Lecture 10

10.1 General Convex Functions

In this part of the course, we are moving on to study general problems with convex components which are not necessarily smooth. The geometry of such problems is directly linked to the notion of *convex sets*, which we are going to review first.

10.1.1 Convex Sets

We say that a set $Q \subseteq \mathbb{R}^n$ is *convex*, if for any two points $x, y \in Q$, the whole segment between these points belong to the set:

$$\lambda x + (1 - \lambda) y \quad \in \quad Q.$$

Basic Properties.

1. *Intersection.* Let $Q_1, Q_2 \subseteq \mathbb{R}^n$ are convex. Then, $Q_1 \cap Q_2$ is also convex. More generally, let $\{Q_\alpha \subseteq \mathbb{R}^n\}$ be *any family* of convex sets indexed by some α. Then their intersection

$$Q \quad = \quad \bigcap_\alpha Q_\alpha$$

is convex. Indeed, let $x, y \in Q$ and $0 \leq \lambda \leq 1$. Consider arbitrary index α. Then, $x, y \in Q_\alpha$ and due to convexity of Q_α, we have $x_\lambda := \lambda x + (1-\lambda) y \in Q_\alpha$. Therefore, $x_\lambda \in Q$.

2. *The convex hull* of a set $X \subseteq \mathbb{R}^n$ is the smallest convex set that contains X:

$$\mathrm{conv}(X) \quad = \quad \bigcap_\alpha Q_\alpha, \qquad \text{for all convex } \; Q_\alpha \subseteq \mathbb{R}^n \;\text{ s.t. }\; X \subseteq Q_\alpha.$$

3. *Scaling* of a convex set by a positive scalar, $a > 0$,

$$aQ \quad = \quad \{ax \; : \; x \in Q\}$$

is convex for convex Q. This is a particular case of the following construction.

4. *Affine image* of a convex set is a convex set. Let $\mathcal{A}(x) := Ax + b$, for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then,

$$\mathcal{A}(Q) \quad = \quad \Big\{ \mathcal{A}(x) \; : \; x \in Q \Big\} \quad \subseteq \quad \mathbb{R}^m$$

is convex. Indeed, let $x, y \in \mathcal{A}(Q)$ and $0 \leq \lambda \leq 1$. Thus, for some $\bar{x}, \bar{y} \in Q$ we have $x = \mathcal{A}(\bar{x})$ and $y = \mathcal{A}(\bar{y})$. By convexity of Q we have that $\bar{x}_\lambda := \lambda \bar{x} + (1-\lambda)\bar{y} \in Q$, and therefore, since affine mapping preserves convex combinations, we have

$$\begin{aligned} x_\lambda \quad &:= \quad \lambda x + (1-\lambda) y \quad = \quad \lambda \mathcal{A}(\bar{x}) + (1-\lambda)\mathcal{A}(\bar{y}) \\ &= \quad \mathcal{A}(\lambda \bar{x} + (1-\lambda)\bar{y}) \quad = \quad \mathcal{A}(\bar{x}_\lambda). \end{aligned}$$

Hence, $x_\lambda \in \mathcal{A}(Q)$.

Examples of Convex Sets.

1. *Hyperplane*: $Q = \{x \in \mathbb{R}^n \;:\; \langle a, x\rangle = b\}$ and *half-space*: $Q = \{x \in \mathbb{R}^n \;:\; \langle a, x\rangle \leq b\}$, for $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, are convex sets.

2. *Affine subspace*: $Q = \bigl\{x \in \mathbb{R}^n \;:\; Ax = b\bigr\}$, for any $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. It is convex as intersection of the hyperplanes.

3. *Polyhedron*: $Q = \{x \in \mathbb{R}^n \;:\; Ax \leq b\}$ is convex as it is a finite intersection of half-spaces. This set is fundamental to linear programming. At the same time, a central fact of convex analysis is that *any* closed convex set can be represented as the intersection (possibly infinite) of half-spaces.

