Recall from the last lecture that a convex optimization problem is a problem of the form:

\[
\begin{align*}
\text{min.} & \quad f(x) \\
\text{s.t.} & \quad g_i(x) \leq 0 \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0 \quad i = 1, \ldots, k
\end{align*}
\]

where

- Each \( h_i \) is affine: \( h_i(x) = a_i^T x - b_i \)
- \( f, g_i, \ldots, g_m \) are convex:
  \[
  f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)
  \]
  \[\forall x, y \in \text{dom}(f), \forall \lambda \in [0,1]\]

Similarly for the \( g_i \)'s.

- Today we start off by proving results that explain why we give special attention to convex optimization problems.
  - In a convex problem, every local minimum is automatically a global minimum. (This is true even for the more abstract definition of a convex optimization problem from the last lecture that only required the feasible set to be a convex set.)
  - In the unconstrained case, every stationary point (i.e., zero of the gradient) is automatically a global minimum.
- We will also see new characterizations for convex functions that make the task of checking convexity somewhat easier, though in general checking convexity can be a very difficult problem [AOPT13].
Let’s recall the definition of local and global minima and generalize them to the constrained setting.

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad x \in \Omega \quad (\text{e.g., } \Omega = \{x \mid g_i(x) \leq 0, h_j(x) = 0\})
\end{align*}
\]

**Definition:** A point \( x^* \in \mathbb{R}^n \) is

- **feasible**, if \( x^* \in \Omega \); i.e., \( g_i(x^*) \leq 0, \forall i, h_j(x^*) = 0, \forall j \)

- a **local minimum**, if feasible, and if \( \exists \delta > 0 \) s.t. \( f(x^*) \leq f(x), \forall x \) s.t. \( x \in \Omega \) and \( ||x - x^*|| \leq \delta \)

- a **strict local minimum**, if feasible, and if \( \exists \delta > 0 \) s.t. \( f(x^*) < f(x), \forall x \neq x^* \) s.t. \( x \in \Omega \) and \( ||x - x^*|| \leq \delta \)

- a **global minimum**, if feasible, and if \( f(x^*) \leq f(x), \forall x \in \Omega \)

- a **strict global minimum**, if feasible, and if \( f(x^*) < f(x), \forall x \in \Omega, x \neq x^* \)

Our next few theorems show the nice features of convex problems in terms of inferring global properties from local ones.
**Theorem.** Consider an optimization problem

\[
\begin{align*}
\text{min.} \quad & f(x) \\
\text{s.t.} \quad & x \in \Omega
\end{align*}
\]

where \( f \) is a convex function and \( \Omega \) is a convex set. Then, every local minimum is also a global minimum.

**Proof.**

Let \( x \) be a local minimum. Suppose for the sake of contradiction that \( x \) is not a global minimum.

\[
\Rightarrow \exists \ y \in \Omega, \ s. \ t. \ f(y) < f(x).
\]

But \( x \in \Omega, y \in \Omega, \ \Omega \text{ convex} \Rightarrow \lambda x + (1 - \lambda)y \in \Omega, \forall \lambda \in [0,1] \)

and \( f \text{ convex} \Rightarrow f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \)

\[
< \lambda f(x) + (1 - \lambda)f(x) = f(x), \forall \lambda \in [0,1].
\]

As \( \lambda \to 1, \ (\lambda x + (1 - \lambda)y) \to x \). So there are points arbitrarily close to \( x \) with a better objective value than \( x \). This contradicts local optimality of \( x \). \( \square \)

**Intuition:**

\[
f(y) < f(x) \Rightarrow f(\text{all line}) < f(x)
\]
First and second order characterization of convex functions

**Theorem.** Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable over its domain. Then, the following are equivalent:

(i) $f$ is convex.
(ii) $f(y) \geq f(x) + \nabla f^T(x)(y - x)$, $\forall x, y \in \text{dom}(f)$.
(iii) $\nabla^2 f(x) \succeq 0$, $\forall x \in \text{dom}(f)$ (i.e., the Hessian is psd $\forall x \in \text{dom}(f)$).

