Convex Functions

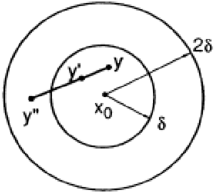
3. Local and Global Behaviour of a Convex Function

4. First- and Second-Order Differentiation

Lemma

Let $f \in Conv\mathbb{R}^n$. For $x_0 \in \mathbb{R}$ and $\delta > 0$, if $\exists\, m, M$ such that $m \leq f(x) \leq M$ for all $x \in B(x_0, 2\delta)$. Then, f is Lipschitzian on $B(x_0, \delta)$, i.e.,

$$|f(y) - f(y')| \leq \frac{M-m}{\delta}\|y - y'\|, \ \forall y, y' \in B(x_0, \delta).$$



proof with two different y and $y' \in B(x_0, \delta)$, take

$$y'' := y' + \delta \frac{y'-y}{\|y'-y\|} \in B(x_0, 2\delta);$$

By construction, y' lies on the segment $[y, y'']$, i.e.,

$$y' = \frac{\|y'-y\|}{\delta+\|y'-y\|} y'' + \frac{\delta}{\delta+\|y'-y\|} y.$$

By the convexity of f and the bounds in assumption, we obtain

$$f(y') - f(y) \leq \frac{\|y'-y\|}{\delta+\|y'-y\|}[f(y'') - f(y)] \leq \frac{1}{\delta}\|y' - y\|(M - m).$$

Theorem

Let $f \in \text{Conv}\mathbb{R}^n$ and $S \subset \text{ri}\,\text{dom} f$ be convex compact. Then, $\exists L(S) > 0$ such that

$$|f(x) - f(x')| \leq L(S)\|x - x'\|, \ \forall x, x' \in S.$$

- f is continuous in ri dom$f$, i.e., $\forall x_0 \in \text{ri}\,\text{dom} f$, if $\text{ri}\,\text{dom} f \ni x \to x_0$, then $f(x) \to f(x_0)$.
- f is locally Lipschitzian on ri dom$f$, i.e., $\forall x_0 \in \text{ri}\,\text{dom} f$, $\exists L(x_0)$, $\delta(x_0)$ such that

$$f(x) - f(x')| \leqslant L(x_0)\|x - x'\|, \ \forall x \in S(x_0), \ x' \in S(x_0),$$

  where $S(x_0) := B(x_0, \delta(x_0)) \cap \text{aff}\,\text{dom} f \subset \text{ri}\,\text{dom} f$.
- S in Theorem can not be weaken as ri dom$f$. ($\because$ convex function may not be Lipschitzian on ri dom$f$).

Theorem (Lipschitzian extension)

Let $f \in \text{Conv}\mathbb{R}^n$ be L-Lipschitzian on convex set $\emptyset \neq C$. Then, $\exists f_1 \in \text{Conv}\mathbb{R}^n$ satisfying $f_1(x) = f(x)$ on C and f_1 is L-Lipschitzian on $\mathbb{R}^n$. Furthermore, the largest f_1 is the infimal convolution
$(f + \iota_C)_{[L]}(x) := [(f + \iota_C) \star (L\|\cdot\|)](x) = \inf\{f(y) + L\|x - y\| : y \in C\}.$

Continuity of convex function

- aff dom$f$ is confined to continuity of f. Continuity (even Lipschitz continuity) holds in ri dom$f$.
- When $x \to$ribd dom$f$, continuity may disappear (e.g. f may be infinity/jump).
- Closing epi$f$ only miss function values with little interest.
- whether f can be assumed upper semi-continuous? This property holds for 1D functions. But, not in general, e.g.

$$f(x) = \sup_{\alpha,\beta}\{\xi\alpha + \eta\beta : \alpha^2 \leqslant 2\beta\}.$$

$f(0) = 0$ and $f \in$ Conv$\mathbb{R}^n$. In fact, "$=$" hold at optimum (i.e. active constraint $\frac{1}{2}\alpha^2 = \beta$), so

$$f(x) = f(\xi,\eta) = \begin{cases} 0, & \xi = \eta = 0 \\ -\frac{\xi^2}{2\eta} & \eta < 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Thus, as $x \to 0$ along path $\eta = -\frac{1}{2}\xi^2$, then $f(x) \equiv 1 > 0 = f(1)$.

