Chapter II. Convex Differentiable Optimization

Lecture 4: Minimization of Smooth Convex Functions

Yurii Nesterov
EPLouvain (UCL)

Outline

- Smooth convex functions
- Lower complexity bounds
- Strongly convex functions
- Lower complexity bounds
- Gradient method

Smooth convex functions

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

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 $\quad f'(\bar{x}) = 0 \quad \Rightarrow \quad f(x) \geq f(\bar{x}) \quad \forall x \in \mathbb{R}^n$.
2. If $f_1, f_2 \in \mathcal{F}$ and $\alpha, \beta \geq 0$ then $\quad \alpha f_1 + \beta f_2 \in \mathcal{F}$.
3. $\alpha + \langle a, x \rangle \in \mathcal{F}$.

What it could be?

Main definition

Let $f \in \mathcal{F}$. Let us fix some $x_0 \in \mathbb{R}^n$ 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 $y \in \mathbb{R}^n$ we have:

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

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

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

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle. \tag{1}$$

Notation: $f \in \mathcal{F}_L^{k,l}(Q)$. The meaning of the indices is the same as for the class $\mathbb{C}_L^{k,l}(Q)$ (see Lecture 2).

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

Properties of convex functions

Theorem 1: If $f \in \mathcal{F}^1(\mathbb{R}^n)$ and $f'(x^*) = 0$ then $f(x) \geq f(x^*) \quad \forall x \in \mathbb{R}^n$.

Proof: In view of the definition for any $x \in \mathbb{R}^n$ we have

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

$\square$

Lemma 1: If $f_1, f_2 \in \mathcal{F}^1(\mathbb{R}^n)$ and $\alpha, \beta \geq 0$ then $f = \alpha f_1 + \beta f_2 \in \mathcal{F}^1(\mathbb{R}^n)$.

Proof: For any $x, y \in \mathbb{R}^n$ 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. $\square$

Affine substitutions

Lemma 2: If $f \in \mathcal{F}^1(\mathbb{R}^m)$, $b \in \mathbb{R}^m$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ then

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

Proof: Indeed, let $x, y \in \mathbb{R}^n$. Denote $\bar{x} = Ax + b$, $\bar{y} = Ay + b$. Since $\phi'(x) = A^T f'(Ax + b)$, we have:

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

$\square$

Equivalent Definitions, I

Theorem 2: Function $f \in \mathcal{F}^1(\mathbb{R}^n)$ if and only if it is continuously differentiable and for any $x, y \in \mathbb{R}^n$ and $\alpha \in [0,1]$ we have:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y). \qquad (2)$$

Proof: Denote $x_\alpha = \alpha x + (1-\alpha)y$. Let $f \in \mathcal{F}^1(\mathbb{R}^n)$. Then

$$f(x_\alpha) \leq f(y) - \langle f'(x_\alpha), y - x_\alpha \rangle = f(y) - \alpha \langle f'(x_\alpha), y - x \rangle,$$

$$f(x_\alpha) \leq f(x) - \langle f'(x_\alpha), x - x_\alpha \rangle = f(x) + (1-\alpha) \langle f'(x_\alpha), y - x \rangle.$$

Multiplying the first inequality by $(1-\alpha)$, the second one by α, and adding the results, we get (2).

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

Equivalent Definitions, II

Theorem 3: Function $f \in \mathcal{F}^1(\mathbb{R}^n)$ if and only if it is continuously differentiable and for any $x, y \in \mathbb{R}^n$ we have:

$$\langle f'(x) - f'(y), x - y \rangle \geq 0. \tag{3}$$

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

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

Adding these inequalities, we get (3).

2. Let (3) holds for all $x, y \in \mathbb{R}^n$. Denote $x_\tau = x + \tau(y - x)$. Then

$$\begin{aligned}
f(y) &= f(x) &&+ \int\limits_0^1 \langle f'(x + \tau(y-x)), y - x \rangle d\tau \\
&= f(x) &&+ \langle f'(x), y - x \rangle + \int\limits_0^1 \langle f'(x_\tau) - f'(x), y - x \rangle d\tau \\
&= f(x) &&+ \langle f'(x), y - x \rangle + \int\limits_0^1 \tfrac{1}{\tau} \langle f'(x_\tau) - f'(x), x_\tau - x \rangle d\tau \\
&\geq f(x) &&+ \langle f'(x), y - x \rangle.
\end{aligned}$$

