When stating a convex optimization problem in standard form we write

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0 \quad \text{for all } i = 1, \ldots, n.
\end{align*}
\]

where \(f_0, f_1, \ldots, f_n\) are convex. Why don’t we use \(\geq\) or \(=\) instead of \(\leq\)?
Review of Convexity

Definition (Convex Set)
A set $S \subseteq \mathbb{R}^d$ is convex if for any $x, y \in S$ and $\theta \in (0, 1)$ we have $(1 - \theta)x + \theta y \in S$.

Definition (Convex Function)
A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if for any $x, y \in \mathbb{R}^d$ and $\theta \in (0, 1)$ we have $f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y)$.
Review of Convexity

Convex Set

Non-convex Set

Convex Function

Non-convex Function
(Sub-)Level Sets of Convex Functions

Definition ((Sub-)Level Sets)

For a function $f : \mathbb{R}^d \to \mathbb{R}$, a level set (or contour line) corresponding to the value $c$ is given by the set of all points $x \in \mathbb{R}^d$ where $f(x) = c$:

$$f^{-1}\{c\} = \{x \in \mathbb{R}^d \mid f(x) = c\}.$$

Analogously, the sublevel set for the value $c$ is the set of all points $x \in \mathbb{R}^d$ where $f(x) \leq c$:

$$f^{-1}(-\infty, c] = \{x \in \mathbb{R}^d \mid f(x) \leq c\}.$$
3D Plot and Contour Plot With Sublevel Set
$f(x) \leq -1$
Theorem

If $f : \mathbb{R}^d \to \mathbb{R}$ is convex then the sublevel sets are convex.
Sublevel Sets of Convex Functions

Theorem

If $f : \mathbb{R}^d \to \mathbb{R}$ is convex then the sublevel sets are convex.

Proof.

Fix a sublevel set $S = \{ x \in \mathbb{R}^d \mid f(x) \leq c \}$ for some fixed $c \in \mathbb{R}$. If $x, y \in S$ and $\theta \in (0, 1)$ then we have

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) \leq (1 - \theta)c + \theta c = c.$$
Plots of Convex Function With Sublevel Set

\[ f(x) \leq -1 \]
Intersection of Convex Sets is Convex
Level Sets and Superlevel Sets Not Convex

\[ f(x) < -1 \]

\[ f(x) = 1 \]

\[ f(x) > -1 \]
Lagrange Duality

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0
\end{align*}
\quad \rightarrow \quad \text{Lagrange Dual} \quad \rightarrow \quad \begin{align*}
\text{maximize} & \quad g(\lambda) \\
\text{subject to} & \quad \lambda \geq 0
\end{align*}
\]
Weak Duality

\[ f_0(x) \]

\[ g(\lambda) \]
**Strong Duality**

\[
p^* = d^*
\]

\[
f_0(x)
\]

\[
g(\lambda)
\]
Gradient Characterization of Convexity

**Theorem**

Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable. Then $f$ is convex iff

$$f(x + v) \geq f(x) + \nabla f(x)^T v$$

hold for all $x, v \in \mathbb{R}^d$. 
Gradient Approximation Gives Global Underestimator

\[ f(x_0) + f'(x)(x - x_0) \]
Subgradients

Definition (Subgradient, Subdifferential, Subdifferentiable)

Let $f : \mathbb{R}^d \to \mathbb{R}$. We say that $g \in \mathbb{R}^d$ is a subgradient of $f$ at $x \in \mathbb{R}^d$ if

$$f(x + v) \geq f(x) + g^T v$$

for all $v \in \mathbb{R}^d$. The subdifferential $\partial f(x)$ is the set of all subgradients of $f$ at $x$. We say that $f$ is subdifferentiable at $x$ if $\partial f(x) \neq \emptyset$ (i.e., if there is at least one subgradient).
Subgradients at $x_0$ and $x_1$
Facts About Subgradients

1. If $f$ is convex and differentiable at $x$, then $\partial f(x) = \{\nabla f(x)\}$. 

2. If $f$ is convex, then $\partial f(x) \neq \emptyset$ for all $x$.

3. The subdifferential $\partial f(x)$ is a convex set. Thus the subdifferential can contain 0, 1, or infinitely many elements.

