Epigraphs, Infimal convolution & Moreau-Yosida envelope

Andersen Ang

ECS, Uni. Southampton, UK

andersen.ang@soton.ac.uk

Homepage angms.science

Version: April 1, 2023

First draft: January 13, 2023

Content

Table of Contents

$$\operatorname{epi} f := \Big\{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \,:\, \alpha \geq f(\boldsymbol{x})\Big\}.$$

Graph and epigraph

- **(Graph of abstract function)** Given two sets X (domain) and Y (codomain), the graph of the function $f : X \to Y$ is the set

$$\mathrm{graph} f := \Big\{(\boldsymbol{x}, \boldsymbol{y}) \in X \times Y \;:\; \boldsymbol{y} = f(\boldsymbol{x})\Big\}.$$

- **(Epigraph of function from $\mathbb{R}^n$ to $\overline{\mathbb{R}}$)** Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ be the extended reals. Now consider $X = \mathbb{R}^n$ and $Y = \overline{\mathbb{R}}$. For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the epigraph of f is the set

$$\mathrm{epi}\, f := \Big\{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \;:\; \alpha \geq f(\boldsymbol{x})\Big\}.$$

- **(Strict epigraph)** $\mathrm{epi}_S f = \mathrm{epi}\, f \setminus \mathrm{graph} f$, or equivalently

$$\mathrm{epi}_S f := \Big\{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \;:\; \alpha > f(\boldsymbol{x})\Big\}.$$

Remarks

$$\operatorname{epi} f := \left\{ (\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \,:\, \alpha \geq f(\boldsymbol{x}) \right\}$$

- $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ but $\operatorname{epi} f$ is defined to be a subset of $\mathbb{R}^n \times \mathbb{R}$, not $\mathbb{R}^n \times \overline{\mathbb{R}}$.

- This is intentional to define $\operatorname{epi} f$ as a subset of $\mathbb{R}^n \times \mathbb{R}$.
  - $\mathbb{R}^n$ is a vector space
  - $\mathbb{R}^n \times \mathbb{R}$ is a vector space
  - $\mathbb{R}^n \times \overline{\mathbb{R}}$ is not a vector space: $\nexists$ additive identity for $\infty + \infty$

  (Being a vector space allows to use tools from real analysis and functional analysis.)

- **(At infinity)** If $f(\boldsymbol{x}_0) = +\infty$ at $\boldsymbol{x} = \boldsymbol{x}_0$, then $(\boldsymbol{x}_0, +\infty) \notin \operatorname{epi} f$.

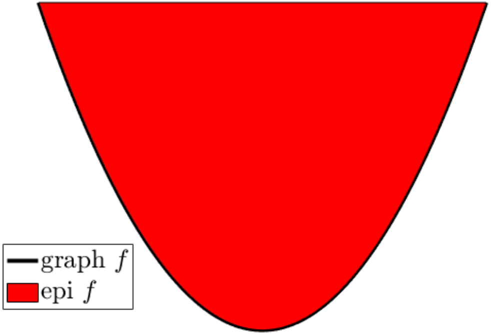
- **Two extreme cases**
  - If $f_{\infty^+}(\boldsymbol{x}) = +\infty \ \forall \boldsymbol{x}$, then $\operatorname{epi} f_{\infty^+}$ is $\varnothing$. **Empty epigraph**
  - If $f_{\infty^-}(\boldsymbol{x}) = -\infty \ \forall \boldsymbol{x}$, then $\operatorname{epi} f_{\infty^-}$ is the whole $\mathbb{R}^n \times \mathbb{R}$ **Whole space epigraph**

  (Convention: $-\infty < +\infty$)

Visualization of $\operatorname{graph} f$ and $\operatorname{epi} f$

- $\operatorname{epi} f$ = **all** the points of $\mathbb{R}^{n+1}$ lying on or above $\operatorname{graph} f$.

- Example: $f(x) = x^2$
  - $n = 1$ (1-dimensional)
  - $\operatorname{graph} f := \Big\{(x, y) \in \mathbb{R} \times \mathbb{R} \, : \, y = f(x)\Big\}$ is a 1d curve in a 2d space.
  - $\operatorname{epi} f := \Big\{(x, \alpha) \in \mathbb{R} \times \mathbb{R} \, : \, \alpha \geq f(x)\Big\}$ is a 2d set in a 2d space.



Level sets: a concept related to epigraph.