4. *Ball in any norm*:
$$Q \;=\; \Bigl\{x \in \mathbb{R}^n \;:\; \|x\| \leq R\Bigr\}.$$
Indeed, let $x, y \in Q$. Then, for $x_\lambda = \lambda x + (1-\lambda)y$ with $0 \leq \lambda \leq 1$ we have
$$\|x_\lambda\| \;\leq\; \lambda\|x\| + (1-\lambda)\|y\| \;\leq\; \lambda R + (1-\lambda)R \;=\; R.$$

5. Ellipsoid:
$$Q \;=\; \Bigl\{\; x \in \mathbb{R}^n \;:\; \langle H(x-x_0), x - x_0\rangle \leq 1 \;\Bigr\}.$$
for some $H = H^\top \succ 0$.
   - It is an image of the unit Euclidean ball under affine transformation:
   $$Q \;=\; \Bigl\{Bu + x_0 \;\;:\;\; u \in \mathbb{R}^n \text{ s.t. } \langle u, u\rangle \leq 1\Bigr\},$$
   where $B := H^{-1/2}$.

6. *Cone of positive definite matrices*: $\mathbb{S}^n_+ = \Bigl\{X \in \mathbb{S}^n \;:\; X \succeq 0\Bigr\}$.
Indeed, let $X, Y \in \mathbb{S}^n_+$. Then $\lambda X \in \mathbb{S}^n_+$ (for any $\lambda \geq 0$) and $X + Y \in \mathbb{S}^n_+$ by the basic properties of eigenvalues. Hence, $\lambda x + (1-\lambda)y \in \mathbb{S}^n_+$ as well, for $0 \leq \lambda \leq 1$. So $\mathbb{S}^n_+$ is a *convex cone.*

7. *Semidefinite programming.* Q is an intersection of $\mathbb{S}^n_+$ cone with affine hyperplanes:
$$Q \;=\; \Bigl\{X \in \mathbb{S}^n_+ \;:\; \langle A_1, X\rangle = b_1, \ldots, \langle A_m, X\rangle = b_m\Bigr\}.$$
If we additionally restrict matrix X to be diagonal, we get the feasible set in *linear programming*:
$$Q \;=\; \Bigl\{x \in \mathbb{R}^n_+ \;:\; \langle a_1, x\rangle = b_1, \ldots, \langle a_m, x\rangle = b_m\}.$$

Epigraph of Convex Function. We recall that a function $f : \mathrm{dom}\, f \to \mathbb{R}$, where $\mathrm{dom}\, f \subseteq \mathbb{R}^n$, is convex, if
$$f(\lambda x + (1-\lambda)y) \;\leq\; \lambda f(x) + (1-\lambda)f(y), \qquad x, y \in \mathrm{dom}\, f, \;\; 0 \leq \lambda \leq 1. \tag{10.1}$$
This definition implies that the domain of a convex function must be a *convex set*, otherwise, the left hand side in (10.1) is not defined. Note that this general definition works even if the function is non-differentiable.

Now, we can look at the following set, called *epigraph*:
$$\mathrm{epi}\, f \;=\; \Bigl\{(x, t) \in Q \times \mathbb{R} \;:\; f(x) \leq t\Bigr\}. \tag{10.2}$$

Proposition 10.1.1. *Function f is convex* $\Leftrightarrow$ $\operatorname{epi} f$ *is a convex set.*

Exercise 10.1.1. Prove Proposition 10.1.1.

Therefore, we can associate any convex function with a convex set, its epigraph. And this is a systematic way to generate convex sets: we can take any convex function we already know, and we obtain a non-trivial convex set (10.2).

In non-smooth convex optimization, we rather view functions through the lens of convex sets and their underlying geometry.

10.1.2 Separation Theorem

The most important geometrical facts about convex sets are the so-called *separation theorems*. There are many variations of these. For our purposes, it is convenient to use the following version, which we state without the proof (see, e.g. [Hör94]).

Theorem 10.1.2. *Let $Q \subseteq \mathbb{R}^n$ be convex and assume that its interior is non-empty:* $\operatorname{int} Q \neq \emptyset$. *Let $x \in \mathbb{R}^n$ do not belong to the interior of Q: $x \notin \operatorname{int} Q$. Then, x can be separated from Q by a linear function, i.e. there exists $\ell \in \mathbb{R}^n$:*

$$\langle \ell, x \rangle \quad \geq \quad \langle \ell, y \rangle, \qquad y \in Q. \tag{10.3}$$

When separation occurs at a boundary point $x \in \partial Q$, we call such hyperplane *supporting* to the set Q. Supporting hyperplanes provide us with the main search directions for algorithms in convex optimization. Another consequence of Theorem 10.1.2 is the following statement.

