using a more sophisticated analysis we can prove the following *exact* lower bound:

$$f(x_k) - f^* \geq \frac{L \parallel x_0 - x^* \parallel^2}{8(k+1)^2}.$$

2. It is possible to prove also that

$$\parallel x_k - x^* \parallel^2 \geq \beta \parallel x_0 - x^* \parallel^2,$$

where the constant β can be *arbitrary* close to one.

Conclusion:

1. The lower bound for the value of the objective is rather optimistic: one hundred iteration could decrease the initial residual in 10^4 times.

2. In general, we cannot guarantee *any* rate of convergence for the minimizing sequence.

Strongly convex functions

Problem:

$$\min_{x \in R^n} f(x), \quad f \in \mathcal{F}^1(R^n).$$

What could we assume to guarantee the following:

- Uniqueness of the solution.
- Fast convergence to the minimizer.

Desired property:

If $f \in \mathcal{F}$ then there exists $\mu > 0$ such that

$$f'(\bar{x}) = 0 \;\Rightarrow\; f(x) \geq f(\bar{x}) + \tfrac{1}{2}\mu \parallel x - \bar{x} \parallel^2 \;\; \forall x \in R^n.$$

Using the similar arguments we come to the following definition.

Definition. A continuously differentiable function $f(x)$ is called *strongly convex* on R^n ($f \in \mathcal{S}^1_\mu(R^n)$) if there exists a constant $\mu > 0$ such that for any $x, y \in R^n$ we have:

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle + \tfrac{1}{2}\mu \parallel y - x \parallel^2 . \quad (2.8)$$

Notation: $f \in \mathcal{S}^{k,l}_{\mu,L}(Q)$. The meaning of the indices k, l and L is the same as for the class $C^{k,l}_L(Q)$.

Properties of strongly convex functions

Theorem 2.7 *If* $f \in \mathcal{S}^1_\mu(R^n)$ *and* $f'(x^*) = 0$ *then*

$$f(x) \geq f(x^*) + \tfrac{1}{2}\mu \parallel x - x^* \parallel^2 \quad \forall x \in R^n.$$

Proof. Since $f'(x^*) = 0$, in view of the definition for any $x \in R^n$ we have

$$\begin{aligned} f(x) &\geq f(x^*) + \langle f'(x^*), x - x^* \rangle + \tfrac{1}{2}\mu \parallel x - x^* \parallel^2 \\ &= f(x^*) + \tfrac{1}{2}\mu \parallel x - x^* \parallel^2 . \end{aligned}$$

□

Lemma 2.4 *If* $f_1 \in \mathcal{S}^1_{\mu_1}(R^n)$, $f_2 \in \mathcal{S}^1_{\mu_2}(R^n)$ *and* α, $\beta \geq 0$ *then*

$$f = \alpha f_1 + \beta f_2 \in \mathcal{S}^1_{\alpha\mu_1 + \beta\mu_2}(R^n).$$

Proof. For any $x, y \in R^n$ we have:

$$\begin{aligned} f_1(y) &\geq f_1(x) + \langle f_1'(x), y - x \rangle + \tfrac{1}{2}\mu_1 \parallel y - x \parallel^2 \\ f_2(y) &\geq f_2(x) + \langle f_2'(x), y - x \rangle + \tfrac{1}{2}\mu_2 \parallel y - x \parallel^2 . \end{aligned}$$

It remains to multiply these equations by α and β and add the results. □

.

Note: $\mathcal{S}^1_0(R^n) \equiv \mathcal{F}^1(R^n)$. Therefore adding a strongly convex function with a convex function we get a strongly convex function with the same constant μ.

Equivalent Definitions

Theorem 2.8 *Function* $f \in \mathcal{S}^1_\mu(R^n)$ *if and only if it is continuously differentiable and for any* $x, y \in R^n$ *and* $\alpha \in [0,1]$ *we have:*

$$\begin{aligned} f(\alpha x + (1-\alpha)y) \leq \; & \alpha f(x) + (1-\alpha) f(y) \\ & -\alpha(1-\alpha)\tfrac{\mu}{2} \parallel x - y \parallel^2 . \end{aligned} \tag{2.9}$$

Theorem 2.9 *Function* $f \in \mathcal{S}^1_\mu(R^n)$ *if and only if it is continuously differentiable and for any* $x, y \in R^n$ *we have:*

$$\langle f'(x) - f'(y), x - y \rangle \geq \mu \parallel x - y \parallel^2 . \tag{2.10}$$

Theorem 2.10 *Function* $f \in \mathcal{S}^2_\mu(R^n)$ *if and only if it is twice continuously differentiable and for any* $x \in R^n$ *we have:*

$$f''(x) \geq \mu I_n. \tag{2.11}$$

The proofs of these theorems are very similar to those of Theorems 2.2 – 2.4.

