CPSC 532D — C. CONVEXITY

Danica J. Sutherland

University of British Columbia, Vancouver

Fall 2025

Here we'll give a quick overview of *convex functions*. More details are available lots of places; in addition to chapters 12-13 of [SSBD14] or Appendix B.2 of [MRT18], the classic super-detailed reference is the book of Rockafellar [Roc70], and Boyd and Vandenbreghe [BV04] is also good (and what I learned from).

Most sources assume functions on $\mathbb{R}^d$; we'll assume a separable Hilbert space $\mathcal{X}$, though the statements e.g. that don't use an inner product will also hold for Banach spaces, and so on. For the results about derivatives, you can use a Fréchet derivative, and have a gradient/Hessian analogue. Don't really worry about any of that, you can just think of everything as on $\mathbb{R}^d$.

Definition C.1. A *set* $C \subseteq \mathcal{X}$ is convex if for all $x_0, x_1 \in C$ and $\alpha \in [0, 1]$, it holds that $(1-\alpha)x_0 + \alpha x_1 \in C$.

Below, we'll use the set $\mathbb{R} \cup \{\infty\}$ a lot. Many of these results hold for the full *extended real line* $\mathbb{R} \cup \{-\infty, \infty\}$, but you often have to exclude $-\infty$ for things to make sense.

It's typical in optimization to, rather than dealing with functions on some restricted domain that's a proper subset of $\mathcal{X}$, instead define $f(x) = \infty$ for x that shouldn't be in the domain. Then $\operatorname{dom} f = \{x \in \mathcal{X} : f(x) < \infty\}$.

Definition C.2. A *function* $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is called

- *convex* if it lies below its chords: for all $x_0, x_1 \in \mathcal{X}$ and $\alpha \in (0, 1)$,
$$f((1-\alpha)x_0 + \alpha x_1) \le (1-\alpha) f(x_0) + \alpha f(x_1);$$
- *strictly convex* if this inequality is strict;
- and *m-strongly convex*, for some $m > 0$, if
$$f((1-\alpha)x_0 + \alpha x_1) \le (1-\alpha) f(x_0) + \alpha f(x_1) - \tfrac{1}{2} m \alpha (1-\alpha) \lVert x_1 - x_0 \rVert^2.$$

A function is convex if and only if its *epigraph*, $\{(x, r) \in \mathcal{X} \times (\mathbb{R} \cup \{\infty\}) : r \ge f(x)\}$, is a convex set.

An m-strongly convex function is m'-strongly convex for any $m' < m$; convexity is equivalent to 0-strong convexity, which we don't call strongly convex. m-strong convexity implies strict convexity, but the reverse is not true. Likewise, strict convexity implies convexity.

A concave/strictly concave/m-strongly concave function is one where $-f$ is convex/strictly convex/m-strongly convex.

For more, visit https://cs.ubc.ca/~dsuth/532D/25w1/.

Any local minimum of a convex function must be a global minimum, since we can connect any two local minima by chords. The set of global minima must be convex, for the same reason. If f is strictly convex, it has only one global minimum.

C.1 FIRST-ORDER CONDITIONS

Proposition C.3. *If $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is differentiable on its convex domain,*

- *it is convex iff it lies above its tangents: for all $x, x' \in \operatorname{dom} f$,*

$$f(x') \geq f(x) + \langle \nabla f(x), x' - x \rangle.$$

- *it is m-strongly convex iff for all $x, x' \in \mathcal{X}$,*

$$f(x') \geq f(x) + \langle \nabla f(x), x' - x \rangle + \tfrac{1}{2} m \|x' - x\|^2.$$

Proof. We'll do this for $m \geq 0$, in which case the $m = 0$ results are for plain convexity.