- For $f : \mathbb{R}^n \to \mathbb{R}$ and $\alpha \in \mathbb{R}$, define

$$\begin{aligned}
\mathrm{lev}_{\le\alpha} f &:= \big\{ \boldsymbol{x} \in \mathbb{R}^n \,:\, f(\boldsymbol{x}) \le \alpha \big\} \\
\mathrm{lev}_{<\alpha} f &:= \big\{ \boldsymbol{x} \in \mathbb{R}^n \,:\, f(\boldsymbol{x}) < \alpha \big\} \\
\mathrm{lev}_{=\alpha} f &:= \big\{ \boldsymbol{x} \in \mathbb{R}^n \,:\, f(\boldsymbol{x}) = \alpha \big\} \\
\mathrm{lev}_{>\alpha} f &:= \big\{ \boldsymbol{x} \in \mathbb{R}^n \,:\, f(\boldsymbol{x}) > \alpha \big\} \\
\mathrm{lev}_{\ge\alpha} f &:= \big\{ \boldsymbol{x} \in \mathbb{R}^n \,:\, f(\boldsymbol{x}) \ge \alpha \big\}
\end{aligned}$$

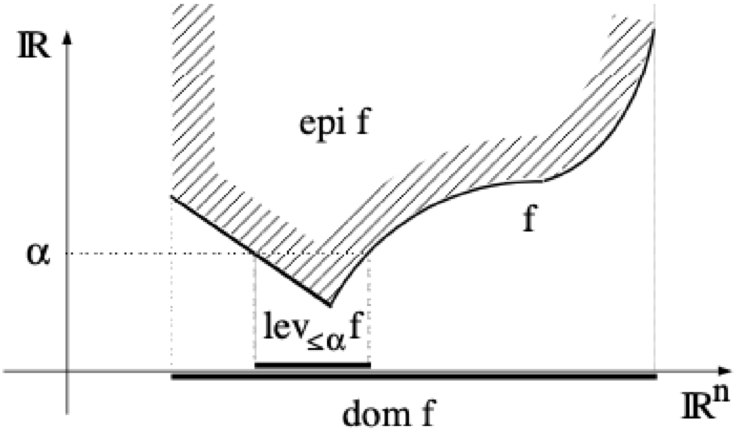
  The important one for minimization is $\mathrm{lev}_{\le\alpha} f$ (named sublevel sets).

- lev is a subset of domain, not codomain.

- If $\alpha = \inf f$, then $\mathrm{lev}_{\le\alpha} f = \mathrm{lev}_{=\alpha} f = \mathrm{argmin}\ f$.

- Level sets can be
  - empty: $\mathrm{lev}_{\le -1}(x^2)$, no x makes $x^2 \le -1$
  - non-continuous: $\mathrm{lev}_{=0} \sin(x)$, then $x = \{n\pi\}_{n\in\mathbb{N}}$ is a set of dots
  - non-convex: $\mathrm{lev}_{=0} \sin(x)$

An illustration of $\mathrm{lev}_{\leq\alpha} f$



Picture from Rockafellar, R. Tyrrell, and Roger J-B. Wets. Variational analysis. Springer, 2009.

Optimization newbie: "Why talk about epigraph?"

Optimization expert: "it is useful!"

The main idea

- Many properties of f has a counterpart in $\operatorname{epi} f$.
- Sometimes it is easier to work with $\operatorname{epi} f$ than with f.

Properties of f — Properties of $\operatorname{epi} f$

Properties of f	Properties of $\operatorname{epi} f$
f is proper	$\operatorname{epi} f$ is nonempty
f is closed	$\operatorname{epi} f$ is a closed set
f is lower semicontinuous on $\mathbb{R}^n$	$\operatorname{epi} f$ is nonempty and closed in $\mathbb{R}^n \times \mathbb{R}$
f is convex	$\operatorname{epi} f$ is a convex set
f is strictly convex	$\operatorname{epi}_S f$ is a convex set
$f \square g$ (Infimal convolution of f and g)	Minkowski sum of $\operatorname{epi} f$ and $\operatorname{epi} g$

Example: f is convex $\iff$ $\operatorname{epi} f$ is convex

(To avoid confusing the bracket of graph and bracket of a number, here we use $\{\boldsymbol{x}_1, y_1\} \in \operatorname{epi} f$)