**Interpretation:**

(i) The first order Taylor expansion at any point is a global under estimator of the function

(ii) Function has *nonnegative curvature* everywhere:

"It curves up".

○ In one dimension: $f''(x) \geq 0$, $\forall x \in \text{dom}(f)$

We prove $(i) \iff (ii)$. For $(ii) \iff (iii)$, see, e.g., Theorem 22.5 of [CZ13].
Proof: ([Tit13])

(i) ⇒ (ii)  If \( f \) convex, by definition
\[
f (\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x), \quad \forall \lambda \in [0,1], x, y \in \text{dom}(f).
\]
After rewriting, we have
\[
f (x + \lambda (y-x)) \leq f(x) + \lambda (f(y) - f(x))
\]
\[
\Rightarrow f(y) - f(x) \geq \frac{f(x + \lambda (y-x)) - f(x)}{\lambda}, \quad \forall \lambda \in [0,1].
\]
As \( \lambda \to 0 \), we get
\[
f(y) - f(x) \geq \nabla f(x) \cdot (y-x). \quad \tag{1}
\]

(ii) ⇒ (i)  Suppose (1) holds \( \forall x, y \in \text{dom}(f) \).

Take any \( x, y \in \text{dom}(f) \) and let \( z = \lambda x + (1-\lambda)y \).

We have
\[
f(z) \geq f(z) + \nabla f(z) \cdot (z - z) \tag{2}
\]
\[
f(y) \geq f(z) + \nabla f(z) \cdot (y - z). \tag{3}
\]
Multiplying (2) by \( \lambda \), (3) by \( (1-\lambda) \) and adding, we get
\[
\lambda f(x) + (1-\lambda)f(y) \geq f(z) + \nabla f(z) \cdot (\lambda x + (1-\lambda)y - z)
\]
\[
= f(z)
\]
\[
= f (\lambda x + (1-\lambda)y).
\]
\[\square\]
**Corollary.** Consider an unconstrained optimization problem:

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad x \in \mathbb{R}^n,
\end{align*}
\]

where \( f \) is convex and differentiable. Then, any point \( \bar{x} \) that satisfies \( \nabla f(\bar{x}) = 0 \), is a global minimum.

**Proof.** From the first order characterization of convexity we have

\[ f(y) \geq f(x) + \nabla f^T(x)(y - x) \quad \forall x, y \]

In particular,

\[ f(y) \geq f(x) + \nabla f^T(x)(y - x) \quad \forall y \]

Since \( \nabla f(x) = 0 \), we get

\[ f(y) \geq f(\bar{x}) \quad \forall y. \quad \square \]

**Remark 1.** Recall that \( \nabla f(x) = 0 \) is always a necessary condition for local optimality in an unconstrained problem. The theorem says that for convex problems \( \nabla f(x) = 0 \) is not only necessary, but also sufficient for local and global optimality.

**Remark 2.** Recall that in absence of convexity, \( \nabla f(x) = 0 \) is not sufficient even for local optimality (e.g., think of \( f(x) = x^3 \) and \( \bar{x} = 0 \)).

**Remark 3.** Recall that another necessary condition for (unconstrained) local optimality of a point \( x \) was: \( \nabla^2 f(x) \succ 0. \)

- Note that a convex function automatically passes this test.
Quadratic functions revisited

- Let \( f(x) = x^T A x + b x + c \) \((A\) symmetric\)

- When is \( f \) convex?
  - Let’s use the second order test:
    \[
    \nabla^2 f(x) = 2A
    \]
    \[
    \Rightarrow f \text{ convex } \iff A \succeq 0 \quad \text{(independent of } b, c)\]

- Consider the unconstrained optimization problem
  \[
  \min_x x^T A x + b x + c \]
  - \( A \not\succ 0 \) \((f \text{ not convex}) \Rightarrow \text{unbounded below (why?)}\)
  - \( A \succ 0 \Rightarrow \text{convex (in fact, } \Rightarrow f \text{ strictly convex as we see next)}\)
    Unique solution: \( x^* = -\frac{1}{2} A^{-1} b \) (why?)
  - \( A \succeq 0 \Rightarrow f \text{ convex. Optimal value may or may not be bounded, and there could be many optimal solutions.} \)
Least squares, revisited.