Corollary 10.1.3. *Any closed (open) convex set $Q \subseteq \mathbb{R}^n$ is equal to the intersection of closed (open) hyperplanes containing it.*

10.1.3 Subgradients

Now, we can look at a convex function through the lens of their epigraphs. Let $f : Q \to \mathbb{R}$ be convex and consider $x \in Q \subset \mathbb{R}^n$. Then, point $(x, f(x))$ belongs to the boundary of the epigraph: $(x, f(x)) \in \partial \operatorname{epi} f$. Hence, there exists a supporting hyperplane such that it separates epigraph. Such hyperplane is called the subgradient.

Definition 10.1.1. We say that a vector $g \in \mathbb{R}^n$ is a subgradient of f at point x if

$$f(y) \quad \geq \quad f(x) + \langle g, y - x \rangle, \qquad y \in \operatorname{dom} f.$$

The set of all subgradients is denoted by $\partial f(x)$ and is called subdifferential of f. We denote by $f'(x) \in \partial f(x)$ any particular selection of a subgradient.

Note that by this definition we might have several subgradients at the same point, which happens when the function is non-differentiable. It might also be the case that there are not subgradients at all: $\partial f(x) = \varnothing$. However, it appears that such unfortunate situations might only happen at the boundary of our domain.

In this course, we will always assume that $f : Q \to \mathbb{R}$ where $Q \subseteq \mathbb{R}^n$ is *open set*. In such situations a subgradient always exists for any $x \in Q$.

Theorem 10.1.4. *Let $f : Q \to \mathbb{R}$ be a convex function defined on an open set $Q \subseteq \mathbb{R}^n$. Then, for any $x \in Q$ we have $\partial f(x) \neq \varnothing$.*

Proof. Consider the point $y = (x, f(x)) \in \partial \operatorname{epi} f \subseteq \mathbb{R}^{n+1}$ from the boundary of the epigraph. By the separation theorem, there exists a non-zero vector $[\ell_0, \ell]^\top \in \mathbb{R}^{n+1}$ where $\ell_0 \in \mathbb{R}$ and $\ell \in \mathbb{R}^n$, such that

$$\ell_0(f(x) - t) + \langle \ell, x - y \rangle \quad \overset{(10.3)}{\geq} \quad 0, \qquad \forall y \in Q \text{ and } \forall t \geq f(y). \tag{10.4}$$

Substituting $y := x$ and $t > f(x)$ we have $\ell_0(f(x) - t) \geq 0$. Therefore, we conclude that $\ell_0 \leq 0$.

Let us prove that $\ell_0 < 0$ (strictly). Assume that $\ell_0 = 0$ and take $y := x + \varepsilon \frac{\ell}{\|\ell\|}$ for a sufficiently small $\varepsilon > 0$ so that $y \in Q$. We have

$$\langle \ell, x - y \rangle \quad = \quad -\varepsilon \|\ell\| \quad < \quad 0,$$

which contradicts (10.4). Therefore, $\ell < 0$. Dividing inequality (10.4) by it and rearranging the terms, we get, for $t := f(y)$:

$$f(y) \quad \geq \quad f(x) + \langle \tfrac{\ell}{\ell_0}, y - x \rangle.$$

Thus, $\frac{\ell}{\ell_0} \in \partial f(x)$. □

Properties of Subdifferentials.

- *Sum of two convex functions.* Let $f(x) = \alpha f_1(x) + \beta f_2(x)$. Then $\partial f(x) = \alpha \partial f_1(x) + \beta \partial f_2(x)$.
- *Pointwise maximum* of any family of convex functions:

  $$f(x) \quad = \quad \max_{\alpha} f_\alpha(x),$$

  is convex as its epigraph is the intersection of convex sets:

  $$\operatorname{epi} f(x) \quad = \quad \bigcap_{\alpha} \operatorname{epi} f_\alpha.$$

  In general, we have:

  $$\partial f(x) \quad \supseteq \quad \operatorname{conv}\Bigl\{ \partial f_\alpha(x) \;:\; \alpha \text{ s.t. } f(x) = f_\alpha(x) \Bigr\}. \tag{10.5}$$