Examples

1. $f(x) = \frac{1}{2} \parallel x \parallel^2 \in \mathcal{S}_1^2(R^n)$ since $f''(x) = I_n$.

2. Let the symmetric matrix A satisfy the condition:

$$\mu I_n \leq A \leq L I_n.$$

Then

$$f(x) = \alpha + \langle a, x \rangle + \frac{1}{2}\langle Ax, x \rangle \in \mathcal{S}_{\mu,L}^{\infty,1}(R^n) \subset \mathcal{S}_{\mu,L}^{1,1}(R^n)$$

(since $f''(x) = A$).

Other examples can be obtained as a sum of convex and strongly convex functions.

Class $\mathcal{S}^{1,1}_{\mu,L}(R^n)$

Conditions:

$$\mu \parallel x - y \parallel^2 \leq \langle f'(x) - f'(y), x - y \rangle, \qquad (2.12)$$

$$\parallel f'(x) - f'(y) \parallel \leq L \parallel x - y \parallel . \qquad (2.13)$$

The value $Q_f = L/\mu$ (≥ 1) is called the *condition number* of the function f.

Theorem 2.11 *If $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$ then for any x, $y \in R^n$ we have:*

$$\begin{aligned} &\langle f'(x) - f'(y), x - y \rangle \\ &\geq \tfrac{\mu L}{\mu + L} \parallel x - y \parallel^2 + \tfrac{1}{\mu + L} \parallel f'(x) - f'(y) \parallel^2 . \end{aligned} \qquad (2.14)$$

Proof. Consider $\phi(x) = f(x) - \frac{1}{2}\mu \parallel x \parallel^2$. Note that $\phi'(x) = f'(x) - \mu x$. Therefore this function is convex (see Theorem 2.3). Moreover, in view of (2.5)

$$\begin{aligned} \phi(y) = f(y) - \tfrac{1}{2}\mu \parallel y \parallel^2 &\leq f(x) + \langle f'(x), y - x \rangle \\ &+ \tfrac{1}{2}L \parallel x - y \parallel^2 - \tfrac{1}{2}\mu \parallel y \parallel^2 \\ = \phi(x) + \langle \phi'(x), y - x \rangle &+ \tfrac{1}{2}(L - \mu) \parallel x - y \parallel^2 . \end{aligned}$$

Therefore $\phi \in \mathcal{F}^{1,1}_{L-\mu}(R^n)$ (see Theorem 2.5). Thus,

$$\langle \phi'(x) - \phi'(y), y - x \rangle \geq \tfrac{1}{L-\mu} \parallel \phi'(x) - \phi'(y) \parallel^2$$

and that is exactly (2.14). □

Lower complexity bounds for $\mathcal{S}^{1,1}_{\mu,L}(R^n)$

Problem formulation:

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

Problem class:

$$f \in \mathcal{S}^{1,1}_{\mu,L}(R^n).$$

Oracle:
First-order black box.

Approximate solution:
Find $\bar{x} \in R^n$ such that

$$f(\bar{x}) - f^* \leq \epsilon, \quad \| \bar{x} - x^* \|^2 \leq \epsilon,$$

Methods:
Generate a sequence $\{x_k\}$:

$$x_k \in x_0 + \mathrm{Lin}\,\{f'(x_0), \ldots, f'(x_{k-1})\}.$$

Simplification:

We allow also the value $n = \infty$.

We will find the lower complexity bounds in terms of *condition number.*

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

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

Let us choose some parameters $\mu > 0$, $Q > 1$.

