PGMO Lecture: Vision, Learning and Optimization

2. Basic notion of convexity

Thomas Pock

Institute of Computer Graphics and Vision

February 11, 2020

Overview

Convexity

An extended real valued function $f : \mathcal{X} \to [-\infty, +\infty]$ is said to be *convex* if and only if its *epigraph*

$$\operatorname{epi} f := \{(x, \lambda) \in \mathcal{X} \times \mathbb{R} : \lambda \geq f(x)\}$$

is a convex set, that is, if when $\lambda \geq f(x)$, $\mu \geq f(y)$, and $t \in [0, 1]$, we have $t\lambda + (1 - t)\mu \geq f(tx + (1 - t)y)$.

It is *proper* if it is not identically $+\infty$ and nowhere $-\infty$: in this case, it is convex if and only if, for all $x, y \in \mathcal{X}$ and $t \in [0, 1]$,

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y).$$

It is *strictly convex* if the above inequality is strict whenever $x \neq y$ and $0 < t < 1$.

Subgradient

Given a convex, extended real valued, l.s.c. function $f : \mathcal{X} \to [-\infty, +\infty]$, we recall that its subgradient at a point x is defined as the set

$$\partial f(x) := \{p \in \mathcal{X} : f(y) \geq f(x) + \langle p, y - x \rangle \ \ \forall y \in \mathcal{X}\}.$$

Fermat's stationary conditions for non-smooth convex functions f is hence generalized as:

$$\boxed{x \in \mathcal{X} \text{ is a global minimizer of } f \text{ if and only if } 0 \in \partial f(x).}$$

Smooth functions

Let us recall that a function f has a L-Lipschitz continuous gradient if there exists a constant $L > 0$ such that for all $x, y \in \mathbb{E}$ one has

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L \|x - y\| .$$

This implies that

$$\langle x - y, \nabla f(x) - \nabla f(y) \rangle \overset{c.s.}{\leq} \|x - y\|_2 \|\nabla f(x) - \nabla f(y)\|_2 \leq L \|x - y\|_2^2$$

Theorem

Let f be a continuously differentiable function over $\mathbb{E}$ with L-Lipschitz continuous gradient. Then, for any $x, y \in \mathbb{E}$

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|x - y\|^2 ,$$

which is known under the name "descent lemma". Moreover, we also have the reverse inequality

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 ,$$

which is called the co-coercivity of the gradient.

Strong convexity

A function f is *strongly convex* or "μ-convex" if in addition, for $x, y \in \mathcal{X}$ and $p \in \partial f(x)$, we have

$$f(y) \geq f(x) + \langle p, y - x \rangle + \frac{\mu}{2} \|y - x\|^2$$

or, equivalently, if $x \mapsto f(x) - \frac{\mu}{2} \|x\|^2$ is convex.

Furthermore, it satisfies

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \mu \frac{t(1-t)}{2} \|y - x\|^2$$

for any x, y and any $t \in [0, 1]$.

Finally, if f is strongly convex and x^* is a minimizer of $f(x)$, then we have

$$f(y) \geq f(x^*) + \frac{\mu}{2} \|y - x^*\|^2$$

for all $y \in \mathcal{X}$. For a L-smooth and μ strongly convex function f, the co-coercivity reads

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{L\mu}{L + \mu} \|x - y\|^2 + \frac{1}{L + \mu} \|\nabla f(x) - \nabla f(y)\|^2 ,$$

which is nothing else then the co-coercivity for the function $f(x) - \frac{\mu}{2} \|x\|^2$, which is $L - \mu$-smooth.

Overview

Legendre-Fenchel conjugate

To any function $f : \mathcal{X} \to [-\infty, +\infty]$ one can associate the *Legendre–Fenchel conjugate* (or convex conjugate)

$$f^*(y) = \sup_{x \in \mathcal{X}} \langle y, x \rangle - f(x)$$

which, as a supremum of linear and continuous functions, is obviously convex and l.s.c.

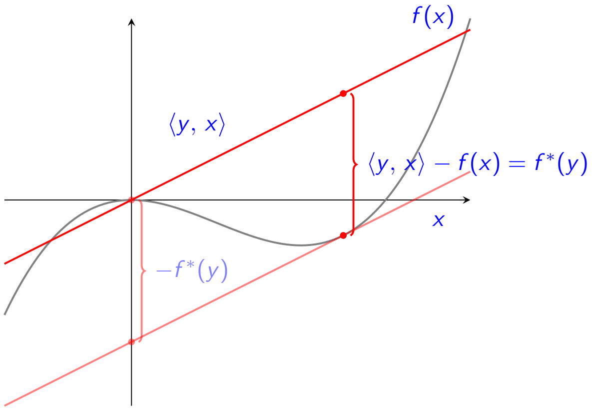


Figure: Illustration of the convex conjugate and its relation to the gradient of smooth functions.