- ($\implies$)
  - f is convex $\iff$ $f\big((1-t)\boldsymbol{x}_1 + t\boldsymbol{x}_2\big) \leq (1-t)f(\boldsymbol{x}_1) + tf(\boldsymbol{x}_2)\ \forall \boldsymbol{x}_1, \boldsymbol{x}_2$ and $t \in [0,1]$. $(*)$
  - $\{\boldsymbol{x}_1, y_1\} \in \operatorname{epi} f \iff f(\boldsymbol{x}_1) \leq y_1$ and $\{\boldsymbol{x}_2, y_2\} \in \operatorname{epi} f \iff f(\boldsymbol{x}_2) \leq y_2$ $(**)$
  - Now $f\Big((1-t)\boldsymbol{x}_1 + t\boldsymbol{x}_2\Big) \overset{(*)}{\leq} (1-t)f(\boldsymbol{x}_1) + tf(\boldsymbol{x}_2) \overset{(**)}{\leq} (1-t)y_1 + ty_2$. So $\Big\{(1-t)\boldsymbol{x}_1 + t\boldsymbol{x}_2,\ (1-t)y_1 + ty_2\Big\} = (1-t)\Big\{\boldsymbol{x}_1,\ y_1\Big\} + t\Big\{\boldsymbol{x}_2,\ y_2\Big\} \overset{(**)}{\in} \operatorname{epi} f$. So $\operatorname{epi} f$ is a convex set.

- ($\impliedby$)
  - $\operatorname{epi} f$ is a convex set $\iff$ $\{\boldsymbol{x}_1, y_1\} \in \operatorname{epi} f$, $\{\boldsymbol{x}_2, y_2\} \in \operatorname{epi} f$ implies for $t \in [0,1]$ we have

$$(1-t)\Big\{\boldsymbol{x}_1,\ y_1\Big\} + t\Big\{\boldsymbol{x}_2,\ y_2\Big\} = \Big\{(1-t)\boldsymbol{x}_1 + t\boldsymbol{x}_2,\ (1-t)y_1 + ty_2\Big\} \in \operatorname{epi} f. \qquad (***)$$

  - By definition of epigraph, $(***)$ is equivalent to

$$f(\boldsymbol{x}_1) \leq y_1 \text{ and } f(\boldsymbol{x}_2) \leq y_2 \quad \text{implies} \quad f\Big((1-t)\boldsymbol{x}_1 + t\boldsymbol{x}_2\Big) \leq (1-t)y_1 + ty_2.$$

    Choose $f(\boldsymbol{x}_1) \leq y_1$ and $f(\boldsymbol{x}_2) \leq y_2$ gives $f\Big((1-t)\boldsymbol{x}_1 + t\boldsymbol{x}_2\Big) \leq (1-t)f(\boldsymbol{x}_1) + tf(\boldsymbol{x}_2)$, so f is convex.

Table of Contents

$$(f_1 \square f_2)(\boldsymbol{x}) \coloneqq \inf_{\boldsymbol{x}_1 + \boldsymbol{x}_2 = \boldsymbol{x}} f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2)$$

Infimal convolution

- Given $f_1 : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$. The infimal convolution $f_1 \Box f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ is

$$\begin{aligned}(f_1 \Box f_2)(\boldsymbol{x}) &:= \inf_{\boldsymbol{x}_1+\boldsymbol{x}_2=\boldsymbol{x}} f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) \\ &= \inf_{\boldsymbol{x}_1} f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}-\boldsymbol{x}_1) \\ &= \inf_{\boldsymbol{x}_2} f_1(\boldsymbol{x}-\boldsymbol{x}_2) + f_2(\boldsymbol{x}_2)\end{aligned}$$

  Convention: $\infty - \infty = \infty$ and $\inf \varnothing = +\infty$

- History

  Earliest(?) work
  Hausdorff, Uber halbstetige Funktionen und deren Verallgemeinerung, Math. Zeit. 5 (1919), 292-309.

  Fenchel, “Convex Cones, Sets, and Functions”, Lecture Notes, Princeton University, Princeton, 1953.

  First systematic study of infimal convolution
  Moreau, Inf-convolution, Sém. d'Math. Montpellier (1963), 3.1-3.48

  Later works by Attouch, Rockafellar, Hiriart-Urruty, etc

  Thomas Stromberg's PhD thesis (1994): a nice summary.

