Convex Functions (in one var is $f''(x) \geq 0$)

A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if for all $x$ and $y$ and all $0 \leq t \leq 1$,
\[ f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \quad (\star) \]

These look like

Or

- linear is convex.

$f$ is strictly convex if
\[ f(tx + (1-t)y) < tf(x) + (1-t)f(y) \]

$\|x\|_2$ is strictly convex, but $\|x\|_0$ and $\|x\|_1$ are merely convex.

$f$ is concave if $-f$ is convex.

I think of a cave.

$\text{dom}(f)$, the domain of $f$, is the set of $x$ for which $f$ is defined.

If $\text{dom}(f) \neq \mathbb{R}^n$, then $f$ is convex if
1. $\text{dom}(f)$ is a convex set, and
2. $(\star)$ holds for all $x, y \in \text{dom}(f)$
Example: \( f(x) = \frac{1}{x^2} \)

is convex if \( \text{dom}(f) = (0, \infty) \),
but not on \( IR \)

Same for \( x^3 \)

If \( f \) is convex but \( \text{dom}(f) \neq IR^n \),
the extension of \( f \), \( \tilde{f} \), is convex on \( IR^n \), where

\[
\tilde{f}(x) = \begin{cases} 
    f(x) & \text{for } x \in \text{dom}(f) \\
    \infty & \text{for } x \notin \text{dom}(f)
\end{cases}
\]

Why? We care about convex optimization problems:

\[
\min f(x) \quad \text{s.t. } x \in C
\]

where \( C \) is a convex set and \( f \) is a convex function.

This is one of the broadest classes of problems
that we can solve efficiently.

If \( f \) is strictly convex, the minimum is unique:

if \( x_1, x_2 \in \text{arg min } f(x) \), but \( x_1 \neq x_2 \),
then \( f(\frac{1}{2}x_1 + \frac{1}{2}x_2) < \frac{1}{2} f(x_1) + f(x_2) \),
contradicting minimality of \( x_1, x_2 \).
Local minima of convex functions are global minima. 
\( x \) is a \textbf{local minimum} of \( f \) if \( \exists \varepsilon > 0 \) s.t. 
\[ \| x - y \| < \varepsilon \implies f(x) \leq f(y) \] 
\( x \) is a \textbf{global minimum} if \( \forall y \) \( f(x) \leq f(y) \).

If \( x \) were a local minimum but \textbf{NOT} global, 
then \( \exists y \) s.t. \( f(y) < f(x) \). 
For every \( \varepsilon > 0 \) there is a \( t \in [0,1] \) s.t. 
\[ \| x - ([1-t]x + ty) \| < \varepsilon. \] 
For this \( t \), 
\[ f([1-t]x + ty) < (1-t)f(x) + tf(y) < f(x). \] 
This \textbf{contradicts} the assertion that \( x \) is a local minimum.

Examples of convex functions.

One variable:
\[ e^{ax} \text{ for } a \in \mathbb{R} \]  
\[ x^2 \text{ for } x \geq 1 \text{ or } x \leq 0, \quad x \in [0,\infty] \]  
\[ -x^2 \text{ for } 0 \leq x \leq 1, \quad x \in [0,\infty] \]  
\[ -\log(x), \quad x \in [0,\infty], \quad \text{because } \log(x) \text{ is concave} \]

Vectors

\[ \| x \| \]

Indicator functions: 
\[ I(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases} \]

\( C \) is convex. (could replace \( \infty \) with 1)
Affine functions: \( f(x) = \alpha^T x + b \)

Rules:

Non-negative sums:
\( f_i, f_n \) convex, \( \omega_1 \ldots \omega_n \geq 0 \Rightarrow \sum \omega_i f_i \) is convex

Maximum: the maximum of convex is convex
\( f(x) = \max (f_1(x), f_2(x)) \)

Ex. maximum of linear is convex

Ex. \( \max (x(1), x(2), \ldots, x(n)) \)

Ex. Sum of largest \( k \) components
\[
= \max \sum_{|S|=k} x(i)
\]

Affine composition:
if \( g \) is convex, then so is \( f(x) = g(Ax + b) \)
Examples: \( \| A x \|_2 \) is convex \( \Rightarrow \) \( \| A x \|_2 \) is convex

\( \Rightarrow x^T (A^T A) x \) is convex.

Every positive semidefinite \( Q = A^T A \), for some \( A \). So, \( x^T Q x \) is convex.

Least squares: \( \| A x - b \|_2^2 \) is convex

\( l_1 \)-regularized: \( \| A x - b \|_2^2 + \lambda \| x \|_1 \) is convex

Logistic loss: \( \sum_i \log (1 + e^{y_i (a_i^T x + b)}) = f(x) \)

\( g(z) = \log (1 + e^{y z}) \) is convex in \( g \), because

\[ g''(z) = \frac{y^2 e^{y z}}{(1 + e^{y z})^2} \]

\( a_i^T x + b \) is affine.