Given: 
\[ A \quad m \times n \text{ matrix } \] 
\[ b \quad m \times 1 \text{ vector } \] 

Solve: 
\[ \min_{\mathbf{x}} \| A\mathbf{x} - b \|^2 \] 

(Assume columns of \( A \) are linearly independent)

Let 
\[ f(\mathbf{x}) = \| A\mathbf{x} - b \|^2 = (A\mathbf{x} - b)^T (A\mathbf{x} - b) \]

\[ = \mathbf{x}^T A^T A\mathbf{x} - 2 \mathbf{x}^T A^T b + b^T b. \]

\[ = \nabla f(\mathbf{x}) = 2 A^T A\mathbf{x} - 2 A^T b \]

\[ \nabla f(\mathbf{x}) = 0 \implies A^T A\mathbf{x} = A^T b \]

Called

\[ \text{"Normal Equations"} \]

\[ \mathbf{x} = (A^T A)^{-1} A^T b \]

\( A^T A \) is invertible b/c its null space is just the origin:

\[ A^T A \mathbf{x} = 0 \implies \mathbf{x}^T A^T A \mathbf{x} = 0 \implies (A\mathbf{x})^T (A\mathbf{x}) = 0 \implies \| A\mathbf{x} \|^2 = 0 \implies A\mathbf{x} = 0 \]

\[ \implies \mathbf{x} = 0. \]

Columns of \( A \) linearly independent:

\[ \nabla^2 f(\mathbf{x}) = 2 A^T A \geq 0 \quad (\text{b/c} \quad \mathbf{x}^T A^T A \mathbf{x} = \| A\mathbf{x} \|^2 \geq 0 \text{ and } = 0 \implies \mathbf{x} = 0) \]

\[ \implies \mathbf{x} = (A^T A)^{-1} A^T b \text{ is a strict local minimum}. \]

This is the best conclusion we could make before without knowing that \( f \) is convex. Now that we know \( f \) is (strictly) convex, we immediately know that the solution \( x = (A^T A)^{-1} A^T b \) is a (strict) global minimum.
Characterization of Strict Convexity

Recall that we say $f: \mathbb{R}^n \to \mathbb{R}$ is strictly convex, if $\forall x, y, x \neq y, \forall \lambda \in (0,1)$,

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y).$$

- $f$ strictly convex $\Rightarrow$ $f$ convex (obvious from the definition)
- $f$ convex $\nRightarrow$ $f$ strictly convex
  
  e.g., $f(x) = x$ ($x \in \mathbb{R}$)

- Second order sufficient condition:
  $$\nabla^2 f(x) > 0 \ \forall x \in \Omega \Rightarrow f$ strongly convex on $\Omega$

- Converse not true:
  $$f(x) = x^4 \ (x \in \mathbb{R})$$
  $f$ is strictly convex (why?).
  But $f''(0) = 0$ (check)

- First order characterization:
  $$f$$ strictly convex on $\Omega \subseteq \mathbb{R}^n$
  $\Leftrightarrow$
  $$f(y) > f(x) + \nabla f^T(x)(y - x), \ \forall x, y \in \Omega, x \neq y$$

- One of the main uses of strict convexity is to ensure uniqueness of optimal solutions. We see this next.
Strict Convexity and Uniqueness of Optimal Solutions

**Theorem.** Consider an optimization problem

\[
\begin{aligned}
\text{min. } & f(x) \\
\text{s.t. } & x \in \Omega,
\end{aligned}
\]

where \( f: \mathbb{R}^n \to \mathbb{R} \) is strictly convex on \( \Omega \) and \( \Omega \) is a convex set. Then, the (optimal) solution is unique (assuming it exists).