How infimal convolution gets its name

$$(f_1 \Box f_2)(\boldsymbol{x}) \coloneqq \inf_{\boldsymbol{x}_1} f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x} - \boldsymbol{x}_2).$$

- Definition of (integral) convolution (Examples: Laplace transform, Fourier transform.)

$$(f * g)(t) \coloneqq \int_{\infty}^{\infty} f(\tau) g(t - \tau) d\tau.$$

- $f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x} - \boldsymbol{x}_2)$ "looks similar" to (integral) convolution
- There is $\inf\limits_{\boldsymbol{w}} \implies$ people name it infimal.
- Deep fact: integral convolution is in $(+, \times)$-algebra
  - Integration $=$ summation
  - You combine f and g by multiplication

  Infimal convolution is in $(\min, +)$-algebra (tropical semi-ring)
  - The summation is replaced by $\min$
  - You combine f and g by addition

What infimal convolution solves: an economics example

- You want to buy totally n hamburgers, from `MacDonald` and `Burger King`.
  Suppose buying n_1 hamburgers from `MacDonald` costs you $f(n_1)$, and if you buy n_2 hamburgers from `Burger King`, the price is $g(n_2)$.

- You want to find the infimum of the total cost $f(n_1) + g(n_2)$ subject to the constraint $n_1 + n_2 = n$. I.e., you want to find the "cheapest way" to buy n hamburgers.

- This problem is exactly: calculate $(f \square g)(n)$

$$(f \square g)(n) = \inf_{n_1 + n_2 = n} f(n_1) + g(n_2) = \underbrace{\underbrace{\inf_{n_1} f(n_1) + g(n - n_1)}_{\text{focus on } n_1} = \underbrace{\inf_{n_2} f(n - n_2) + g(n_2)}_{\text{focus on } n_2}}_{\text{this also means infimal convolution is commutative}}.$$

Infimal convolution is commutative: $f \square g = g \square f$.

Example. Infimal convolution of two indicator functions

$$
\begin{aligned}
(f_1 \Box f_2)(\boldsymbol{x}) &:= \inf_{\boldsymbol{x}_1+\boldsymbol{x}_2=\boldsymbol{x}} f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) \\
&= \inf_{\boldsymbol{x}_1} \quad f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x} - \boldsymbol{x}_1) \\
&= \inf_{\boldsymbol{x}_2} \quad f_1(\boldsymbol{x} - \boldsymbol{x}_2) + f_2(\boldsymbol{x}_2)
\end{aligned}
$$

- Given two sets C_1, C_2 and two indicator functions i_{C_1}, i_{C_2}.

$$
(i_{C_1} \Box i_{C_2})(\boldsymbol{x}) = \inf_{\boldsymbol{x}_1+\boldsymbol{x}_2=\boldsymbol{x}} i_{C_1}(\boldsymbol{x}_1) + i_{C_2}(\boldsymbol{x}_2) = i_{C_1 \oplus C_2}(\boldsymbol{x})
$$

  $\oplus$ is Minkowski sum of sets: $P \oplus Q := \big\{ p + q \mid p \in P, q \in Q \big\}$.

- Minkowski sum keeps convexity of sets, so $C_1 \oplus C_2$ is a convex set and $i_{C_1 \oplus C_2}$ is a convex function. Here we see that inf-convolution of two convex functions is a convex function.

- In general, if $f_1 : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ are convex, then $f_1 \Box f_2$ is also convex.
  Proof. We prove inf-convolution preserves convexity using definition and operations that preserve convexity. By definition, $(f_1 \Box f_2)(x) = \inf_{x_1} h(x, x_1)$ where $h(x, x_1) = f_1(x_1) + f_2(x - x_1)$. By assumption $f_1(x_1)$ is convex and $f_2(x_1)$ is convex. The function $f_2(x - x_1)$ is $f_2(x_1)$ with argument x_1 under a translation to $x - x_1$ so $f_2(x - x_1)$ is convex. Now $h(x, x_1)$ is the sum of two convex functions on x_1, thus it is convex.

Infimal convolution is also called epi-addition

$$(f_1 \Box f_2)(\boldsymbol{x}) \ := \ \inf_{\boldsymbol{x}_2} f_1(\boldsymbol{x} - \boldsymbol{x}_2) + f_2(\boldsymbol{x}_2).$$