Biconjugate and properties

The *biconjugate* f^{**} is the largest convex l.s.c. function below f. In general $f \geq f^{**}$ but if f is already convex and l.s.c. we have $f^{**} = f$.

Recall the celebrated Legendre-Fenchel identity:

$$y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y) \Leftrightarrow f(x) + f^*(y) = \langle y, x \rangle .$$

We have the following basic properties

- For $f(x) = g(ax)$ with $a \neq 0$, the convex conjugate is given by $f^*(y) = g^*(y/a)$.
- For $f(x) = g(x + b)$, the convex conjugate is given by $f^*(y) = g^*(y) - \langle b, \ y \rangle$.
- for $f(x) = ag(x)$ with $a > 0$, the convex conjugate is given by $f^*(y) = ag^*(y/a)$.
- Note that we also have the property:

$$f(x) \geq g(x) \Longleftrightarrow g^*(y) \geq f^*(y)$$

Monotone operator

The subgradient of a convex function is a *monotone operator*, that is it satisfies

$$\langle p - q, x - y \rangle \geq 0 \quad \forall (x, y) \in \mathcal{X}^2,\ p \in \partial f(x),\ q \in \partial f(y),$$

It is *strongly monotone* if f is strongly convex:

$$\langle p - q, x - y \rangle \geq \mu \|x - y\|^2 \quad \forall (x, y) \in \mathcal{X}^2,\ p \in \partial f(x),\ q \in \partial f(y).$$

An important remark is that f is μ-strongly convex if and only if its conjugate f^* is continuously differeniable with $1/\mu$-Lipschitz gradient, i.e.

$$\|\nabla f^*(p) - \nabla f^*(q)\| \leq \frac{1}{\mu} \|p - q\| .$$

Hence, duality allows to trade strong convexity with smoothness.

Overview

The infimal convolution

Definition

Let $f, g : \mathcal{X} \to (-\infty, \infty]$ be two proper functions. The infimal convolution between f and g is defined by

$$(f \Box g)(x) = \min_u f(u) + g(x - u)$$

- Interpretation: Take the function f and convolve it with the minimum of the function g. The function $f \Box g$ is given by the resulting envelope function.
- Remark: In case f and g are positive and one-homogeneous functions (e.g. norms or semi-norms) the infimal convolution is equivalent to the convex envelope (i.e. biconjugate) of the minimum of the two functions

$$(f \Box g) = (\min\{f(x), g(x)\})^{**} .$$

Graphical interpretation

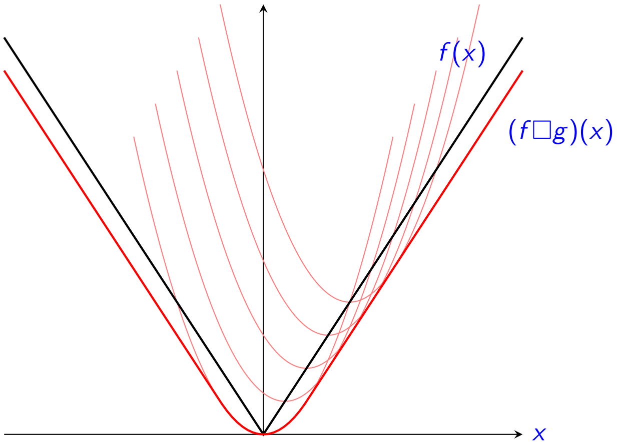


Figure: Illustration of the infimal convolution of the function $f(x) = |x|$ and the quadratic function $g(x) = \frac{1}{2}x^2$.

Properties of the infimal convolution

- The infimal convolution of two convex functions enjoys the following two fundamental properties.

Proposition

Let f and g be two proper convex functions, then the following holds:

- $f \Box g$ *is a convex function.*
- $f \Box g = g \Box f$.

- Proof: The convexity follows from the convexity of partial minimization. The symmetry follows from the fact that the infimal convolution can also be written as

$$(f \Box g)(x) = \min_u f(u) + g(x - u) = \min_{x=u+v} f(u) + g(v).$$

Infimal convolution and convex conjugates

There are interesting connections between the convex conjugate and the infimal convolution.

Theorem

Let $f, g : \mathcal{X} \to (-\infty, \infty]$ be two proper functions and let f^*, g^* be their convex conjugates. It holds that

$$(f \Box g)^* = f^* + g^*.$$

The second direction requires additional assumptions such as convexity.