Theorem

Let Conv$\mathbb{R}^n \ni \{f_k\}_{k=1}^{\infty} \xrightarrow{pointwisely} f : \mathbb{R}^n \to \mathbb{R}$. Then, $f \in$ Conv$\mathbb{R}^n$ and $f_k \xrightarrow{uniformly} f$ on any compact set S.

? what is the behaviour of f as $x \to \infty$?

- Let $f \in \overline{\text{Conv}}\mathbb{R}^n$. The asymptotic cone is
$$(\text{epi}f)_\infty = \{(d,p) : \text{epi}f + t(d,p) \subset \text{epi}f,\ \forall t > 0\}$$
$$= \{(d,p) : \text{epi}f + (d,p) \subset \text{epi}f\} = \text{epi}f \underline{*} \text{epi}f = \text{epi}f \bar{\nabla} \text{epi}f$$
is closed convex cone containing half-line $\{0\} \times \mathbb{R}^+$.

Proposition

Let $f \in$ Conv$\mathbb{R}^n$ and any $x_0 \in$ dom$f$. Then, $(\text{epi}f)_\infty$ is the epigraph of $f'_\infty \in$ Conv$\mathbb{R}^n$ (called asymptotic/recession/$\cdots$ function of f), where

$$f'_\infty(d) := \sup_{t>0} \frac{f(x_0+td) - f(x_0)}{t} = \lim_{t\to\infty} \frac{f(x_0+td) - f(x_0)}{t}.$$

proof $\because\ (x_0, f(x_0)) \in \text{epi}f$, $\therefore\ (d,\rho) \in (\text{epi}f)_\infty$

$\therefore\ f(x_0+td) \le f(x_0) + t\rho,\ \forall t > 0$. Thus, $\sup_{t>0} \frac{f(x_0+td)-f(x_0)}{t} \leqslant \rho$.

- f'_∞ is positively homogeneous.

e.g. Let C be closed convex. Then, $\iota_C(x_0+td)=0$ for all $t>0$ iff $d\in C_\infty$. Thus, $(\iota_C)'_\infty=\iota_{(C_\infty)}$.

e.g. Let $f\in \mathrm{Conv}\mathbb{R}^n$. For $x_0\in \mathrm{dom} f$, $\varphi(d)=f(x_0+d)-f(x_0)$ is convex. Thus, $0\in \mathrm{dom}\varphi$ and its perspective function is r. The cl$r$ can be computed as

$$(\mathrm{cl}r)(0,d)=\lim_{\alpha\downarrow 0}\alpha[f(x_0+d'-d+d/\alpha)-f(x_0)]$$
$$=\lim_{t\to+\infty}\frac{f(x_0+d'+td)-f(x_0+d')}{t}=f'_\infty(d).$$

Proposition

Let $f\in \mathrm{Conv}\mathbb{R}^n$. If $S_r(f)\neq\emptyset$, then $[S_r(f)]_\infty=\{d\in\mathbb{R}^n: f'_\infty(d)\leqslant 0\}$. Particularly, the following statements are equivalent

1. *There is r for which $S_r(f)$ is nonempty compact;*
2. *all sublevel-sets of f are compact;*
3. *$f'_\infty(d)>0$ for all $0\neq d\in\mathbb{R}^n$.*

proof d is in asymptotic cone of $\emptyset \neq S_r(f)$ iff

$$x \in S_r(f) \implies [x + td \in S_r(f),\ \forall t > 0],$$

which amounts to $(x, r) \in \text{epi} f \implies (x + td, r + tx_0) \in \text{epi} f,\ \forall t > 0.$

It means $(d, 0) \in (\text{epi} f)_\infty = \text{epi} f'_\infty$, which proves ❶.

Particularly, if $S_0(f'_\infty)$ is reduced to singleton $\{0\}$, which implies ❸. It amounts to $[S_r(f)]_\infty = \{0\}$ for all $r \in \mathbb{R}$ with $S_r(f) \neq \emptyset$, which means $S_r(f)$ is compact.