- $\operatorname{epi}_S(f_1 \Box f_2) = \operatorname{epi}_S f_1 \oplus \operatorname{epi}_S f_2$ $\iff$ inf-convolution of convex functions is convex[1]
- $\operatorname{epi}(f_1 \Box f_2) \supseteq \operatorname{epi} f_1 \oplus \operatorname{epi} f_2$
- $\operatorname{epi}(f_1 \Box f_2) = \operatorname{epi} f_1 \oplus \operatorname{epi} f_2$ if inf-convolution is exact
  Exact means the $\inf$ is gone: $(f_1 \Box f_2)(\boldsymbol{x}) = f_1(\boldsymbol{x} - \boldsymbol{x}_2) + f_2(\boldsymbol{x}_2)$.
- For proof, see Jean Jacques Moreau. Inf-convolution, sous-additivité, convexité des fonctions numériques. Journal de Mathématiques Pures et Appliquées, 1970. `https://hal.archives-ouvertes.fr/hal-02162006`

[1] Remark 2.3.3 in Urruty, Jean-Baptiste Hiriart, and Claude Lemaréchal. Convex analysis and minimization algorithms. Springer-Verlag, 1993

The proof of $\text{epi}\,(f_1 \Box f_2) \supseteq \text{epi}\, f_1 \oplus \text{epi}\, f_2$

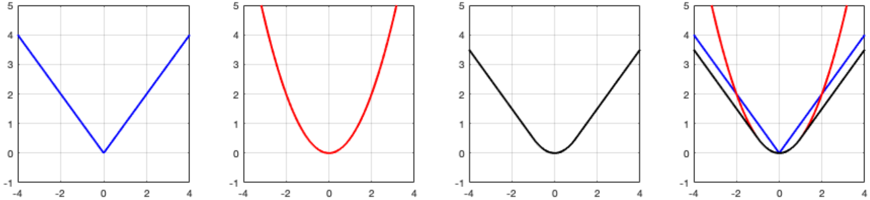
- Take $\{\boldsymbol{x}, \alpha\} \in \text{epi}\, f_1 \oplus \text{epi}\, f_2$. Since the element $\{\boldsymbol{x}, \alpha\}$ is inside the sum of two sets ($\text{epi}\, f_1$ and $\text{epi}\, f_2$), that means we can decompose $\{\boldsymbol{x}, \alpha\}$ as the sum of element from each set. I.e., we have $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$ and $\alpha = \alpha_1 + \alpha_2$ that $\{\boldsymbol{x}_1, \alpha_1\} \in \text{epi}\, f_1$ and $\{\boldsymbol{x}_2, \alpha_2\} \in \text{epi}\, f_2$.
  (It means given $\{\boldsymbol{x}, \alpha\}$, there is exist the pair $\{\boldsymbol{x}_1, \alpha_1\}$, $\{\boldsymbol{x}_2, \alpha_2\}$ that fulfill the above conditions)
- implies $f_1(\boldsymbol{x}_1) \leq \alpha_1$, implies $f_2(\boldsymbol{x}_2) \leq \alpha_2$ and $f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) \leq \alpha_1 + \alpha_2 = \alpha$.
- Now consider $(f_1 \Box f_2)(\boldsymbol{x}) := \inf\limits_{\boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{x}} f_1(\boldsymbol{y}_1) + f_2(\boldsymbol{y}_2)$. As $\boldsymbol{x}_1 + \boldsymbol{x}_2 = \boldsymbol{x} = \boldsymbol{y}_1 + \boldsymbol{y}_2$, the infimum $\inf\limits_{\boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{x}}$ is the smallest among all pair that sum to $\boldsymbol{x}$, so

$$(f_1 \Box f_2)(\boldsymbol{x}) := \inf_{\boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{x}} f_1(\boldsymbol{y}_1) + f_2(\boldsymbol{y}_2) \leq f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) \leq \alpha.$$

  So $\{\boldsymbol{x}, \alpha\} \in \text{epi}\,(f_1 \Box f_2)$.
- What we just showed is $\{\boldsymbol{x}, \alpha\} \in \text{epi}\, f_1 \oplus \text{epi}\, f_2 \implies \{\boldsymbol{x}, \alpha\} \in \text{epi}\,(f_1 \Box f_2)$, so in set language we have $\text{epi}\,(f_1 \Box f_2) \supseteq \text{epi}\, f_1 \oplus \text{epi}\, f_2$.