**Proof.** Suppose there were two optimal solutions \( x, y \in \mathbb{R}^n \). This means that \( x, y \in \Omega \) and

\[
\forall z \in \Omega \quad f(x) = f(y) \leq f(z) \tag{1}
\]

But consider \( z = \frac{x+y}{2} \). By convexity of \( \Omega \), we have \( z \in \Omega \). By strict convexity,

we have

\[
f(z) = f\left(\frac{x+y}{2}\right) < \frac{1}{2} f(x) + \frac{1}{2} f(y) = \frac{1}{2} f(x) + \frac{1}{2} f(x) = f(x).
\]

But this contradicts (1). □

**Practice:** for each function below, determine whether it is convex, strictly convex, or neither.

- \( f(x) = (x_1 - 3x_2)^2 \)
- \( f(x) = (x_1 - 3x_2)^2 + (x_1 - 2x_2)^2 \)
- \( f(x) = (x_1 - 3x_2)^2 + (x_1 - 2x_2)^2 + x_1^3 \)
- \( f(x) = |x| \ (x \in \mathbb{R}) \)
- \( f(x) = ||x|| \ (x \in \mathbb{R}^n) \)
An Optimality Condition for Convex Problems

Theorem. Consider an optimization problem

\[
\min f(x) \\
\text{s.t. } x \in \Omega,
\]

where \( f: \mathbb{R}^n \to \mathbb{R} \) is convex and differentiable and \( \Omega \) is a convex. Then, a point \( x \) is optimal if and only if \( x \in \Omega \) and

\[
\nabla f(x)^T (y - x) \geq 0, \forall y \in \Omega.
\]

- **What does this mean?**
  - If you move from \( x \) towards any feasible \( y \), you will increase \( f \) locally.
  - \( -\nabla f(x) \) (assuming it is nonzero) serves as a hyperplane that "supports" the feasible set \( \Omega \) at \( x \) (see figure below).

- The necessity of the condition holds independent of convexity of \( f \).
- Convexity is used in establishing sufficiency.
- If \( \Omega = \mathbb{R}^n \), can you see how the condition above reduces to our first order unconstrained optimality condition \( \nabla f(x) = 0 \)?
  - Hint: take \( y = x - \nabla f(x) \).
Proof.

(Sufficiency)

Suppose \( x \in \Omega \)

satisfies \( \nabla f(x) \cdot (y - x) > 0, \quad \forall y \in \Omega. \quad (1) \)

By the first order characterization of convexity, we have:

\( f(y) > f(x) + \nabla f(x) \cdot (y - x), \quad \forall y \in \Omega \quad (2) \)

\[ (1) + (2) \Rightarrow f(y) > f(x), \quad \forall y \in \Omega. \]

\[ \Rightarrow x \text{ is optimal.} \]

(necessity)

Suppose \( x \) is optimal, but for some \( y \in \Omega \) we had

\[ \nabla f(x) \cdot (y - x) < 0. \]

Consider \( g(\alpha) := f(x + \alpha (y - x)) \).

Because \( \Omega \) is convex, \( \forall \alpha, \ x + \alpha (y - x) \in \Omega. \)

Observe that \( g'(\alpha) = (y - x)^T \nabla f(x + \alpha (x - y)). \)

\[ \Rightarrow g'(0) = (y - x)^T \nabla f(x) < 0. \]

\[ \Rightarrow \exists \delta > 0 \text{ s.t. } g(\alpha) < g(0) \quad \forall \alpha \in (0, \delta). \]

\[ \Rightarrow f(x + \alpha (y - x)) < f(x) \quad \forall \alpha \in (0, \delta). \]

But this contradicts optimality of \( x. \) \( \Box \)
Further reading for this lecture can include the first few pages of Chapters 2, 3, 4 of [BV04]. Your [CZ13] book defines convex sets in Section 4.3. Convex optimization appears in Chapter 22. The relevant sections are 22.1-22.3.

References:


