Definition 1. The minimum of a function $f : \mathbb{R}^d \to \mathbb{R}$ is written as $\min_x f(x)$, and has the property that $\min_x f(x) \leq f(y)$ for any $y$.

Definition 2. The value $x^*$ such that $f(x^*) = \min_x f(x)$ is called a minimizer.

Example 1. For the parabola $f(x) = x^2 + 4x - 1 = (x + 2)^2 - 5$, the minimum is $-5$ and the minimizer is $x = -2$.

Definition 3. A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if for any $0 \leq \alpha \leq 1$, we have

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

for any $x$ and $y$.

Definition 4. A function $f$ is concave if $-f$ is convex.

Example 2. If $f$ is convex, then

$$f(x) \geq f(y) + \nabla f(y)^\top (x - y)$$

for any $x$ and $y$.

We can arrange the following

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

into

$$f(y) + \frac{f(y + \alpha(x - y)) - f(y)}{\alpha} \leq f(x).$$

Remember that this holds for any $0 \leq \alpha \leq 1$. In particular, if we take the limit,

$$f(y) + \lim_{\alpha \to 0} \frac{f(y + \alpha(x - y)) - f(y)}{\alpha} = f(y) + \nabla f(y)^\top (x - y) \leq f(x).$$
**Definition 5.** A matrix $A$ is positive semidefinite if $v^\top Av \geq 0$ for all $v$, and is written as $A \succeq 0$.

**Example 3.** A function is convex if its Hessian is positive semidefinite.

The proof relies on mean-value theorem. It’s not difficult, but is beyond the scope of this course.

**Example 4.** Show that the mean-squared error $\ell(y, \hat{y}) = (y - \hat{y})^2$ is convex in $\hat{y}$.

\[
\frac{\partial^2}{\partial \hat{y}^2} \ell = 2 \geq 0. \tag{6}
\]

**Example 5.** Show that the function

\[ f(x) = x^\top \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} x \tag{7} \]

is convex.

The Hessian of $f$ is $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$. For any $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, we have

\[ \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 2v_1^2 + 3v_2^2 \geq 0 \tag{8} \]

The Hessian of $f$ is positive semidefinite.

**Example 6.** Show that the Hessian of $f(x) = \|x\|_2^2$ is $2I$, and hence $\|x\|_2^2$ is convex in $x$.

\[ \frac{\partial^2}{\partial x_i \partial x_j} f = 0 \quad \frac{\partial^2}{\partial x_i^2} f = 2 \tag{9} \]

**Example 7.** Show that if $f$ is convex, then $g(x) = f(Ax + b)$ is also convex.

\[ g(\alpha x + (1 - \alpha)y) = f(\alpha(Ax + b) + (1 - \alpha)(Ay + b)) \]
\[ \leq \alpha f(Ax + b) + (1 - \alpha)f(Ay + b) = \alpha g(x) + (1 - \alpha)g(y) \tag{10} \]

**Example 8.** Show that if $f_1, \ldots, f_k$ are convex, then $f = \beta_1 f_1 + \cdots + \beta_k f_k$ is also convex when $\beta_1, \ldots, \beta_k \geq 0$. 

2
\[ f(\alpha x + (1 - \alpha)y) = \beta_1 f_1(\alpha x + (1 - \alpha)y) + \cdots + \beta_k f_k(\alpha x + (1 - \alpha)y) \] 
\[ \leq \beta_1 \alpha_1 f_1(x) + \beta_1(1 - \alpha)f_1(y) + \cdots + \beta_k \alpha f_k(x) + \beta_k(1 - \alpha)f_k(y) \] 
\[ = \alpha(\beta_1 f_1(x) + \cdots + \beta_k f_k(x)) + (1 - \alpha)(\beta_1 f_1(y) + \cdots + \beta_k f_k(y)) \] 
\[ = \alpha f(x) + (1 - \alpha)f(y) \] 
\[ \leq \beta_1 \alpha f_1(x) + \beta_1(1 - \alpha)f_1(y) + \cdots + \beta_k \alpha f_k(x) + \beta_k(1 - \alpha)f_k(y) \] 
\[ = f(\alpha x + (1 - \alpha)y) \]

**Exercise 1.** Given a data set of \( n \) samples \( \{(x_1, y_1), \ldots, (x_n, y_n)\} \), show that

\[ L = \sum_{i=1}^{n} (w^\top x_i - y_i)^2 = \|Xw - y\|_2^2 \] 

if we have

\[ X = \begin{bmatrix} -x_1 & \cdots & -x_n \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\
\vdots \\
y_n \end{bmatrix}. \]

**Exercise 2.** Given a data set of \( n \) samples \( \{(x_1, y_1), \ldots, (x_n, y_n)\} \), show that the mean-squared error

\[ L = \|Xw - y\|_2^2 \]

is convex.