  Indeed, let us fix $x \in Q$ and an α s.t. the maximum is achieved: $f(x) = f_\alpha(x)$. Then, for any $y \in Q$ it holds:

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

  Hence, $f'_\alpha(x) \in \partial f(x)$. The exact equation in (10.5) holds, e.g. when the family $\{f_\alpha\}$ is *finite*.
- *Differentiable function.* Let $f : Q \to \mathbb{R}$ be differentiable and convex. Then,

  $$f(x + h) - f(x) \quad \geq \quad \langle g, h \rangle \quad \text{for} \quad g \in \partial f(x)$$

  and

  $$f(x + h) - f(x) \quad = \quad \langle \nabla f(x), h \rangle + o(\|h\|).$$

  Subtracting this equation from the inequality above, we get

  $$0 \quad \geq \quad \langle g - \nabla f(x), h \rangle + o(\|h\|), \qquad h \in \mathbb{R}^n.$$

  We conclude that $g = \nabla f(x)$. So $\partial f(x) = \{\nabla f(x)\}$.

Examples.

1. Let $f(x) = \|x\|_2 = \sqrt{\langle x, x \rangle}$. This function is differentiable everywhere except 0. Therefore, we have
$$\partial \| \cdot \|_2(x) \quad = \quad \{\nabla f(x)\} \quad = \quad \Big\{ \tfrac{1}{\|x\|_2} x \Big\}, \qquad x \neq 0. \tag{10.6}$$
Computing the subdifferential at 0 means to find all vectors $s \in \mathbb{R}^n$ (subgradients) such that
$$\|x\|_2 \quad \geq \quad \langle s, x \rangle, \qquad x \in \mathbb{R}^n. \tag{10.7}$$
By Cauchy-Schwarz inequality, we know that any $s \in \mathbb{R}^n$ such that $\|s\|_2 \leq 1$ satisfies (10.7). At the same time, plugging in $x := s$ into (10.7) we ensure that this is also a necessary condition for s to be a subgradient. Hence, we justified that
$$\partial \| \cdot \|_2(0) \quad = \quad \Big\{ s \in \mathbb{R}^n \; : \; \|s\|_2 \leq 1 \Big\}. \tag{10.8}$$

2. The formula (10.8) works for an arbitrary norm $\| \cdot \|$:
$$\partial \| \cdot \|(0) \quad = \quad \Big\{ s \in \mathbb{R}^n \; : \; \|s\|_* \leq 1 \Big\}, \tag{10.9}$$
where $\| \cdot \|_*$ is the dual norm. However, formula (10.6) is no longer true, as $f(x) = \|x\|$ is, in general, might nor be differentiable (e.g. consider $\| \cdot \|_1$ or $\| \cdot \|_\infty$ norm).

3. Let $f(x) = \max\limits_{1 \leq i \leq m} \Big[\langle a_i, x \rangle - b_i \Big] = \max\limits_{1 \leq i \leq m} f_i(x)$, where $f_i(x) = \langle a_i, x \rangle - b_i$ is affine. We have $\nabla f_i(x) = a_i$ and
$$\partial f(x) \quad = \quad \mathrm{conv} \Big\{ a_i \; : \; 1 \leq i \leq m \;\; \text{s.t.} \;\; f(x) = \langle a_i, x \rangle - b_i \Big\}.$$

4. Let $f(X) = \lambda_{\max}(X)$, for $X \in \mathbb{S}^n$. It is convex as an (infinite) maximum of linear functions:
$$f(X) \quad = \quad \max_{u \in \mathbb{R}^n \,:\, \|u\| = 1} \langle Xu, u \rangle \quad = \quad \max_{u \in \mathbb{R}^n \,:\, \|u\| = 1} \mathrm{tr}(Xuu^\top)$$
From this representation, we immediately obtain a way to compute its subgradients:
$$\partial f(X) \quad \supseteq \quad \mathrm{conv} \Big\{ uu^\top \; : \; u \in \mathbb{R}^n \text{ s.t. } Xu = \lambda_{\max}(X) u \Big\}.$$

10.1.4 Stationary Condition

Consider the following problem of *additive composite optimization*, that often appears in practice:
$$\min_{x \in Q} \Big[F(x) \quad = \quad f(x) + \psi(x) \Big] \tag{10.10}$$

We can set $Q := \mathrm{dom}\, \psi$ and assume that ψ is a *general convex function* (possibly non-differentiable). At the same time, f is differentiable. We can directly prove the following stationary condition for a minimum of this problem.