- Convexity of f ensure that all $\emptyset \neq S_r f$ have the same asymptotic cone.
- If f is closed quasi-convex, then all $S_r(f)$ are convex, and as such they have asymptotic cones, which normally depend on the level.

Definition (0-coercivity: "increase at infinity")

$f \in \text{Conv}\mathbb{R}^n$ satisfying ❶ or ❷ or ❸ are called 0-coercive, i.e.,

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

The closed convex 0-coercive functions has minimum over $\mathbb{R}^n$.

Definition (1-coercivity)

$f'_\infty(d) = +\infty$ for all $d \neq 0$ (i.e. $f'_\infty = \iota_{\{0\}}$) $\Longleftrightarrow$ $\lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|} = +\infty$.

► e.g. $f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle b, x\rangle + c$ with $Q \succeq 0$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then

$$f'_\infty(d) = \begin{cases} \langle b, d\rangle & \text{if } d \in \ker Q, \\ +\infty, & \text{if not.} \end{cases}$$

In this case, f is 0-coercive $\Longleftrightarrow$ f is 1-coercive $\Longleftrightarrow$ $Q \succ 0$.

Proposition

$f \in \overline{\text{Conv}}\mathbb{R}^n$ is Lipschitzian on $\mathbb{R}^n$ iff f'_∞ is finite on $\mathbb{R}^n$. Furthermore, the Lipschitz constant of f is $\sup\{f'_\infty(d) : \|d\| = 1\}$.

Proposition

Let $f_i \in \overline{\text{Conv}}\mathbb{R}^n$ *and* $t_i > 0$ $(i = 1, \dots, m)$*. Assume that* $\exists x_0$ *such that all* f_i*'s is finite.*

$$\text{If } f := \sum_{i=1}^{m} t_i f_i, \text{ then } f'_\infty = \sum_{i=1}^{m} t_i (f'_i)_\infty.$$

Let $\{f_j\}_{j\in J}$ *be in* $\overline{\text{Conv}}\mathbb{R}^n$*, Assume that* $\exists x_0$ *such that* $\sup_{j\in J} f_j < +\infty$*.*

$$\text{If } f := \sum_{j\in J} t_j f_j, \text{ then } f'_\infty = \sup_{j\in J} (f'_j)_\infty.$$

Let $A : \mathbb{R}^n \to \mathbb{R}^m$ *be affine with linear part* A_0*, and* $f \in \overline{\text{Conv}}\mathbb{R}^m$*. Assume that* $A(\mathbb{R}^n) \cap \text{dom} f \neq \emptyset$*. Then* $(f \circ A)'_\infty = f'_\infty \circ A_0$*.*

First- and Second-Order Differentiation

Let $\emptyset \neq C \subset \mathbb{R}^n$ be convex and $f : C \to \mathbb{R} \cup \{+\infty\}$ ($f(x) < +\infty$ for all $x \in C$).

- **?** When f is convex differentiable?
- **?** If f is differentiable, can we characterize its convexity by of ∇f?
- **?** If f is convex, what is the behaviors of first- and second-differentiability?

Theorem

Let f be differentiable on open set $\Omega \subset \mathbb{R}^n$ and $C \subset \Omega$ be convex. Then

1. *f is convex on C iff*

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle, \ \forall (x, x_0) \in C \times C;$$

2. *f is strictly convex on C iff "$>$" holds in above inequality whenever $x \neq x_0$;*
3. *f is c-strongly convex on C iff*

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2} c \|x - x_0\|^2, \ \forall (x_0, x) \in C \times C.$$

proof ❶ $\because$ $f \in \mathrm{Conv}C$, $\therefore$ $\forall (x_0, x) \in C \times C$ and $\alpha \in [0,1]$, we have from the definition of convexity

$$f(\alpha x + (1-\alpha)x_0) - f(x_0) \le \alpha[f(x) - f(x_0)].$$

Divide by α and let $\alpha \downarrow 0$, the left-hand side $\to \langle \nabla f(x_0), x - x_0 \rangle$.
Conversely, take $x_1 \in C$, $x_2 \in C$, $\alpha \in [0,1]$, and set $x_0 := \alpha x_1 + (1-\alpha)x_2 \in C$. By assumption,

$$f(x_i) \ge f(x_0) + \langle \nabla f(x_0), x_i - x_0 \rangle \quad \text{for } i = 1, 2.$$

