Convex functions

The domain \( \text{dom} \ f \) of a functional \( f : \mathbb{R}^N \to \mathbb{R} \) is the subset of \( \mathbb{R}^N \) where \( f \) is well-defined.

A function(al) \( f \) is \textbf{convex} if \( \text{dom} \ f \) is a convex set, and
\[
f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)
\]
for all \( x, y \in \text{dom} \ f \) and \( 0 \leq \theta \leq 1 \).

\( f \) is \textbf{concave} if \(-f\) is convex.

\( f \) is \textbf{strictly convex} if \( \text{dom} \ f \) is convex and
\[
f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)
\]
for all \( x \neq y \in \text{dom} \ f \) and \( 0 < \theta < 1 \).

The domain matters. For example,
\[
f(x) = x^3
\]
is convex if \( \text{dom} \ f = \mathbb{R}_+ = [0, \infty] \) but not if \( \text{dom} \ f = \mathbb{R} \).
We define the **extension** of $f$ from $\text{dom } f$ to all of $\mathbb{R}^N$ as

$$\tilde{f}(x) = f(x), \quad x \in \text{dom } f, \quad \tilde{f}(x) = +\infty, \quad x \notin \text{dom } f.$$ 

If $f$ is convex on $\text{dom } f$, then its extension is also convex on $\mathbb{R}^N$.

Here are some standard examples for functions on $\mathbb{R}$:

- **affine functions** $f(x) = ax + b$ are both convex and concave for $a, b \in \mathbb{R}$,
- **exponentials** $f(X) = e^{ax}$ are convex for all $a \in \mathbb{R}$,
- **powers** $x^\alpha$ are convex on $\mathbb{R}_+$ for $\alpha \geq 1$, concave for $0 \leq \alpha \leq 1$, and convex for $\alpha \leq 0$,
- $|x|^\alpha$ is convex on all of $\mathbb{R}$ for $\alpha \geq 1$.
- the entropy function $x \log x$ is concave on $\mathbb{R}_{++},$
- **logarithms**: $\log x$ is concave on $\mathbb{R}_{++}$.

Here are some standard examples for functionals on $\mathbb{R}^N$:

- **affine functions** $f(x) = \langle x, a \rangle + b$ are both convex and concave on all of $\mathbb{R}^N$,
- **any valid norm** $f(x) = \|x\|$ is convex on all of $\mathbb{R}^N$
- etc
A functional $f : \mathbb{R}^N \to \mathbb{R}$ is convex if and only if the function $g_v : \mathbb{R} \to \mathbb{R}$,

$$g_v(t) = f(\mathbf{x} + tv), \quad \text{dom } g = \{ t : \mathbf{x} + tv \in \text{dom } f \}$$

is convex for every $\mathbf{x} \in \text{dom } f$, $v \in \mathbb{R}^N$.

**Example:** Let $f(\mathbf{X}) = -\log \det \mathbf{X}$ with $\text{dom } f = S_{++}^N$. For any $\mathbf{X} \in S_{++}^N$, we know that

$$\mathbf{X} = \mathbf{U} \Lambda \mathbf{U}^T,$$

for some diagonal, positive $\Lambda$, so we can define

$$\mathbf{X}^{1/2} = \mathbf{U} \Lambda^{1/2} \mathbf{U}^T, \quad \text{and } \mathbf{X}^{-1/2} = \mathbf{U} \Lambda^{-1/2} \mathbf{U}^T.$$ 

Now consider

$$g_v(t) = -\log \det(\mathbf{X} + tv\mathbf{V}) = \log \det(\mathbf{X}^{1/2}(\mathbf{I} + t\mathbf{X}^{-1/2}\mathbf{V}\mathbf{X}^{-1/2})\mathbf{X}^{1/2})$$

$$= -\log \det \mathbf{X} - \log \det(\mathbf{I} + t\mathbf{X}^{-1/2}\mathbf{V}\mathbf{X}^{-1/2})$$

$$= -\log \det \mathbf{X} - \sum_{n=1}^{N} \log(1 + \sigma_i t),$$

where the $\sigma_i$ are the (positive) eigenvalues of $\mathbf{X}^{-1/2}\mathbf{V}\mathbf{M}^{-1/2}$. The function $-\log(1 + \sigma_i t)$ is convex, so the above is a sum of convex functions, which is convex.
First-order conditions for convexity

We say that $f$ is **differentiable** if $\text{dom } f$ is an open set (all of $\mathbb{R}^N$, for example), and the gradient

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_N} \end{bmatrix}$$

exists for each $x \in \text{dom } f$.