Theorem 10.1.5. *Let $x^\star$ be a global minimum of* (10.10). *Then*
$$\langle \nabla f(x^\star), x - x^\star \rangle + \psi(x) \quad \geq \quad \psi(x^\star), \qquad x \in Q. \tag{10.11}$$

Proof. Indeed, if (10.11) holds, then, using convexity of f we have:

$$\begin{aligned} F(x) &= f(x) + \psi(x) \\ &\geq f(x^\star) + \langle \nabla f(x^\star), x - x^\star \rangle + \psi(x) \\ &\overset{(10.11)}{\geq} f(x^\star) + \psi(x^\star) \;=\; F(x^\star). \end{aligned}$$

Hence, $x^\star$ is the global minimum.

Now, assume that $x^\star$ is the global minimum of (10.10) and our goal is to prove (10.11). For a sufficiently small $\alpha > 0$, we have

$$\begin{aligned} \langle \nabla f(x^\star), x - x^\star \rangle + \psi(x) - \psi(x^\star) &= \tfrac{1}{\alpha}\Big[f(x^\star + \alpha(x - x^\star)) - f(x^\star) \Big] + \psi(x) - \psi(x^\star) + o(1) \\ = \tfrac{1}{\alpha}\Big[F(x^\star + \alpha(x - x^\star)) - F(x^\star) \Big] &+ \tfrac{1}{\alpha}\Big[\alpha\psi(x) + (1-\alpha)\psi(x^\star) - \psi(x^\star + \alpha(x - x^\star)) \Big] + o(1) \\ \geq 0. \end{aligned}$$

□

Corollary 10.1.6. *We have proved that*

$$-\nabla f(x^\star) \in \partial \psi(x^\star).$$

In practice, it implies that the rule "set gradient to zero" works as well:

$$F'(x^\star) = \nabla f(x^\star) + \psi'(x^\star) = 0,$$

where $\psi'(x^\star) \in \partial\psi(x^\star)$ is some subgradient.

Corollary 10.1.7. *From the proof we see that if f is a non-convex differentiable function, and $x^\star$ is a local minimum of* (10.10)*, then* (10.11) *holds, as a necessary condition for optimality.*

Corollary 10.1.8. *Let $\psi(x)$ be the indicator of a convex set Q:*

$$\psi(x) = \begin{cases} 0, & x \in Q \\ +\infty, & x \notin Q. \end{cases}$$

Then, our problem is $\min\limits_{x \in Q} f(x)$ and condition (10.11) *implies that:*

$$\langle \nabla f(x^\star), x - x^\star \rangle \geq 0, \qquad x \in Q.$$

Geometric interpretation of this inequality is that the gradient at the optimum, $\nabla f(x^\star)$, separates Q from the sublevel set $\mathcal{F} = \Big\{ x \in \operatorname{dom} f \;:\; f(x) \leq f(x^\star) \Big\}$.

10.2 Binary Search Algorithm

10.2.1 Problem Class

We consider 1-dimensional optimization problem:

$$\min_{a \leq x \leq b} f(x),$$

where $f : [a, b] \to \mathbb{R}$ is a convex continuous function, and our feasible set $Q = [a, b]$ is the segment.

- Oracle: $x \mapsto (f(x), f'(x))$ where $f'(x)$ is some subgradient $f'(x) \in \partial f(x) \subseteq \mathbb{R}$.
- The goal: to find $\bar{x}$ s.t. $f(\bar{x}) - f^{\star} \leq \varepsilon$.

10.2.2 Algorithm

Let us analyze the simplest binary search algorithm, which is familiar to everyone.

We start with the initial segment $\ell_0 = a, r_0 = b$. Set $x_0 = \frac{\ell_0 + r_0}{2}$, the middle point, and compute $f'(x_0)$. By convexity, we have the following inequality,

$$f(y) \quad \geq \quad f(x_0) + f'(x_0)(y - x_0), \qquad y \in [a, b].$$

There are the following three options:

1. $f'(x_0) = 0$. Then x_0 is the desirable global minimum: $x_0 = x^{\star}$. However, in practice it is better to never conduct such exact check due to machine precision errors.