By convex combination the above two inequalities for $i = 1, 2$, we obtain

$$\alpha f(x_1) + (1-\alpha) f(x_2) \ge f(x_0) + \langle \nabla f(x_0), \alpha x_1 + (1-\alpha)x_2 - x_0 \rangle = f(x_0),$$

which implies f is convex.

proof ❷ $\because$ f is strictly convex, $\therefore \forall x_0 \neq x$ and $\alpha \in [0,1]$, we have

$$f(x_0 + \alpha(x - x_0)) - f(x_0) < \alpha[f(x) - f(x_0)];$$

but f is convex and we can use

$$\langle \nabla f(x_0), \alpha(x - x_0) \rangle \le f(x_0 + \alpha(x - x_0)) - f(x_0),$$

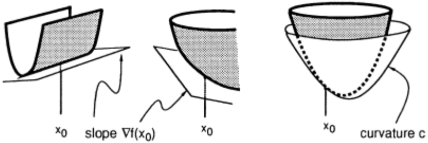
so "¡" holds.

proof ❸ apply to $f - \frac{1}{2}c\|\cdot\|^2$.

- if f is differentiable convex, then $\forall x_0$ (tangent points), f is minorized by its first-order approximation $f(x_0) + \langle \nabla f(x_0), \cdot - x_0 \rangle$.
- f is strictly convex if tangent point is the singleton.
- f is strongly convex if it is minorized by quadratic convex function

$$q(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{c}{2}\|x - x_0\|^2,$$

whose gradient at x_0 is $\nabla f(x_0)$.



- if $f \in \mathrm{Conv}\mathbb{R}^n$, the remainder in development $r(x_0, \cdot) = f(\cdot) - f(x_0) - \langle \nabla f(x_0), \cdot - x_0 \rangle$ is nonnegative convex.

Definition

Let $C \subset \mathbb{R}^n$ be convex. The mapping $F : C \to \mathbb{R}^n$ is said to be

monotone if: $\langle F(x) - F(x'), x - x' \rangle \geq 0, \ \forall x, x' \in C$

strictly monotone if: $\langle F(x) - F(x'), x - x' \rangle > 0, \ \forall x, x' \in C$

c-strongly monotone if: $\langle F(x) - F(x'), x - x' \rangle > 0 \geq c\|x - x_0\|^2, \ \forall x, x' \in C$

- "monotone" function reduces to "increasing" function in 1D case.

Theorem

Let f be differentiable on open set $\Omega \subset \mathbb{R}^n$ and $C \subset \Omega$ be convex. Then, f is [resp. strictly, c-strongly] convex on C iff ∇f is [resp. strictly, c-strongly] on C.

proof If f is c-strongly convex on C, then $\forall x \in C$, $x_0 \in C$,

$$f(x) \geqslant f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{c}{2}\|x - x_0\|^2,$$
$$f(x_0) \geqslant f(x) + \langle \nabla f(x), x_0 - x \rangle + \frac{c}{2}\|x - x_0\|^2.$$

By adding both inequalities, strongly monotone of ∇f holds.
Conversely, let $(x_0, x_1) \in C \times C$ and $\varphi(t) = f(x_t)$ with $x_t = x_0 + t(x_1 - x_0)$.
φ is differentiable, and $\varphi'(t) = \langle \nabla f(x_t), x_1 - x_0 \rangle$. Thus,
$\varphi'(t) - \varphi'(t') = \langle \nabla f(x_t) - \nabla f(x_{t'}), x_1 - x_0 \rangle = \frac{1}{t-t'}\langle \nabla f(x_t) - \nabla f(x_{t'}), x_t - x_{t'} \rangle$.
∇f is monotone $\Longrightarrow$ φ' is increasing $\Longrightarrow$ φ is convex.
By setting $t' = 0$ in above inequality and use the strong monotonicity of ∇f, we have $\varphi'(t) - \varphi'(0) \geq \frac{c}{t}\|x_t - x_0\|^2 = tc\|x_1 - x_0\|^2$.
$\because$ differentiable convex function is the integral of its derivative,
$\therefore$ $\varphi(1) - \varphi(0) - \varphi'(0) = \int_0^1 [\varphi'(t) - \varphi'(0)]dt \geq \frac{c}{2}\|x_1 - x_0\|^2$.