If $f(x') \geq f(x) + \langle \nabla f(x), x' - x \rangle + \frac{1}{2} m \|x' - x\|^2$ for all x, x', then

$$\begin{aligned}
f((1-\alpha)x + \alpha x') &\leq (1-\alpha) f(x) + \alpha f(x') - \tfrac{1}{2} m \alpha (1-\alpha) \|x' - x\|^2 \\
\frac{1}{\alpha}\Big[f((1-\alpha)x + \alpha x') - f(x) \Big] &\leq f(x') - f(x) - \tfrac{1}{2} m (1-\alpha) \|x' - x\|^2 \\
\lim_{\alpha \to 0} \frac{f(x + \alpha(x' - x)) - f(x)}{\alpha} &\leq f(x') - f(x) - \tfrac{1}{2} m \|x' - x\|^2,
\end{aligned}$$

and that limit is exactly the directional derivative given by $\langle \nabla f(x), x' - x \rangle$.

In the other direction, let $x_\alpha = (1-\alpha) x_0 + \alpha x_1$, and note $x_\alpha - x_0 = \alpha(x_1 - x_0)$, $x_\alpha - x_1 = -(1-\alpha)(x_1 - x_0)$. Then

$$\begin{aligned}
f(x_\alpha) &\leq f(x_0) + \langle \nabla f(x_\alpha), x_\alpha - x_0 \rangle - \tfrac{1}{2} m \|x_\alpha - x_0\|^2 \\
&= f(x_0) + \alpha \langle \nabla f(x_\alpha), x_1 - x_0 \rangle - \tfrac{1}{2} m \alpha^2 \|x_1 - x_0\|^2
\end{aligned}$$

and

$$\begin{aligned}
f(x_\alpha) &\leq f(x_1) + \langle \nabla f(x_\alpha), x_\alpha - x_1 \rangle - \tfrac{1}{2} m \|x_\alpha - x_1\|^2 \\
&= f(x_1) - (1-\alpha) \langle \nabla f(x_\alpha), x_1 - x_0 \rangle - \tfrac{1}{2} m (1-\alpha)^2 \|x_1 - x_0\|^2.
\end{aligned}$$

Adding $1 - \alpha$ times the first inequality plus α times the second yields

$$f(x_\alpha) \leq (1-\alpha) f(x_0) + \alpha f(x_1) - \tfrac{1}{2} m \alpha (1-\alpha)(\alpha + 1 - \alpha) \|x_1 - x_0\|^2. \qquad \square$$

Proposition C.4. *If $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is continuously differentiable on its convex domain,*

- *it is convex iff $\forall x, x' \in \operatorname{dom} f$, $\langle \nabla f(x) - \nabla f(x'), x - x' \rangle \geq 0$;*
- *it is m-strongly convex iff $\forall x, x' \in \operatorname{dom} f$, $\langle \nabla f(x) - \nabla f(x'), x - x' \rangle \geq m \|x - x'\|^2$.*

This result is important for convex optimization: if x^* is a minimizer, $\nabla f(x^*) = 0$, and so then $m \|x - x^*\|^2 \leq \langle \nabla f(x), x - x^* \rangle \leq \|\nabla f(x)\| \, \|x - x^*\|$, i.e. $\|x - x^*\| \leq \frac{1}{m} \|\nabla f(x)\|$, and if we know $m > 0$ then the right-hand side is something we can actually measure for any point x and upper-bound how far we can be from the minimizer.

Proof. We'll again use $m = 0$ for plain convexity.

If f is convex/m-strongly convex, then

$$f(x) \geq f(x') + \langle \nabla f(x'), x - x' \rangle + \tfrac{1}{2} m \|x - x'\|^2$$
$$f(x') \geq f(x) - \langle \nabla f(x), x - x' \rangle + \tfrac{1}{2} m \|x - x'\|^2$$

and so

$$f(x) + f(x') \geq f(x') + f(x) + \langle \nabla f(x') - \nabla f(x), x - x' \rangle + m\|x - x'\|^2.$$