Example. Pictorial illustration of epi-addition (and Minkowski sum)

- $f = |\cdot|$ $\qquad$ $\operatorname{epi} f = \big\{(x,t) \in \mathbb{R} \times \mathbb{R} \,:\, |x| \le t\big\}$
- $g = \frac{1}{2}(\cdot)^2$ $\qquad$ $\operatorname{epi} g = \big\{(x,t) \in \mathbb{R} \times \mathbb{R} \,:\, \frac{1}{2}x^2 \le t\big\}$
- $(f \square g)(x) = \inf_w |w| + \dfrac{1}{2}(x-w)^2$ $\qquad$ $\operatorname{epi}(f \square g) = \left\{(x,t) \in \mathbb{R} \times \mathbb{R} : \left(\inf_w |w| + \dfrac{1}{2}(x-w)^2\right) \le t\right\}$



"Epi-addition: sliding the blue curve on red curve and perform union operation gives the black curve"

What about $f_1 \Box f_2 \Box f_3$? Inf-convolution is associative

$$
\begin{aligned}
f_1 \Box (f_2 \Box f_3)(t) &= \inf_{x+y=t} \left\{ f_1(x) + (f_2 \Box f_3)(y) \right\} && \text{by definition} \\
&= \inf_{x+y=t} \left\{ f_1(x) + \left\{ \inf_{z+w=y} f_2(w) + f_3(z) \right\} \right\} && \text{by definition} \\
&= \inf_{\substack{x+y=t \\ z+w=y}} \left\{ f_1(x) + f_2(w) + f_3(z) \right\} && \text{you can move inf} \\
&= \inf_{x+z+w=t} \left\{ f_1(x) + f_2(w) + f_3(z) \right\} && \text{combine } x+y=t, z+w=y \\
&= \inf_{\substack{r+z=t \\ x+w=r}} \left\{ f_1(x) + f_2(w) + f_3(z) \right\} && \text{let } t=r+z, x+w=r \\
&= \inf_{r+z=t} \left\{ \left\{ \inf_{x+w=r} f_1(x) + f_2(w) \right\} + f_3(z) \right\} && \text{you can move inf} \\
&= \inf_{r+z=t} \left\{ (f_1 \Box f_2)(r) + f_3(z) \right\} && \text{by definition} \\
&= (f_1 \Box f_2) \Box f_3(t) && \text{by definition}
\end{aligned}
$$

Properties of inf-convolution

We already see

- $f \Box g = g \Box f$ — commutative
- $f \Box g \Box h = (f \Box g) \Box h = f \Box (g \Box h)$ — associative
- f, g convex $\implies f \Box g$ convex — inf-convolution preserves convexity

Useful table

f	g	$f \Box g$
f	0	$\inf_x f(x)$
i_C	$\\| \cdot \\|_2$	d_C
i_{C_1}	i_{C_2}	$i_{C_1 \oplus C_2}$
f	i_x	$f(\cdot - x)$
f	$\langle s, \cdot \rangle$	$\langle s, \cdot \rangle - f^*(s)$
f convex	f	$2f\left(\frac{\cdot}{2}\right)$

- Distance function $d_C(x) = \inf_{c \in C} \|x - c\|_2$
- Indicator function $i_C(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}$
- Conjugate $f^*(x) = \sup_u \langle u, x \rangle - f(x)$

Inf-convolution and conjugate: $(f \Box g)^* = f^* + g^*$

Let $f, g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper functions.

$$\begin{aligned}
(f \Box g)^*(\boldsymbol{y}) &= \sup_{\boldsymbol{x}} \left\{ \langle \boldsymbol{y}, \boldsymbol{x} \rangle - (f \Box g)(\boldsymbol{x}) \right\} && \text{by definition of conjugate} \\
&= \sup_{\boldsymbol{x}} \left\{ \langle \boldsymbol{y}, \boldsymbol{x} \rangle - \inf_{\boldsymbol{u}} \left[f(\boldsymbol{u}) + g(\boldsymbol{x} - \boldsymbol{u}) \right] \right\} && \text{by definition of inf-convolution} \\
&= \sup_{\boldsymbol{x}} \sup_{\boldsymbol{u}} \left\{ \langle \boldsymbol{y}, \boldsymbol{x} \rangle - \left[f(\boldsymbol{u}) + g(\boldsymbol{x} - \boldsymbol{u}) \right] \right\} && -\inf_{\boldsymbol{u}} = +\sup_{\boldsymbol{u}} \\
&= \sup_{\boldsymbol{x}, \boldsymbol{u}} \left\{ \langle \boldsymbol{y}, \boldsymbol{x} {\color{red} - \boldsymbol{u} + \boldsymbol{u}} \rangle - f(\boldsymbol{u}) - g(\boldsymbol{x} - \boldsymbol{u}) \right\} \\
&= \sup_{\boldsymbol{x}, \boldsymbol{u}} \left\{ \langle \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{u} \rangle + \langle \boldsymbol{y}, \boldsymbol{u} \rangle - f(\boldsymbol{u}) - g(\boldsymbol{x} - \boldsymbol{u}) \right\} \\
&= \sup_{\boldsymbol{x}, \boldsymbol{u}} \left\{ \langle \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{u} \rangle - g(\boldsymbol{x} - \boldsymbol{u}) \right\} + \sup_{\boldsymbol{u}} \left[\langle \boldsymbol{y}, \boldsymbol{u} \rangle - f(\boldsymbol{u}) \right] \\
&= g^*(\boldsymbol{y}) + f^*(\boldsymbol{y})
\end{aligned}$$