$\square$

Equivalent Definitions, III

Theorem 3: Function $f \in \mathcal{F}^2(\mathbb{R}^n)$ if and only if it is twice continuously differentiable and for any $x \in \mathbb{R}^n$ we have:

$$f''(x) \succeq 0. \tag{4}$$

Proof: 1. Let $f \in \mathbb{C}^2(\mathbb{R}^n)$ be convex. Denote $x_\tau = x + \tau s$, $\tau > 0$. Then, in view of (3), we have:

$$\begin{aligned} 0 &\leq \tfrac{1}{\tau^2}\langle f'(x_\tau) - f'(x), x_\tau - x\rangle = \tfrac{1}{\tau}\langle f'(x_\tau) - f'(x), s\rangle \\ &= \tfrac{1}{\tau}\int\limits_0^\tau \langle f''(x+\lambda s)s, s\rangle d\lambda, \end{aligned}$$

and we get (4) by tending $\tau \to 0$.

2. Let (4) holds for all $x \in \mathbb{R}^n$. Then

$$\begin{aligned} f(y) &= f(x) + \langle f'(x), y - x\rangle \\ &\quad + \int\limits_0^1 \int\limits_0^\tau \langle f''(x + \lambda(y-x))(y-x), y-x\rangle d\lambda d\tau \\ &\geq f(x) + \langle f'(x), y - x\rangle. \qquad \square \end{aligned}$$

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

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 $\mathcal{F}^1(\mathbb{R})$:

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

and many others.

Geometric Programming: $f(x) = \sum\limits_{i=1}^{m} e^{\alpha_i + \langle a_i, x \rangle}$.

L_p**-approximation:** $f(x) = \sum\limits_{i=1}^{m} |\, \langle a_i, x \rangle - b_i \,|^p$.

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

Theorem 4: For inclusion $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$ all the conditions below are equivalent:

$$0 \leq f(y) - f(x) - \langle f'(x), y - x \rangle \leq \frac{L}{2} \| x - y \|^2 . \tag{5}$$

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle + \frac{1}{2L} \| f'(x) - f'(y) \|^2 . \tag{6}$$

$$\langle f'(x) - f'(y), x - y \rangle \geq \frac{1}{L} \| f'(x) - f'(y) \|^2 . \tag{7}$$

Proof: Indeed, (5) follows from Lemma 2.3. Further, let us fix $x_0 \in \mathbb{R}^n$. Consider the function $\phi(y) = f(y) - \langle f'(x_0), y \rangle$.

Note that $\phi \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$ and $y^* = x_0$. Therefore, in view of (5), we have:

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

And we get (6) since $\phi'(y) = f'(y) - f'(x_0)$.

We obtain (7) from (6) by adding two inequalities with x and y interchanged.

Finally, from (7) we get $\| f'(x) - f'(y) \| \leq L \| x - y \|$. $\square$

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

Problem formulation: $\min_{x\in\mathbb{R}^n} f(x)$.

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

Oracle: First-order black box.

Approximate solution: Find $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) - f^* \leq \epsilon$.

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

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

Worst function in the world, I

Consider the family of functions

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

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

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

Note that for any $h \in \mathbb{R}^n$ we have:

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

and

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

Thus, $0 \preceq f_k''(x) \preceq L\, I_n$. Therefore $f_k(x) \in \mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$.

Structure of the solution

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

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

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

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

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

Note that

$$\begin{aligned} \| \bar{x}_k \|^2 &= \sum_{i=1}^{n} (\bar{x}_k^{(i)})^2 = \sum_{i=1}^{k} \left(1 - \frac{i}{k+1}\right)^2 = k - \frac{2}{k+1} \sum_{i=1}^{k} i + \frac{1}{(k+1)^2} \sum_{i=1}^{k} i^2 \\ &= k - \frac{2}{k+1} \cdot \frac{k(k+1)}{2} + \frac{1}{(k+1)^2} \cdot \frac{k(k+1)(2k+1)}{6} = \frac{k(2k+1)}{6(k+1)} \le \frac{k+1}{3}. \end{aligned}$$

Behavior of the minimization sequence

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

Lemma 3: Let $x_0 = 0$. Then for any sequence $\{x_k\}_{k=0}^{p}$:

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

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

Proof: 1. Since $x_0 = 0$, we have $f_p'(x_0) = -\frac{L}{4} e_1 \in \mathbb{R}^{1,n}$. Therefore $\mathcal{L}_1 \equiv \mathbb{R}^{1,n}$.