If $f$ is differentiable, then it is convex if and only if

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

for all $x, y \in \text{dom } f$.

This means that the linear approximation

$$g(y) = f(x) + \nabla f(x)^T(y - x)$$

is a **global underestimator** of $f(y)$.
It is easy to show that $f$ convex, differentiable $\Rightarrow$ (1). Since $f$ is convex,
\[
f(x + t(y - x)) \leq (1 - t)f(x) + tf(y), \quad 0 \leq t \leq 1,
\]
and so
\[
f(y) \geq f(x) + \frac{f(x + t(y - x)) - f(x)}{t}, \quad \forall 0 < t \leq 1.
\]
Taking the limit as $t \to 0$ on the right yields (1).

It is also true that (1) $\Rightarrow$ $f$ convex. For a proof, see [BV04, p. 70].

**Second-order conditions for convexity**

We say that $f : \mathbb{R}^N \to \mathbb{R}$ is **twice differentiable** if dom $f$ is an open set, and the $N \times N$ Hessian matrix

\[
\nabla^2 f(x) = \begin{bmatrix}
\frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_N} \\
\frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \frac{\partial^2 f(x)}{\partial x_N \partial x_1} \\
\frac{\partial^2 f(x)}{\partial x_N \partial x_2} & \cdots & \cdots & \frac{\partial^2 f(x)}{\partial x_N^2}
\end{bmatrix}
\]

exists for every $x \in \text{dom } f$.

If $f$ is twice differentiable, then it is convex if and only if
\[
\nabla^2 f(x) \succeq 0 \quad (\text{i.e. } \nabla^2 f(x) \in S^N_+).
\]
for all $x \in \text{dom } f$. It is strictly convex if an only if
\[
\nabla^2 f(x) \succ 0 \quad (\text{i.e. } \nabla^2 f(x) \in S^N_{++}).
\]
You will prove this on the next homework.
Standard examples (from [BV04])

**Quadratic functionals:**

\[ f(x) = \frac{1}{2} x^T P x + q^T x + r, \]

where \( P \) is symmetric, has

\[ \nabla f(x) = P x + q, \quad \nabla^2 f(x) = P, \]

so \( f(x) \) is convex iff \( P \succeq 0 \).

**Least-squares:**

\[ f(x) = \|Ax - b\|_2^2, \]

where \( A \) is an arbitrary \( M \times N \) matrix, has

\[ \nabla f(x) = 2A^T(Ax - b), \quad \nabla^2 f(x) = 2A^T A, \]

and is convex for any \( A \).

**Quadratic-over-linear:** In \( \mathbb{R}^2 \), if

\[ f(x) = x_1^2/x_2, \]

then

\[ \nabla f(x) = \begin{bmatrix} 2x_1/x_2 \\ -x_1^2/x_2^2 \end{bmatrix}, \quad \nabla^2 f(x) = \frac{2}{x_2^3} \begin{bmatrix} x_2^2 & -x_1 x_2 \\ -x_1 x_2 & x_1 \end{bmatrix} = \frac{2}{x_2^3} \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix} \begin{bmatrix} x_2 & -x_1 \end{bmatrix}, \]

and so \( f \) is convex on \([0, \infty] \times \mathbb{R} \) \((x_1 \geq 0, x_2 \in \mathbb{R})\).
**Epigraph**

The *epigraph* of a functional $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is the subset of $\mathbb{R}^{N+1}$ created by filling in the space above $f$:

$$
\text{epi } f = \left\{ \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} \in \mathbb{R}^{N+1} : \mathbf{x} \in \text{dom } f, \; f(\mathbf{x}) \leq t \right\}.
$$

$f$ is convex if and only if $\text{epi } f$ is a convex set.

The gradient of $f$ at $\mathbf{x}$, when it exists, is a supporting hyperplane of $\text{epi } f$ at $\begin{bmatrix} \mathbf{x} \\ f(\mathbf{x}) \end{bmatrix}$. 

\[
\text{epi } f
\]
References