In the other direction, again using $x_\alpha = (1-\alpha)x_0 + \alpha x_1$ we know that

$$\begin{aligned}
f(x_1) = f(x_0) &+ \int_0^1 \langle f(x_\alpha), x_1 - x_0 \rangle \mathrm{d}\alpha \\
&= f(x_0) + \langle \nabla f(x_0), x_1 - x_0 \rangle + \int_0^1 \langle f(x_\alpha) - \nabla f(x_0), x_1 - x_0 \rangle \mathrm{d}\alpha \\
&= f(x_0) + \langle \nabla f(x_0), x_1 - x_0 \rangle + \int_0^1 \frac{1}{\alpha} \langle f(x_\alpha) - \nabla f(x_0), x_\alpha - x_0 \rangle \mathrm{d}\alpha \\
&\geq f(x_0) + \langle \nabla f(x_0), x_1 - x_0 \rangle + \int_0^1 \frac{1}{\alpha} m \|x_\alpha - x_0\|^2 \mathrm{d}\alpha \\
&= f(x_0) + \langle \nabla f(x_0), x_1 - x_0 \rangle + m\|x_1 - x_0\|^2 \int_0^1 \alpha \mathrm{d}\alpha \\
&= f(x_0) + \langle \nabla f(x_0), x_1 - x_0 \rangle + \tfrac{1}{2} m \|x_1 - x_0\|^2.
\end{aligned}$$

□

C.2 SECOND-ORDER CONDITIONS

The notation $\mathrm{A} \succeq 0$ means that the square matrix (or Hilbert-space operator) A is positive semi-definite; $\mathrm{A} \succeq \mathrm{B}$ means that $\mathrm{A} - \mathrm{B} \succeq 0$. Thus $\mathrm{A} \succeq m\mathrm{I}$ means that all eigenvalues of A are at least m. The notation $\nabla^2 f$ denotes the Hessian, the matrix of all second derivatives. (This is a $\mathcal{F} \to \mathcal{F}$ operator in Hilbert spaces.)

If f is a function on scalars, $\nabla^2 f(x) \succeq m\mathrm{I}$ exactly means than $f''(x) \geq m$.

Proposition C.5. *If $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is continuously twice-differentiable on its convex domain,*

- *it is convex iff* $\forall x \in \operatorname{dom} f$, $\nabla^2 f \succeq 0$;
- *it is m-strongly convex iff* $\forall x \in \operatorname{dom} f$, $\nabla^2 f \succeq m\mathrm{I}$.

Proof. Again use $m = 0$ for the plain convexity case, and $x_\alpha = (1-\alpha)x_0 + \alpha x_1$.

If f is convex / m-strongly convex, then using Proposition C.4 gives

$$\begin{aligned}
m\|x_1 - x_0\|^2 &\leq \langle \nabla f(x_1) - \nabla f(x_0), x_1 - x_0 \rangle \\
&= \left\langle \int_0^1 \nabla^2 f(x_\alpha)(x_1 - x_0)\,\mathrm{d}\alpha, x_1 - x_0 \right\rangle \\
&= \left\langle x_1 - x_0, \left[\int_0^1 \nabla^2 f(x_\alpha)\,\mathrm{d}\alpha\right](x_1 - x_0) \right\rangle \\
0 &\leq \left\langle x_1 - x_0, \left[\int_0^1 \nabla^2 f(x_\alpha)\,\mathrm{d}\alpha - m\mathrm{I}\right](x_1 - x_0) \right\rangle.
\end{aligned}$$

Now, let x_0 be any point in the interior of the domain and let $x_1 = x_0 + \varepsilon v$, getting

$$\begin{aligned}
\left\langle \varepsilon v, \left[\int_0^1 \nabla^2 f(x_0 + \varepsilon\alpha v)\,\mathrm{d}\alpha - m\mathrm{I}\right] \varepsilon v \right\rangle \geq 0 \\
\left\langle v, \left[\int_0^1 \nabla^2 f(x_0 + \varepsilon\alpha v)\,\mathrm{d}\alpha - m\mathrm{I}\right] v \right\rangle \geq 0.
\end{aligned}$$

As $\varepsilon \to 0$, we have that $\int_0^1 \nabla^2 f(x_0 + \varepsilon\alpha v)\mathrm{d}\alpha \to \nabla^2 f(x_0)$ since $\nabla^2 f$ is continuous. Thus $\langle v, (\nabla^2 f(x) - m\mathrm{I})v \rangle \geq 0$ for all x in the interior of the domain and all v. This is exactly the condition that $\nabla^2 f(\mathrm{X}) \geq m\mathrm{I}$.