▶e.g. If $f(x) = \langle Ax, x\rangle + \langle b, x\rangle$ with $A \succeq 0$, then
$\langle \nabla f(x) - \nabla f(x'), x - x'\rangle \geq \lambda_n \|x - x'\|^2$.
$\therefore$ ∇f is λ_n-strongly monotone. Moreover,
$f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0\rangle = \frac{1}{2}\langle A(x - x_0), x - x_0\rangle \geq \frac{1}{2}\lambda_n \|x - x'\|^2$.

- strong and strict convexity are equivalent if $A \succ 0$.
- monotone mapping $\not\Leftrightarrow$ gradient mapping. Indeed, if $F : \Omega \to \mathbb{R}^n$ is differentiable, then F is a gradient iff its Jacobian is symmetric.

Proposition

Let $f \in \text{Conv}\mathbb{R}^n$ and $x \in \text{int dom} f$. Then, the followings are equivalent:

1. $q(d) := \lim_{t\downarrow 0} \frac{f(x+td)-f(x)}{t}$ *is linear in d;*
2. *for some basis of $\mathbb{R}^n$ in which $x = (\xi^1, \ldots, \xi^n)$, the $\frac{\partial f}{\partial \xi^i}(x)$ exist at x for $i = 1, \ldots, n$;*
3. *f is differentiable at x.*

- let $\varphi_d(t) = f(x + td)$. If $\exists$ n independent directions $\{d_i\}_{i=1}^n$ such that φ_{d_i} has derivative at $t = 0$, then the same property holds for all $d \in \mathbb{R}^n$.
- "radial" differentiability of φ_d suffices to guarantee "global" (i.e. Fréchet) differentiability of f at x.

- If f is convex differentiable in $\mathcal{B}(x)$, then ∇f is continuous at x. Hence, if Ω is open convex and f is convex, then f differentiable on $\Omega \iff f \in C^1(\Omega)$.
- The largest set on which f can be differentiable is interior of its domain.
- If f is locally Lipschitzian on an open set Ω, then f is a.e. differentiable on Ω.

Theorem

Let $f \in \mathrm{Conv}\mathbb{R}^n$. The subset of $\mathrm{intdom} f$ where f is non-differentiable is of zero (Lebesgue) measure.

Theorem

Let f be twice differentiable on an open convex set Ω. Then

1. *f is convex on $\Omega \iff \nabla^2 f(x_0) \succeq 0$ for all $x_0 \in \Omega$;*
2. *if $\nabla^2 f(x_0) \succ 0$ for all $x_0 \in \Omega \implies f$ is strictly convex on Ω;*
3. *f is c-strongly convex on $\Omega \iff \lambda_{\min}[\nabla^2 f(x_0)]$ is minorized by c on Ω, i.e., $\langle \nabla^2 f(x_0)d, d\rangle \geq c\|d\|^2,\ \forall x_0 \in \Omega,\ d \in \mathbb{R}^n$.*

proof For $x_0 \in \Omega$, $d \in \mathbb{R}^n$ and $t \in \mathbb{R}$ such that $x_0 + td \in \Omega$, let $x_t := x_0 + td$ and $\varphi(t) := f(x_t) = f(x + td)$. $\therefore\ \varphi''(t) = \langle \nabla^2 f(x_t)d, d \rangle$.

1. If f is convex on Ω, then $\forall (x_0, d) \in \Omega \times \mathbb{R}^n$ with $d \neq 0$, φ is convex on open interval $I := \{t \in \mathbb{R} : x_0 + td \in \Omega\}$.

   $\therefore\ 0 \leq \varphi''(t) = \langle \nabla^2 f(x_t)d, d \rangle,\ \forall t \in I \ni 0$ and $\nabla^2 f(x_0) \succeq 0$.