Recall $\Box$ is similar to convolution: let $\mathcal{F}$ denotes Fourier transform and let $\star$ denotes integral convolution. Then $\mathcal{F}(f \star g) = \mathcal{F}(f) \cdot \mathcal{F}(g)$. Note the correspondence between $(+, \times)$-algebra and $(\min, +)$-algebra

Table of Contents

$$f \,\square\, \frac{1}{2\mu}\|\cdot\|^2$$

Jean Moreau (1923-2014)

Kosaku Yosida (1909-1990)

Moreau-Yosida envelope

- Moreau-Yosida envelope is the special case of infimal convolution $f \Box \frac{1}{2\mu} \| \cdot \|^2$
- Denote e_f^μ the Moreau-Yosida envelope of f under smoothing parameter $\mu > 0$,

$$e_f^\mu(\boldsymbol{x}) \coloneqq \inf_{\boldsymbol{x}_1 + \boldsymbol{x}_2 = \boldsymbol{x}} f(\boldsymbol{x}_1) + \frac{1}{2\mu} \|\boldsymbol{x}_2\|_2^2.$$

  In optimization, usually the following form is used

$$e_f^\mu(\boldsymbol{x}) = \inf_{\boldsymbol{w}} f(\boldsymbol{w}) + \frac{1}{2\mu} \|\boldsymbol{x} - \boldsymbol{w}\|_2^2 \overset{\|\boldsymbol{x}-\boldsymbol{w}\|_2^2 = \|\boldsymbol{w}-\boldsymbol{x}\|_2^2}{=} \inf_{\boldsymbol{w}} f(\boldsymbol{w}) + \frac{1}{2\mu} \|\boldsymbol{w} - \boldsymbol{x}\|_2^2$$

- The point-to-point map associated with Moreau-Yosida envelope is called the Proximal operator

$$\operatorname{prox}_f^\mu(\boldsymbol{x}) \coloneqq \operatorname*{argmin}_{\boldsymbol{w}} f(\boldsymbol{w}) + \frac{1}{2\mu} \|\boldsymbol{w} - \boldsymbol{x}\|_2^2$$

- Why study e_f^μ and $\operatorname{prox}_f^\mu$: they form the basis of modern convex optimization toolbox!
- Remark: $\inf$ becomes $\min$ if f is closed ($\operatorname{epi} f$ is closed) and convex

Why Moreau-Yosida envelope is useful

"Smoothing a non-smooth function to ease optimization"

- Consider minimizing $f(x) = |x|$ using gradient descent $x^+ = x - \alpha \nabla f(x)$
- Gradient descent requires differentiable f, while $|x|$ is not differentiable at $x = 0$.
- Now instead of $\min f(x)$, consider $\min e_f^\mu(x)$, here is the magic:
  - e_f^μ is always differentiable — can use gradient descent!
  - $\min e_f^\mu(x)$ and $\min f(x)$ share the same minimizer

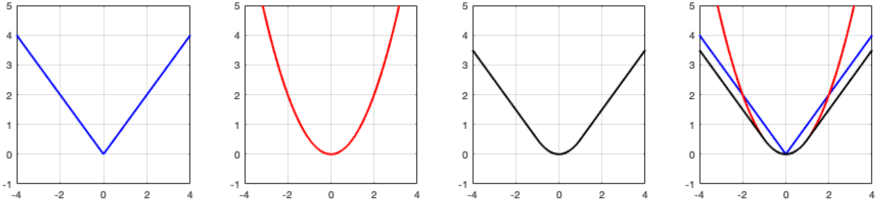
Properties of Moreau-Yosida envelope and proximal operator