Theorem

Let $f : \mathcal{X} \to (-\infty, \infty]$ be a proper convex function and let $g : \mathcal{X} \to \mathbb{R}$ be a real-valued convex function. Then

$$(f + g)^* = f^* \Box g^*.$$

The Moreau envelope

- An interesting special case of the infimal convolution is the Moreau envelope.
- It is obtained by choosing one function in the infimal convolution as a quadratic function

$$x \mapsto f_\lambda(x) = (f \Box \tfrac{1}{2\lambda} \|\cdot\|_2^2)(x) = \min_u f(u) + \frac{1}{2\lambda} \|x - u\|^2 ,$$

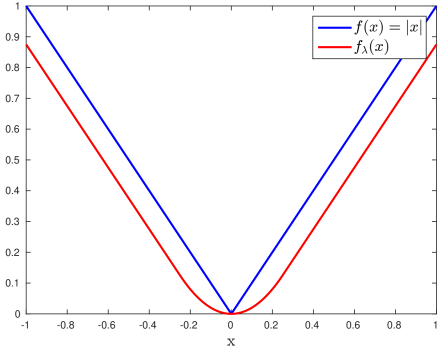
  where $\lambda > 0$ is a "smoothing" parameter.
- It can be used to "smooth" a non-smooth convex function without destroying its minimizer.

Example

- Let $f(x) = |x|$. The Moreau-envelope $f_\lambda(x)$ is given by

$$f_\lambda(x) = \begin{cases} \frac{x^2}{2\lambda} & \text{if } |x| \le \lambda \\ |x| - \frac{\lambda}{2} & \text{else.} \end{cases}$$

- This is exactly the Huber function from robust statistics!



Overview

Convex functions

Legendre-Fenchel conjugate

Infimal convolution

Proximal map

Duality

The proximal map

- Let us consider again the Moreau envelope of a function f:

$$\min_x f(x) + \frac{1}{2\lambda} \|x - y\|^2$$

- The point that attains the minimum in the Moreau envelope is called the proximal map.

Definition

Let $\lambda > 0$ be a parameter. The proximal map with parameter λ is defined as

$$\text{prox}_{\lambda f}(y) = \arg\min_x f(x) + \frac{1}{2\lambda} \|x - y\|^2$$

- It can be seen as a generalization of the Euclidean projection of a point y on a convex set C.

Optimality condition of the proximal map

- The optimality condition of the proximal map is given by

$$\lambda\partial f(x) + x - y \ni 0 \iff y - x \in \lambda\partial f(x)$$

- In operator notation, the proximal map is written as

$$\begin{aligned} x - y + \lambda\partial f(x) \ni 0 \\ \Leftrightarrow x + \lambda\partial f(x) \ni y \\ \Leftrightarrow (I + \lambda\partial f)x \ni y \\ \Leftrightarrow x = (I + \lambda\partial f)^{-1}(y) \\ \Leftrightarrow x = \operatorname{prox}_{\lambda f}(y) \end{aligned}$$

Examples

- Let $f(x) = \frac{1}{2}x^T A x + b^T x + c$. The proximal map is given by

$$\operatorname{prox}_{\lambda f}(y) = (I + \lambda A)^{-1}(y - \lambda b)$$

- Let $f(x) = \|x\|_1$, the proximal map is given by

$$(\operatorname{prox}_{\lambda f}(y))_i = \max(0, |y_i| - \lambda)\operatorname{sgn}(y_i)$$

- Let $f(x) = -\sum_{i=1}^n \log x_i$, the proximal map is given by

$$(\operatorname{prox}_{\lambda f}(y))_i = \frac{y_i + \sqrt{y_i^2 + 4\lambda}}{2}$$

- Let $f(x) = \delta_{\|\cdot\|_\infty \le 1}(x)$, the proximal map is given by

$$(\operatorname{prox}_f(y))_i = \frac{y_i}{\max(1, |y_i|)}$$

Proximal map calculus

- Let $f(x,y) = f_1(x) + f_2(y)$, then the proximal map with respect to f is given by

$$\text{prox}_f(u,v) = (\text{prox}_{f_1}(u), \text{prox}_{f_2}(v)).$$

- If $f(x) = \alpha g(x) + b$, with $\alpha > 0$, then

$$\text{prox}_f(u) = \text{prox}_{\alpha f}(u)$$

- If $f(x) = g(\alpha x + b)$, with $\alpha \neq 0$ then

$$\text{prox}_f(u) = \frac{1}{\alpha}\left(\text{prox}_{\alpha^2 g}(\alpha u + b) - b\right)$$

- If $f(x) = g(Qx)$, with Q orthogonal (such that $QQ^T = Q^T Q = I$), then

$$\text{prox}_f(u) = Q^T \text{prox}_g(Qu)$$

Proximal map calculus

- If $f(x) = g(x) + a^T x + b$, then

$$\text{prox}_f(u) = \text{prox}_g(u - a)$$

- If $f(x) = g(x) + \frac{\gamma}{2}\|x - a\|^2$, then

$$\text{prox}_f(u) = \text{prox}_{\tilde{\gamma}g}\left(\tilde{\gamma}u + \tilde{\gamma}\gamma a\right),$$

with $\tilde{\gamma} = 1/(1 + \gamma)$.