   Conversely, take any $[x_0, x_1] \subset \Omega$, set $d = x_1 - x_0$ and assume $\nabla^2 f(x_t) \succeq 0$, i.e., $\varphi''(t) \geq 0$ for $t \in [0, 1]$. Then φ is convex on $[0, 1]$, i.e. f is convex on $[x_0, x_1]$. The result follows since x_0 and x_1 are arbitrary in Ω.

2. Take any $[x_0, x_1] \subset \Omega$ with $x_1 \neq x_0$ and $d = x_1 - x_0$. By mean-value theorem to φ' on [0,1]: $\exists \tau \in]0, 1[$, $\varphi'(1) - \varphi'(0) = \varphi''(\tau) = \langle \nabla^2 f(x_\tau)d, d \rangle > 0$.

3. By applying 1 to $f - \frac{c}{2}\|\cdot\|^2$, whose Hessian is $\nabla^2 f - cI_n$ and eigenvalues is $\lambda(\nabla^2 f) - c$.

- 2 is sufficient condition but not necessary. e.g. $f(x) = \frac{1}{4}x^4$
- Affine approximation of f around x_0 is actually a global minorization. But, quadratic approximation of f around x_0 can not minorizes f. E.g. $f(x) = \frac{1}{2}x^2 - \frac{1}{4}x^4$ is convex on $\{x : |x|^2 \leq \frac{1}{3}\}$.
- C must be open. A claim "f is convex on $C \subset \Omega$ iff $\nabla^2 f(\cdot) \succeq 0$ on C" is wrong when C is not open. e.g. $f(\xi, \eta) := \xi^2 - \eta^2$ is convex on $C = \mathbb{R} \times \{0\}$ but $\nabla^2 f \not\succeq 0$.

- To check convexity of f on non-open set C, we first use $\nabla^2 f$ on $\operatorname{int} C$ and then pass to the limit: property $C \subset \overline{C} \subset \operatorname{int} C$.

e.g. Let $\Omega = \{x = (\xi^1, \ldots, \xi^n) : x > 0\}$ and $f(x) = -(\xi^1 \xi^2 \cdots \xi^n)^{1/n}$ on Ω.

Then, $\frac{\partial^2 f}{\partial \xi^i \partial \xi^j}(x) = \frac{f(x)}{n^2 \xi^i \xi^j}(1 - n\delta_{ij})$, where δ_{ij} is Kronecker's symbol. Thus,

$$\langle \nabla^2 f(x) d, d \rangle = \frac{f(x)}{n^2} \left[\left(\sum_{i=1}^{n} \frac{d^i}{\xi^i} \right)^2 - n \sum_{i=1}^{n} \left(\frac{d^i}{\xi^i} \right)^2 \right], \ \forall d = (d^1, \ldots, d^n) \in \mathbb{R}^n.$$

$\because$ $\|\cdot\|_1 \leq \sqrt{n}\|\cdot\|_2$ and $f < 0$ on Ω, the above expression is $\geqslant 0$. Thus, f is convex. Furthermore, f is positively homogeneous.

- Flat Domain. If $\operatorname{dom} f$ is not full-dimensional, by defining $f_0(y) := f(x_0 + y)$ with $x_0 \in \operatorname{dom} f$ and y varies in subspace V parallel to $\operatorname{affdom} f$.

  $f_0 \in \operatorname{Conv} V$ and $\operatorname{dom} f_0$ is full-dimensional in V. Equipping V with $\langle \cdot, \cdot \rangle$ and Lebesgue measure, the results above can be reproduced, i.e.,

  a.e. $x_0 + y \in \operatorname{intdom} f$, $\exists s \in V$ (i.e., $\nabla f_0(y)$)

$$\forall h \in V, \quad f(x_0 + y + h) = f(x_0 + y) + \langle s, h \rangle + o(\|h\|_V)$$

  (the remainder $o(\|h\|_V)$ is nonnegative).

  s is "relative gradient" of f at $x = x_0 + y$; it exists at $x_0 + y$ iff $t \mapsto f(x_0 + y + td)$ has a derivative at $t = 0$ for all $d \in V$.