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

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

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

Corollary 1: For any sequence $\{x_k\}_{k=0}^{p}$ such that $x_0 = 0$ and $x_k \in \mathcal{L}_k$, we have $f_p(x_k) \ge f_k^*$.

Proof: Indeed, $x_k \in \mathbb{R}^{k,n}$ and therefore $f_p(x_k) = f_k(x_k) \ge f_k^*$. $\square$

Lower complexity bounds

Theorem 5: 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 $\mathcal{M}$ generating a sequence

$$x_k \in x_0 + \mathrm{Lin}\{f'(x_0), \dots, f'(x_{k-1})\}, \text{ we have}$$

$$f(x_k) - f^* \ge \tfrac{3L\|x_0 - x^*\|^2}{32(k+1)^2}, \quad \| x_k - x^* \|^2 \ge \tfrac{1}{8} \| x_0 - x^* \|^2.$$

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

Then $x^* = \bar{x}_{2k+1}$, $f^* = f^*_{2k+1}$, and $f(x_k) = f_{2k+1}(x_k) = f_k(x_k) \ge f_k^*$.

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

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

$$\begin{aligned}
\| x_k - x^* \|^2 &\ge \sum_{i=k+1}^{2k+1} (\bar{x}^{(i)}_{2k+1})^2 = \sum_{i=k+1}^{2k+1} \left(1 - \tfrac{i}{2k+2}\right)^2 \\
&= k + 1 - \tfrac{1}{k+1} \sum_{i=k+1}^{2k+1} i + \tfrac{1}{4(k+1)^2} \sum_{i=k+1}^{2k+1} i^2 \\
&= k + 1 - \tfrac{1}{k+1} \cdot \tfrac{(k+1)(3k+2)}{2} + \tfrac{14k^2 + 19k + 6}{24(k+1)} \\
&= \tfrac{2k^2 + 7k + 6}{24(k+1)} \ge \tfrac{2k+2}{24} \ge \tfrac{1}{8} \| x_0 - \bar{x}_{2k+1} \|^2 . \qquad \square
\end{aligned}$$

Remarks

NB: Using a more sophisticated analysis, it is possible to prove that

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

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

Conclusion:

1. The lower bound for objective function is quite optimistic: one hundred iterations can 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: $\min\limits_{x \in \mathbb{R}^n} f(x), \quad f \in \mathcal{F}^1(\mathbb{R}^n).$

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

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

By the same arguments as in the convex case, we come to

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

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle + \tfrac{1}{2}\mu \parallel y - x \parallel^2 . \tag{8}$$

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 $\mathbb{C}^{k,l}_L(Q)$ (see Lecture 2).

Properties of strongly convex functions

Theorem 6: If $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$ and $f'(x^*) = 0$, then

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

Proof: Since $f'(x^*) = 0$, in view of Definition, for any $x \in \mathbb{R}^n$ we have

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

Lemma 4: If $f_1 \in \mathcal{S}^1_{\mu_1}(\mathbb{R}^n)$, $f_2 \in \mathcal{S}^1_{\mu_2}(\mathbb{R}^n)$ and α, $\beta \geq 0$, then

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

Proof: For any $x, y \in \mathbb{R}^n$ 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 add these inequalities multiplied by α and β. $\square$.

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

Equivalent Definitions

Theorem 7: Function $f \in \mathcal{S}_{\mu}^{1}(\mathbb{R}^n)$ if and only if it is continuously differentiable and for any $x, y \in \mathbb{R}^n$ and $\alpha \in [0,1]$ we have:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \tfrac{\alpha(1-\alpha)}{2}\mu \parallel x - y \parallel^2 \qquad (9)$$

Theorem 8: Function $f \in \mathcal{S}_{\mu}^{1}(\mathbb{R}^n)$ if and only if it is continuously differentiable and for any $x, y \in \mathbb{R}^n$ we have:

$$\langle f'(x) - f'(y), x - y \rangle \geq \mu \parallel x - y \parallel^2 . \qquad (10)$$

Theorem 9: Function $f \in \mathcal{S}_{\mu}^{2}(\mathbb{R}^n)$ if and only if it is twice continuously differentiable and for any $x \in \mathbb{R}^n$ we have:

$$f''(x) \succeq \mu I_n. \qquad (11)$$

The proofs of these theorems are very similar to proofs of Theorems 2–4.

Examples

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

2. Let the symmetric matrix A satisfy the condition: $\mu I_n \preceq A \preceq L\, I_n$. Then

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

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

Other examples can be obtained by adding a convex function to a strongly convex function.