4. If the zero vector is a subgradient of $f$ at $x$, then $x$ is a global minimum.

5. If $g$ is a subgradient of $f$ at $x$, then $(g, -1)$ is orthogonal to the underestimating hyperplane $\{ (x + v, f(x) + g^T v) | v \in \mathbb{R}^d \}$ at $(x, f(x))$. 
Facts About Subgradients

1. If $f$ is convex and differentiable at $x$ then $\partial f(x) = \{\nabla f(x)\}$.
2. If $f$ is convex then $\partial f(x) \neq \emptyset$ for all $x$. 
Facts About Subgradients

1. If \( f \) is convex and differentiable at \( x \) then \( \partial f(x) = \{\nabla f(x)\} \).

2. If \( f \) is convex then \( \partial f(x) \neq \emptyset \) for all \( x \).

3. The subdifferential \( \partial f(x) \) is a convex set. Thus the subdifferential can contain 0, 1, or infinitely many elements.
Facts About Subgradients

1. If $f$ is convex and differentiable at $x$ then $\partial f(x) = \{\nabla f(x)\}$.

2. If $f$ is convex then $\partial f(x) \neq \emptyset$ for all $x$.

3. The subdifferential $\partial f(x)$ is a convex set. Thus the subdifferential can contain 0, 1, or infinitely many elements.

4. If the zero vector is a subgradient of $f$ at $x$, then $x$ is a global minimum.
Facts About Subgradients

1. If $f$ is convex and differentiable at $x$ then $\partial f(x) = \{\nabla f(x)\}$.
2. If $f$ is convex then $\partial f(x) \neq \emptyset$ for all $x$.
3. The subdifferential $\partial f(x)$ is a convex set. Thus the subdifferential can contain 0, 1, or infinitely many elements.
4. If the zero vector is a subgradient of $f$ at $x$, then $x$ is a global minimum.
5. If $g$ is a subgradient of $f$ at $x$, then $(g, -1)$ is orthogonal to the underestimating hyperplane $\{(x + v, f(x) + g^T v) \mid v \in \mathbb{R}^d\}$ at $(x, f(x))$. 
Compute the Subdifferentials of $f(x) = |x|$
Compute $\partial f(3, 0)$ For $f(x_1, x_2) = |x_1| + 2|x_2|$
Compute $\partial f(3, 0)$ For $f(x_1, x_2) = |x_1| + 2|x_2|$
Compute $\partial f(3,0)$ For $f(x_1, x_2) = |x_1| + 2|x_2|$

$$\partial f(3,0) = \{(1, b)^T \mid b \in [-2, 2]\}$$
Gradient Lies Normal To Contours

**Theorem**

If \( f : \mathbb{R}^d \rightarrow \mathbb{R} \) is continuously differentiable and \( x_0 \in \mathbb{R}^d \) with \( \nabla f(x_0) \neq 0 \) then \( \nabla f(x_0) \) is normal to the level set \( S = \{ x \in \mathbb{R}^d \mid f(x) = f(x_0) \} \).
Gradient Lies Normal To Contours

\[ \nabla f(x) \]
Normal Plane to Subgradient Splits Space

\[ g^T (y - v) < 0 \]

\[ f(y) \geq f(v) + g^T (y - v) > f(v) \]
1. Let $x^{(0)}$ denote the initial point.

2. For $k = 1, 2, \ldots$
   - Assign $x^{(k)} = x^{(k-1)} - \alpha_k g$, where $g \in \partial f(x^{(k-1)})$ and $\alpha_k$ is the step size.
   - Set $f_{\text{best}}^{(k)} = \min_{i=1, \ldots, k} f(x^{(i)})$. (Used since this isn’t a descent method.)
Convergence of Subgradient Descent

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and Lipschitz with constant $G$, and let $x^*$ be a minimizer. For a fixed step size $t$, the subgradient method satisfies:

$$\lim_{k \to \infty} f(x^{(k)}_{\text{best}}) \leq f(x^*) + \frac{G^2 t}{2}.$$  

For step sizes respecting the Robbins-Monro conditions,

$$\lim_{k \to \infty} f(x^{(k)}_{\text{best}}) = f(x^*).$$