Consider the function

$$f(x) = \frac{1}{2}\mu \| x \|^2$$

$$+\frac{\mu(Q-1)}{4}\left\{\frac{1}{2}[(x^{(1)})^2 + \sum_{i=1}^{\infty}(x^{(i)} - x^{(i+1)})^2] - x^{(1)}\right\}.$$

Denote

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

Then $f''(x) = \frac{\mu(Q-1)}{4}A + \mu I$. We have seen that

$$0 \leq A \leq 4I.$$

Therefore

$$\mu I \leq f''(x) \leq (\mu(Q-1) + \mu)I = \mu Q I.$$

Thus, $f \in \mathcal{S}^{\infty,1}_{\mu,\mu Q}(R^\infty)$ and its condition number is

$$Q_f = \frac{\mu Q}{\mu} = Q.$$

Let us find the minimum of $f(x)$:

$$f'(x) \equiv \left(\frac{\mu(Q-1)}{4}A + \mu I\right)x - \frac{\mu(Q-1)}{4}e_1 = 0.$$

That can be written as:

$$\left(A + \frac{4}{Q-1}I\right)x = e_1.$$

The coordinate form of that is

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

Let q be the smallest root of the equation

$$q^2 - 2\frac{Q+1}{Q-1}q + 1 = 0,$$

that is $q = \frac{\sqrt{Q}-1}{\sqrt{Q}+1}$. Then the sequence

$$(x^*)^{(k)} = q^k, \quad k = 1, 2, \ldots,$$

satisfies our system.

Thus, we come to the following result.

Theorem 2.12 *For any $x_0 \in R^\infty$ and any constants $\mu > 0$, $Q > 1$ there exists a function $f \in \mathcal{S}^{\infty,1}_{\mu,\mu Q}(R^\infty)$ such that for any first order method $\mathcal{M}$ generating a sequence*

$$x_k \in x_0 + \mathrm{Lin}\,\{f'(x_0), \ldots, f'(x_{k-1})\},$$

we have

$$\| x_k - x^* \|^2 \geq \left(\frac{\sqrt{Q}-1}{\sqrt{Q}+1}\right)^{2k} \| x_0 - x^* \|^2,$$

$$f(x_k) - f^* \geq \frac{\mu}{2}\left(\frac{\sqrt{Q}-1}{\sqrt{Q}+1}\right)^{2k} \| x_0 - x^* \|^2 .$$

Proof. Indeed, let $x_0 = 0$. Then

$$\| x_0 - x^* \|^2 = \sum_{i=1}^{\infty} [(x^*)^{(i)}]^2 = \sum_{i=1}^{\infty} q^{2i} = \frac{q^2}{1-q^2}.$$

Since $f''(x)$ is a 3-diagonal operator and $f'(0) = e_1$, we have $x_k \in R^{k,\infty}$. Therefore

$$\begin{aligned}\| x_k - x^* \|^2 &\geq \sum_{i=k+1}^{\infty} [(x^*)^{(i)}]^2 = \sum_{i=k+1}^{\infty} q^{2i} \\ &= \frac{q^{2(k+1)}}{1-q^2} = q^{2k} \| x_0 - x^* \|^2 .\end{aligned}$$

The second estimate follows from the first one and the definition. □

Gradient Method

Problem:

$$\min_{x \in R^n} f(x), \quad f \in \mathcal{F}_L^{1,1}(R^n).$$

Scheme:

0. Choose $x_0 \in R^n$.

1. kth iteration ($k \geq 0$).

   a). Compute $f(x_k)$ and $f'(x_k)$.

   b). Find $x_{k+1} = x_k - h_k f'(x_k)$

In what follows we analyze this scheme in the simplest case:

$$h_k = h > 0.$$

Theorem 2.13 *If* $f \in \mathcal{F}_L^{1,1}(R^n)$ *and* $0 < h < \frac{2}{L}$ *then*

$$f(x_k) - f^* \leq \frac{2(f(x_0)-f^*)\|x_0-x^*\|^2}{2\|x_0-x^*\|^2+(f(x_0)-f^*)h(2-Lh)k}.$$

Proof. Denote $r_k = \| x_k - x^* \|$. Then

$$\begin{aligned} r_{k+1}^2 &= \| x_k - x^* - hf'(x_k) \|^2 \\ &= r_k^2 - 2h\langle f'(x_k), x_k - x^*\rangle + h^2 \| f'(x_k) \|^2 \\ &\leq r_k^2 - h(\tfrac{2}{L} - h) \| f'(x_k) \|^2 \end{aligned}$$

(we use (2.7) and $f'(x^*) = 0$). Therefore $r_k \leq r_0$.