Composition rules (one of many) 
If \( f(x) = h(g(x)) \) where \( h : \mathbb{R} \rightarrow \mathbb{R} \), \( g : \mathbb{R}^n \rightarrow \mathbb{R} \),

\( g \) convex & \( h \) convex, non-decreasing \( \Rightarrow f \) convex

idea: in one var, \( f'(x) = g'(x) h'(g(x)) \)

\[ f''(x) = g''(x) h'(g(x)) + g'(x) h''(g(x)) \]

\[ > 0 \quad > 0 \quad > 0 \]
Functions & Sets:

The \( \alpha \)-sublevel set of \( f \) is \( \{ x : f(x) \leq \alpha \} \).

If \( f \) is convex then so are its \( \alpha \)-sublevel sets.

The graph of a function \( f \) is \( \{(x, f(x)) : x \in \text{dom}(f)\} \).

The epigraph of \( f \) is the region above:

\[ \{(x, t) : x \in \text{dom}(f), \; f(x) \leq t\} \]

like the ice cream cone, or:

\[ \text{convex} \quad \text{not convex} \]

\( f \) is a convex function iff its epigraph is a convex set.

In one variable, a twice-differentiable function \( f(x) \) is convex iff \( f''(x) \geq 0, \; \forall x \in \text{dom}(f) \).

This says that the function always lies above tangent lines, like
**Theorem 1**

In \( \mathbb{R}^n \), if \( f \) is differentiable then \( f \) is convex iff

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

Note that \( h(y) = f(x) + \nabla f(x)^T (y-x) \) is the linear function in \( y \) such that \( h(x) = f(x) \), and is a supporting hyperplane of epigraph at \( (x, f(x)) \).

**Proof of Theorem 1**

We first prove \( (*) \Rightarrow \text{convex} \)

Let \( x, y \in \text{dom}(f), \ 0 < \lambda < 1, \ z = \lambda x + (1-\lambda) y \).

Let \( g = \nabla f(z) \).

\[ (*) \Rightarrow \lambda f(x) \geq \lambda f(z) + \lambda g^T (x-z) \]

\[ (1-\lambda) f(y) \geq (1-\lambda) f(z) + (1-\lambda) g^T (y-z) \]

\[ \Rightarrow \lambda f(y) + (1-\lambda) f(y) = f(z) + g^T (\lambda x - \lambda z + (1-\lambda) y - (1-\lambda) z) \]

Convex \( \Rightarrow (* \)

For \( x, y \in \text{dom}(f), \ 0 < \lambda < 1 \),

\[ (1-\lambda) f(x) + \lambda f(y) \leq f((1-\lambda) x + \lambda y) = f(x + \lambda (y-x)) \]

\[ \Rightarrow f(y) \geq f(x) + \frac{f(x + \lambda (y-x)) - f(x)}{\lambda} \]

Taking \( \lim \) as \( \lambda \to 0 \) gives

\[ f(y) + \nabla f(x)^T (y-x) \]
Thus, if \( f \) is twice differentiable,

\[
f \text{ is convex iff } \nabla^2 f(x) \text{ is psd where } \nabla^2 f(x) \text{ is matrix with entries } \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \nabla^2 f(x) \text{ is matrix with entries } \frac{\partial^2 f}{\partial x_i \partial x_j}(x)
\]

Computing \( \|A\|_2^2 = \max_{\|x\|_2=1} \|Ax\|_2^2 = \max_{\|x\|_2=1} x^T (ATA)x \)

Does not seem a convex program, because

1. \( \|x\|_2=1 \) is not a convex set.
2. \( \|x\|_2=1 \) is not a convex set.
 Restricting to \( \|x\|_2 \leq 1 \) solves ii but not i.

Solution: write as

\[
\min t \text{ s.t. } tI - A \succeq 0
\]

\( (M \succeq 0 \text{ iff } M \text{ is positive semidefinite}) \).

Is a convex cone.

Can check if \( M \succeq 0 \) by trying to compute

a Cholesky factorization: \( L \) s.t. \( LL^\top = M \).

\( S_n^+ \) = set of symmetric \( n \times n \) positive semidefinite.
If $A$ is symmetric, but $A \notin S^+_n$, then $\nu^T A \nu < 0$.

A hyperplane separating $A$ from $S^+_n$ is given by

$$\{ \text{symmetric } X : \nu^T X \nu = 0 \}$$

$\nu^T X \nu \geq 0$ for $X \in S^+_n$, $\nu^T A \nu < 0$.

And is a hyperplane because

$$\nu^T X \nu = \sum_{1 \leq i \leq n} X(i,i) \nu(i) \nu(i)$$

is linear in $X$. In space of symmetric matrices.