For the other direction, we have that

$$\begin{aligned}
f(x') &= f(x) + \langle \nabla f(x), x' - x \rangle + \int_0^1 \int_0^\alpha \langle x' - x, \nabla^2 f(x_\tau)(x' - x) \rangle \mathrm{d}\tau\,\mathrm{d}\alpha \\
&\geq f(x) + \langle \nabla f(x), x' - x \rangle + \int_0^1 \int_0^\alpha m\|x' - x\|^2 \mathrm{d}\tau\,\mathrm{d}\alpha \\
&= f(x) + \langle \nabla f(x), x' - x \rangle + \tfrac{1}{2} m\|x' - x\|^2
\end{aligned}$$

since $\int_0^\alpha \mathrm{d}\tau = \alpha$, and $\int_0^1 \alpha\,\mathrm{d}\alpha = \frac{1}{2}$. □

C.3 PROPERTIES

Proposition C.6. *If f, g, and f_y for all $y \in \mathcal{Y}$ are all convex functions, then so are*

- *αf for any $\alpha \geq 0$;*
- *$f + g$, or more generally $\int f_y \mathrm{d}w(y)$ if w is any (nonnegative) measure on $\mathcal{Y}$;*
- *$x \mapsto f(\mathrm{A}x + b)$ for any A, b;*
- *$x \mapsto g(f(x))$ if $g : \mathbb{R} \to \mathbb{R}$ is also nondecreasing;*

- $x \mapsto \max(f(x), g(x))$, *or more generally* $x \mapsto \sup_{y \in \mathcal{Y}} f_y(x)$;
- $x \mapsto \inf_{y \in \mathcal{Y}} f(x, y)$ *if* $f(x, y)$ *is convex in* (x, y), *and* $\mathcal{Y}$ *is a nonempty convex set.*

The proofs are mostly straightforward, and omitted here.

Similarly, the sum of an m-strongly convex and an m'-strongly convex function is $(m + m')$-strongly convex, and the sum of an m-strongly convex function with a convex function is m-strongly convex. Scaling an m-strongly convex function by $\alpha > 0$ gives you an $m\alpha$-strongly convex function.

Notice that the square loss, hinge loss, and logistic loss are all convex functions of the function h.

Theorem C.7 (Jensen's inequality)**.** *If* $f : \mathcal{X} \to \mathbb{R}$ *is convex and* X *a random variable on* $\mathcal{X}$ *such that the expectations exist,* $f(\mathbb{E}\,\mathrm{X}) \le \mathbb{E}\, f(\mathrm{X})$.

Proposition C.8. *The function* $w \mapsto \|w\|^2$ *is 2-strongly convex.*

Proof. Its gradient is $2h$, and so its Hessian is $2\mathrm{I}$. Or, more directly,

$$
\begin{aligned}
\|(1-\alpha)w + \alpha v\|^2 + \tfrac{1}{2} \cdot 2 \cdot \alpha(1-\alpha)\|w - v\|^2 \\
&= (1-\alpha)^2\|w\|^2 + \alpha^2\|v\|^2 + 2\alpha(1-\alpha)\langle w, v\rangle \\
&\qquad + \alpha(1-\alpha)\|w\|^2 + \alpha(1-\alpha)\|v\|^2 - 2\alpha(1-\alpha)\langle w, v\rangle \\
&= (1-\alpha+\alpha)(1-\alpha)\|w\|^2 + \alpha(\alpha+1-\alpha)\|v\|^2 \\
&= (1-\alpha)\|w\|^2 + \alpha\|v\|^2. \qquad \square
\end{aligned}
$$

Thus $\frac{1}{2}\|\cdot\|^2$ is 1-strongly convex.