In view of (2.5) we have:

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k\rangle \\ &+\tfrac{L}{2} \| x_{k+1} - x_k \|^2 = f(x_k) - \omega \| f'(x_k) \|^2 \end{aligned}$$

with $\omega = h(1 - \frac{L}{2}h)$. Denote $\Delta_k = f(x_k) - f^*$. Then

$$\Delta_k \leq \langle f'(x_k), x_k - x^*\rangle \leq r_0 \| f'(x_k) \| .$$

Therefore $\Delta_{k+1} \leq \Delta_k - \frac{\omega}{r_0^2}\Delta_k^2$. Thus,

$$\frac{1}{\Delta_{k+1}} \geq \frac{1}{\Delta_k} + \frac{\omega}{r_0^2} \cdot \frac{\Delta_k}{\Delta_{k+1}} \geq \frac{1}{\Delta_k} + \frac{\omega}{r_0^2}.$$

Summarizing these inequalities we get

$$\frac{1}{\Delta_{k+1}} \geq \frac{1}{\Delta_0} + \frac{\omega}{r_0^2}(k+1).$$

□

Optimal step size

We need to maximize the function

$$\phi(h) = h(2 - Lh)$$

with respect to h.

$$\phi'(h^*) = 0 \quad \Rightarrow \quad 2 - 2Lh^* = 0.$$

Thus, $h^* = \frac{1}{L}$ and we get:

$$f(x_k) - f^* \leq \frac{2L(f(x_0) - f^*) \parallel x_0 - x^* \parallel^2}{2L \parallel x_0 - x^* \parallel^2 + (f(x_0) - f^*)k}. \quad (2.15)$$

Further, in view of (2.5) we have

$$\begin{aligned} f(x_0) &\leq f^* + \langle f'(x^*), x_0 - x^* \rangle + \tfrac{L}{2} \parallel x_0 - x^* \parallel^2 \\ &= f^* + \tfrac{L}{2} \parallel x_0 - x^* \parallel^2 . \end{aligned}$$

Since the right hand side of (2.15) is increasing in $f(x_0) - f^*$, we get the following

Corollary 2.2 *If* $h = \frac{1}{L}$ *and* $f \in \mathcal{F}_L^{1,1}(R^n)$ *then*

$$f(x_k) - f^* \leq \frac{2L \parallel x_0 - x^* \parallel^2}{k + 4}. \quad (2.16)$$

Theorem 2.14 *If* $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$ *and*

$$0 < h \le \frac{2}{\mu + L}$$

then

$$\| x_k - x^* \|^2 \le \left(1 - \frac{2h\mu L}{\mu + L}\right)^k \| x_0 - x^* \|^2 .$$

If $h = \frac{2}{\mu+L}$ *then*

$$\| x_k - x^* \| \le \left(\frac{Q-1}{Q+1}\right)^k \| x_0 - x^* \|,$$

$$f(x_k) - f^* \le \frac{L}{2}\left(\frac{Q-1}{Q+1}\right)^{2k} \| x_0 - x^* \|^2,$$

where $Q = L/\mu$.

Proof. Denote $r_k = \| x_k - x^* \|$. Then

$$\begin{aligned} r_{k+1}^2 &= \| x_k - x^* - hf'(x_k) \|^2 \\ &= r_k^2 - 2h\langle f'(x_k), x_k - x^* \rangle + h^2 \| f'(x_k) \|^2 \\ &\le \left(1 - \frac{2h\mu L}{\mu+L}\right) r_k^2 + h\left(h - \frac{2}{\mu+L}\right) \| f'(x_k) \|^2 \end{aligned}$$

(we use (2.14) and $f'(x^*) = 0$).

The second inequality follows from the previous one and (2.5). □

Conclusion:

1. The gradient method is not optimal for $\mathcal{F}_L^{1,1}(R^n)$.

1. The gradient method is not optimal for $\mathcal{S}_{\mu,L}^{1,1}(R^n)$.