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

Conditions:
$$\mu \| x-y \|^2 \leq \langle f'(x)-f'(y), x-y\rangle, \qquad (12)$$
$$\| f'(x)-f'(y) \| \leq L \| x-y \| . \qquad (13)$$

The value $Q_f = L/\mu$ (≥ 1) is called *condition number* of function f.
Theorem 10: If $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, then for any x, $y \in \mathbb{R}^n$ we have:

$$\langle f'(x)-f'(y), x-y\rangle \geq \tfrac{\mu L \|x-y\|^2}{\mu+L} + \tfrac{1}{\mu+L} \| f'(x)-f'(y) \|^2 \qquad (14)$$

Proof: Define $\phi(x) = f(x) - \frac{1}{2}\mu \| x \|^2$. Note that $\phi'(x) = f'(x) - \mu x$. Therefore this function is convex (Theorem 3). Moreover, in view of (5)

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

Therefore $\phi \in \mathcal{F}^{1,1}_{L-\mu}(\mathbb{R}^n)$ (see Theorem 5). Thus,

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

and that is exactly (14). $\square$

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

Problem formulation: $\min_{x \in \mathbb{R}^n} f(x)$.

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

Oracle: First-order black box.

Approximate solution: Find $\bar{x} \in \mathbb{R}^n$ 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 consider the case $n = \infty$.

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

Worst function in the world, II

Let $\mathbb{R}^\infty \equiv \ell_2$, 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

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

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 $0 \preceq A \preceq 4I$.

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

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

Structure of the solution

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

Coordinate form of this equation is as follows:

$$\begin{aligned} 2\tfrac{Q+1}{Q-1}x^{(1)} - x^{(2)} &= 1, \\ x^{(k+1)} - 2\tfrac{Q+1}{Q-1}x^{(k)} + x^{(k-1)} &= 0, \quad k = 2, \dots . \end{aligned}$$

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 $(x^*)^{(k)} = q^k, \quad k = 1, 2, \dots,$ satisfies our system.

Thus, we come to the following result.

Lower Complexity Bound

Theorem 11: For any $x_0 \in \mathbb{R}^\infty$ and any constants $\mu > 0$ and $Q > 1$ there exists a function $f \in \mathcal{S}^{\infty,1}_{\mu,\mu Q}(\mathbb{R}^\infty)$ such that for any first order method $\mathcal{M}$ generating a sequence

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

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

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

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 matrix, and $f'(0) = e_1$, we have $x_k \in \mathbb{R}^{k,\infty}$. Therefore

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

The second estimate follows from the first one and Definition of strongly convex functions. $\square$

Gradient Method

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

Scheme:

- Choose $x_0 \in \mathbb{R}^n$.
- **kth iteration ($k \geq 0$).**

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

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

  (see Lecture 2 for the step-size rules).

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

$$h_k \equiv h > 0.$$

Convergence

Theorem 12: If $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$ 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 (7) and $f'(x^*) = 0$). Therefore $r_k \leq r_0$. In view of (5),

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

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

Summing up 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:

$$f(x_k) - f^* \quad \leq \quad \frac{2L(f(x_0) - f^*)\|x_0 - x^*\|^2}{2L\|x_0 - x^*\|^2 + (f(x_0) - f^*)k}. \tag{15}$$

Further, in view of (5) we have

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

Since the right hand side of (15) is increasing in $f(x_0) - f^*$, we get the following

Corollary 2: If $h = \frac{1}{L}$ and $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$. then

$$f(x_k) - f^* \quad \leq \quad \frac{2L\|x_0 - x^*\|^2}{k+4}. \tag{16}$$

Strongly Convex Case

Theorem 13: If $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ and $0 < h \leq \frac{2}{\mu+L}$, then

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

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

$$\begin{aligned} \| x_k - x^* \| \quad &\leq \quad \left(\frac{Q-1}{Q+1}\right)^k \| x_0 - x^* \|, \\ f(x_k) - f^* \quad &\leq \quad \frac{L}{2}\left(\frac{Q-1}{Q+1}\right)^{2k} \| x_0 - x^* \|^2, \end{aligned}$$

where $Q = L/\mu$.

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

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

(we use (14) and $f'(x^*) = 0$).

The second inequality follows from the previous one and (5). $\square$

Conclusion

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

Note:

All standard NLP methods (conjugate gradients, variable metric, etc.) have the similar *lower* efficiency estimates.