**Example 9.** Show that if \( f \) is convex and \( \nabla f(x^*) = 0 \) for a point \( x^* \), then \( x^* \) is the minimizer of \( f \).

Because \( f \) is convex, we have for any \( x \) and \( y \),

\[ f(x) \geq f(y) + \nabla f(y)^\top (x - y). \] 

In particular, if we let \( y = x^* \),

\[ f(x) \geq f(x^*) + \nabla f(x^*)^\top (x - x^*) = f(x^*). \]

**Example 10.** Show that \( \nabla_x (x^\top Ax) = (A^\top + A)x \).

We see that \( x^\top Ax \) is a real value. If we take the derivative of \( x^\top Ax \), we get

\[ \frac{\partial}{\partial x_k} \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij}x_i x_j = \sum_{i\neq j} a_{ik} x_i + \sum_{j \neq i} a_{kj} x_j + \sum_{i=1}^{d} 2a_{ii} x_i \] 
\[ = \sum_{i=1}^{d} a_{ik} x_i + \sum_{j=1}^{d} a_{kj} x_j = a_{ik}^\top x + a_k x \]
where \( a_{k} \) is the \( k \)-th column of \( A \) and \( a_{k} \) is the \( k \)-th row of \( A \).

**Example 11.** Show that \( w^* = (X^\top X)^{-1}X^\top y \) is the minimizer for \( L = \|Xw - y\|_2^2 \).

\[
L = (Xw - y)^\top (Xw - y) = w^\top X^\top Xw - 2y^\top Xw + y^\top y \tag{23}
\]
\[
\nabla L = (X^\top X + X^\top X)w - 2X^\top y = 0 \tag{24}
\]

If \( w^* = (X^\top X)^{-1}X^\top y \), then \( \nabla L(w^*) = 0 \). Because \( L \) is convex in \( w \), \( w^* \) is a minimizer of \( L \).

**Example 12.** Show that \( \ell(s) = \log(1 + \exp(-s)) \) is convex in \( s \).

\[
\frac{\partial \ell}{\partial s} = \frac{-\exp(-s)}{1 + \exp(-s)} = \frac{1}{1 + \exp(-s)} - 1 \tag{25}
\]
\[
\frac{\partial^2 \ell}{\partial s^2} = \frac{-1}{1 + \exp(-s)} \frac{-\exp(-s)}{1 + \exp(-s)} = \frac{1}{1 + \exp(-s)} \left( 1 - \frac{1}{1 + \exp(-s)} \right) \geq 0 \tag{26}
\]

**Exercise 3.** Given a data set of \( n \) samples \( \{(x_1, y_1), \ldots, (x_n, y_n)\} \), show that the log loss

\[
L = \sum_{i=1}^{n} \log \left( 1 + \exp(-y_i w^\top x_i) \right) \tag{27}
\]

is convex.

**Definition 6.** A function \( f : \mathbb{R}^d \to \mathbb{R} \) is called strictly convex if for \( 0 \leq \alpha \leq 1 \), we have

\[
f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y) \tag{28}
\]

for any \( x \neq y \).

**Exercise 4.** A function \( f : \mathbb{R}^d \to \mathbb{R} \) is strictly convex if

\[
f(x) > f(y) + \nabla f(y)^\top (x - y) \tag{29}
\]

for any \( x \neq y \).

**Definition 7.** A matrix \( A \) is positive definite if \( v^\top Av > 0 \) for any \( v \neq 0 \).
**Exercise 5.** A function $f : \mathbb{R}^d \to \mathbb{R}$ is strictly convex if its Hessian is positive definite.

**Example 13.** Show that if $f$ is strictly convex, then $f$ has a unique minimizer.

Suppose $x^*$ is a minimizer of $f$, i.e., $\nabla f(x^*) = 0$. The inequality

$$f(x) > f(y) + \nabla f(y)^\top (x - y).$$  \hspace{1cm} (30)

holds for any $x \neq y$. In particular, if we let $y = x^*$,

$$f(x) > f(x^*) + \nabla f(x^*)^\top (x - x^*) = f(x^*).$$  \hspace{1cm} (31)