2. $f'(x_0) < 0$ (the function is decreasing at x_0). Thus for any $y \in [a, x_0]$ we have $f'(x_0)(y - x_0) \geq 0$ and hence
$$f(y) \quad \geq \quad f(x_0).$$

3. $f'(x_0) > 0$ (the function is increasing at x_0). Then, for any $y \in [x_0, b]$ we have
$$f(y) \quad \geq \quad f(x_0).$$

   In both cases, we know how to switch to a smaller segment.

Algorithm 10.1: *Binary Search Algorithm.*

```
Initialization: $\ell_0 = a$, $r_0 = b$. Fix $K \geq 1$.
For $k = 0 \ldots K - 1$ iterate:
    1. Set $x_k = \frac{1}{2}(\ell_k + r_k)$
    2. Compute $f'(x_k) \in \partial f(x_k) \subseteq \mathbb{R}$
    3. If $f'(x_k) < 0$ then set $\ell_{k+1} = x_k$ and $r_{k+1} = r_k$ else set $\ell_{k+1} = \ell_k$ and $r_{k+1} = x_k$.
Return a point $\bar{x}_K$ among $\{x_0, \ldots, x_K\}$ with the smallest function value: $f(\bar{x}_K) = \min_{0 \leq i \leq K} f(x_i)$.
```

Note that returning the last point x_K is not a good idea in general, if we are interested in a small functional residual.

10.2.3 Analysis

The analysis of the method is based on the following simple observations, which are immediate to check:

Proposition 10.2.1. *Denote by $G_k := [\ell_k, r_k]$ our localization set. It holds $|G_k| = r_k - \ell_k = \frac{b-a}{2^k}$.*

Proposition 10.2.2. *For any solution $x^\star$, we have $x^\star \in G_k$.*

Proposition 10.2.3. *For any $y \in Q \setminus G_k$, we have $f(y) \geq f(\bar{x}_k)$ (the function value outside the localizer is always greater than the best seen point).*

Thus, from the construction of the binary search we immediately obtain very fast linear rate of decrease of the localizer set $|G_k| \to 0$ (Proposition 10.2.1). However, our initial goal was to establish convergence in terms of the functional residual. For that, we employ the following simple machinery.

We denote by V the *variation of the function* over our initial set $Q = [a, b]$:

$$V \quad = \quad \max_{x \in Q} f(x) - \min_{x \in Q} f(x) \quad = \quad \max_{x \in Q} f(x) - f^\star$$

For some $\gamma \in [0, 1]$ consider the *contraction* of the initial set:

$$Q_\gamma \quad := \quad \gamma Q + (1 - \gamma) x^\star.$$

We have: $|Q_\gamma| = \gamma |Q| = \gamma(b - a)$. Let $1 \geq \gamma > 2^{-k}$. Then, there exists

$$y \quad = \quad \gamma z + (1 - \gamma) x^\star \quad \in \quad Q_\gamma, \;\; z \in Q,$$

such that $y \notin G_k$. Then, we get, by convexity:

$$f(\bar{x}_k) \quad \leq \quad f(y) \quad \leq \quad \gamma f(z) + (1 - \gamma) f^\star \quad = \quad \gamma(f(z) - f^\star) + f^\star \quad \leq \quad \gamma V + f^\star.$$

By taking the limit $\gamma \to 2^{-k}$ we prove the following theorem.

Theorem 10.2.4. *After $K \geq 0$ iterations of the binary search algorithm, it holds:*

$$f(\bar{x}_K) - f^\star \quad \leq \quad \tfrac{V}{2^K}.$$

We see that this is a very fast linear rate with a constant factor. In order to achieve $f(\bar{x}_K) - f^\star \leq \varepsilon$ it is enough to perform

$$K \quad = \quad \log_2 \tfrac{V}{\varepsilon}$$

oracle calls.

- It appears that this complexity is *optimal* for the univariate case (there is no better algorithm than binary search in general for one-dimensional convex minimization). See Section 1 in [Nem95].
- Next lecture: we study a generalization of the binary search to multivariate case, called the *ellipsoid method.*

Literature

[Hör94] Lars Hörmander. *Notions of convexity.* Springer, 1994.

[Nem95] Arkadi Nemirovski. *Information-based complexity of convex programming.* Lecture notes, 1995.