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 **convex** if $\text{dom}\, f$ is a convex set, and

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \;\le\; \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \text{dom}\, f$ and $0 \le \theta \le 1$.

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

f is **strictly convex** if $\text{dom}\, f$ is convex and

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \;<\; \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$

for all $\boldsymbol{x} \neq \boldsymbol{y} \in \text{dom}\, f$ 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 $\mathrm{dom}\, f$ to all of $\mathbb{R}^N$ as

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

If f is convex on $\mathrm{dom}\, f$, 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^α 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 $f(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{a} \rangle + b$ are both convex and concave on all of $\mathbb{R}^N$,
- any valid norm $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ is convex on all of $\mathbb{R}^N$
- etc

A functional $f : \mathbb{R}^N \to R$ is convex if and only if the function $g_v : \mathbb{R} \to \mathbb{R}$,

$$g_v(t) = f(\boldsymbol{x} + t\boldsymbol{v}), \quad \operatorname{dom} g = \{t \ : \ \boldsymbol{x} + t\boldsymbol{v} \in \operatorname{dom} f\}$$

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

Example: Let $f(\boldsymbol{X}) = -\log\det \boldsymbol{X}$ with $\operatorname{dom} f = S^N_{++}$. For any $\boldsymbol{X} \in S^N_{++}$, we know that

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}},$$

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

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

Now consider

$$\begin{aligned} g_V(t) = -\log\det(\boldsymbol{X} + t\boldsymbol{V}) &= -\log\det(\boldsymbol{X}^{1/2}(\mathbf{I} + t\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2})\boldsymbol{X}^{1/2}) \\ &= -\log\det \boldsymbol{X} - \log\det(\mathbf{I} + t\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2}) \\ &= -\log\det \boldsymbol{X} - \sum_{n=1}^{N} \log(1 + \sigma_i t), \end{aligned}$$

where the σ_i are the (positive) eigenvalues of $\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2}$. 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 dom f is an open set (all of $\mathbb{R}^N$, for example), and the gradient

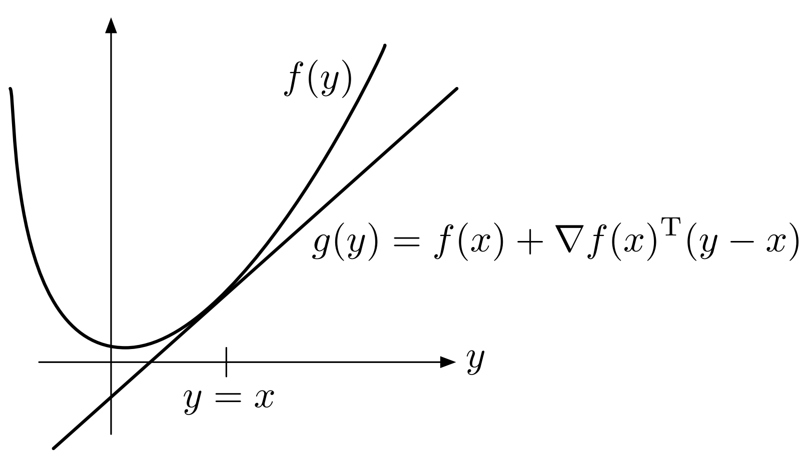
$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_N} \end{bmatrix}$$

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

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

$$f(\boldsymbol{y}) \;\geq\; f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \tag{1}$$

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



This means that the linear approximation
$g(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x})$
is a **global underestimator** of $f(\boldsymbol{y})$.

It is easy to show that f convex, differentiable $\Rightarrow$ (1). Since f is convex,

$$f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})) \ \leq \ (1-t)f(\boldsymbol{x}) + tf(\boldsymbol{y}), \quad 0 \leq t \leq 1,$$

and so

$$f(\boldsymbol{y}) \ \geq \ f(\boldsymbol{x}) + \frac{f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{t}, \quad \forall 0 < t \leq 1.$$

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 $\operatorname{dom} f$ is an open set, and the $N \times N$ Hessian matrix

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

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

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

$$\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0} \quad (\text{i.e. } \nabla^2 f(\boldsymbol{x}) \in S_+^N).$$

for all $\boldsymbol{x} \in \operatorname{dom} f$. It is strictly convex if an only if

$$\nabla^2 f(\boldsymbol{x}) \succ \mathbf{0} \quad (\text{i.e. } \nabla^2 f(\boldsymbol{x}) \in S_{++}^N).$$

You will prove this on the next homework.

Standard examples (from [BV04])

Quadratic functionals:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x} + r,$$

where $\boldsymbol{P}$ is symmetric, has

$$\nabla f(\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}, \quad \nabla^2 f(\boldsymbol{x}) = \boldsymbol{P},$$

so $f(\boldsymbol{x})$ is convex iff $\boldsymbol{P} \succeq \boldsymbol{0}$.

Least-squares:

$$f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2,$$

where $\boldsymbol{A}$ is an arbitrary $M \times N$ matrix, has

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}), \quad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A},$$

and is convex for any $\boldsymbol{A}$.

Quadratic-over-linear: In $\mathbb{R}^2$, if

$$f(\boldsymbol{x}) = x_1^2/x_2,$$

then

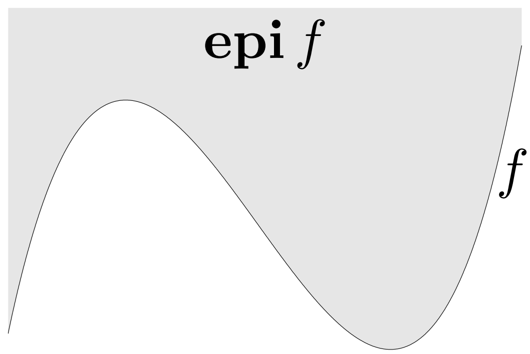
$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1/x_2 \\ -x_1^2/x_2^2 \end{bmatrix}, \quad \nabla^2 f(\boldsymbol{x}) = \frac{2}{x_2^3}\begin{bmatrix} x_2^2 & -x_1 x_2 \\ -x_1 x_2 & x_1 \end{bmatrix} = \frac{2}{x_2^3}\begin{bmatrix} x_2 \\ -x_1 \end{bmatrix}\begin{bmatrix} x_2 & -x_1 \end{bmatrix},$$

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 $f : \mathbb{R}^N \to \mathbb{R}$ is the subset of $\mathbb{R}^{N+1}$ created by filling in the space above f:

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



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

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

References

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.