$(\mu > 0)$

- e_f^μ is always differentiable — smoothness
- $\min e_f^\mu(x) = \min f(x)$ and $\operatorname{argmin}\, e_f^\mu(x) = \operatorname{argmin}\, f(x)$ — same minimum and minimizer
- f is L-Lipschitz, then $0 \le f(x) - e_f^\mu(x) \le L^2\mu$ — e_f^μ is a lower bound of f
- $\nabla e_f^\mu(v) = \frac{1}{\mu}\big(v - \mathrm{prox}_f^\mu(v)\big) \in \partial f\big(\mathrm{prox}_f^\mu(v)\big)$ — relationship between e_f^μ and prox_f^μ
- Proximal point algorithm = gradient descent on $\min e_f^\mu(x)$ — Proximal point algorithm
- f is nonconvex, then $\mathrm{prox}_f^\mu : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ — non-uniqueness of prox_f^μ
- f is convex, then $\mathrm{prox}_f^\mu : \mathbb{R}^n \to \mathbb{R}^n$ — uniqueness of prox_f^μ
- $e_f^\mu : \mathbb{R}^n \to \mathbb{R}$ — uniqueness of e_f^μ
- $\Big\langle \mathrm{prox}_f^\mu(x) - \mathrm{prox}_f^\mu(y), (\mathrm{Id} - \mathrm{prox}_f^\mu)(x) - (\mathrm{Id} - \mathrm{prox}_f^\mu)(y) \Big\rangle \ge 0$ — prox_f^μ is firmly non-expansive
- $\mathrm{Fix}\,\mathrm{prox}_f^\mu = \underset{y}{\operatorname{argmin}}\ f(y)$ — fixed point
- Let $T = \mathrm{prox}_f^\mu$, then $\big\{T^k x\big\}_{k\in\mathbb{N}} \rightharpoonup \underset{\xi}{\operatorname{argmin}}\ f(\xi)$ — weakly convergence
- $\mathrm{prox}_f^\mu(x) + \mathrm{prox}_{f^*}^\mu(x) = x$ — Moreau decomposition

Example: Moreau-Yosida envelope of absolute value = Huber function

$$f(x) = \begin{cases} -x & x \leq 0 \\ x & x \geq 0 \end{cases}, \quad g = \frac{1}{2\mu}|\cdot|^2, \quad e_f^1(x) = (f \square g)(x) = \begin{cases} -x - \frac{1}{2} & x < -1 \\ \frac{1}{2}x^2 & x \in [-1, +1] \\ x - \frac{1}{2} & x > 1 \end{cases}$$



Example: $\min e_f^\mu(x) = \min f(x)$ and $\operatorname{argmin}\ e_f^\mu(x) = \operatorname{argmin}\ f(x)$

Proof by definition.

$$
\begin{aligned}
\min_x e_f^\mu(x) &= \min_x \left\{ \min_y \left\{ f(y) + \frac{1}{2\mu}\|x-y\|_2^2 \right\} \right\} && \text{by definition of } e_f^\mu \\
&= \min_y \left\{ \min_x \left\{ f(y) + \frac{1}{2\mu}\|x-y\|_2^2 \right\} \right\} && \text{you can swap the order of two min} \\
&= \min_y \left\{ \left\{ f(y) + \min_x \frac{1}{2\mu}\|x-y\|_2^2 \right\} \right\} && f(y) \text{ is constant for } \min_x \\
&= \min_y f(y) && x = y \text{ minimizes } \|x-y\|_2^2 \\
&= \min_x f(x) && \text{rename } y \text{ as } x
\end{aligned}
$$

Similar proof for $\operatorname{argmin}\ e_f^\mu(x) = \operatorname{argmin}\ f(x)$.

Last page - summary

- The epigraph of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is the set $\operatorname{epi} f := \Big\{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \,:\, \alpha \geq f(\boldsymbol{x})\Big\}$.
- Level sets
- Given $f_1 : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$. The infimal convolution $f_1 \Box f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined as

$$
\begin{aligned}
(f_1 \Box f_2)(\boldsymbol{x}) &:= \inf_{\boldsymbol{x}_1 + \boldsymbol{x}_2 = \boldsymbol{x}} f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) \\
&= \inf_{\boldsymbol{x}_1} f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x} - \boldsymbol{x}_1) \\
&= \inf_{\boldsymbol{x}_2} f_1(\boldsymbol{x} - \boldsymbol{x}_2) + f_2(\boldsymbol{x}_2)
\end{aligned}
$$

- Infimal convolution is also called epi-addition
- Moreau-Yosida envelope $f \Box \frac{1}{2\mu} \| \cdot \|^2$ as the foundation of modern optimization toolbox

End of document