Moreau identity

- The Moreau identity [Moreau '65] connects the proximal map of a function with the proximal map of its convex conjugate.

Theorem

Let f be a convex function and let f^* be its convex conjugate. Then

$$x = \mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x)$$

- One also has that

$$x = \mathrm{prox}_{\tau f}(x) + \tau \mathrm{prox}_{\frac{1}{\tau} f^*}\left(\frac{x}{\tau}\right)$$

- From a practical point of view, it shows that the proximal map of a function f is as easy to compute as the proximal map of its convex conjugate.
- Example: $f(x) = |x|$. Its convex conjugate is $f^*(y) = \delta_{[-1,1]}(y)$ and $\mathrm{prox}_{f^*}(y) = \max(-1, \min(1, x))$, hence

$$\mathrm{prox}_f(x) = x - \max(-1, \min(1, x)).$$

The gradient of the Moreau envelope

- It turns out that the the gradient of the Moreau envelope of a convex function f has a particularly appealing structure.

Proposition

Let f be a convex function and let $\lambda > 0$. Then the Moreau envelope of f defined through

$$(f \square \frac{1}{2\lambda} \|\cdot\|_2^2)(x) = \min_y f(y) + \frac{1}{2\lambda} \|x - y\|_2^2$$

is continuously differentiable and its gradient

$$\nabla \left(f \square \frac{1}{2\lambda} \|\cdot\|_2^2 \right) = \frac{1}{\lambda}(I - \mathrm{prox}_{\lambda f})$$

is λ^{-1} Lipschitz continuous.

Overview

Convex functions

Legendre-Fenchel conjugate

Infimal convolution

Proximal map

Duality

Fenchel-Rockafellar Duality

- Let us consider the following (primal) optimization problem

$$\min_x f(Kx) + g(x),$$

where f, g are closed convex functions and $K \in \mathbb{R}^{m \times n}$ is a matrix. Using the fact that $f = f^{**}$ we have

$$\min_x f(Kx) + g(x) = \min_x \max_y \langle y,\, Kx\rangle - f^*(y) + g(x).$$

- Under very mild conditions, we can swap the min and max which yields

$$\max_y \min_x \langle y,\, Kx\rangle - f^*(y) + g(x) = \max_y - \max_x - \langle x,\, K^* y\rangle + f^*(y) - g(x).$$

- Using the definition of g^* we obtain the Rockafellar-Fenchel dual

$$\max_y -f^*(y) - g^*(-K^* y),$$

Strong duality

- Let us denote by

$$\mathcal{P}(x) = f(Kx) + g(x), \quad \mathcal{D}(y) := -f^*(y) - g^*(-K^*y),$$

  the primal and dual problems.
- From the Fenchel-Rockafellar duality it follows that strong dualty holds, i.e.

$$\max_y \mathcal{D}(y) = \min_x \mathcal{P}(x)$$

- Moreover,

$$\mathcal{D}(y) \leq \mathcal{P}(x), \quad \forall (x, y).$$

- Hence, it is natural to define the primal-dual gap

$$\mathcal{G}(x, y) = \mathcal{P}(x) - \mathcal{D}(y) \geq 0,$$

  which vanishes if and only if (x, y) is an optimal solution pair of the primal and dual problems.

Saddle point formulation

- Let us define the Lagrangian function

$$\mathcal{L}(x, y) := \langle y, Kx\rangle - f^*(y) + g(x)$$

- If x^* is a solution to the primal problem and y^* is the solution of the dual problem, then

$$\max_y \mathcal{L}(x^*, y) = \mathcal{P}(x^*) = \mathcal{L}(x^*, y^*) = \mathcal{D}(y^*) = \min_x \mathcal{L}(x, y^*)$$

- Observing that

$$\mathcal{L}(x^*, y) \leq \mathcal{P}(x^*), \quad \mathcal{D}(y^*) \leq \mathcal{L}(x, y^*),$$

  we obtain

$$\mathcal{L}(x^*, y) \leq \mathcal{L}(x^*, y^*) \leq \mathcal{L}(x, y^*),$$

  hence the optimal solution pair (x^*, y^*) is a saddle point of the Lagrangian.

Minimax Theorem

The mixed primal-dual formulation that appeared in the derivation of the Fenchel-Rockafellar dual problem belongs to the class of convex-concave minimax problems.

Theorem

Let $X \subset \mathcal{X}$ and $Y \subset \mathcal{X}^*$ be convex and compact sets. Moreover, let $f(\cdot, y)$ be convex for all fixed y and $f(x, \cdot)$ concave for all fixed x. Then, we have that:

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y)$$

- Proved by John von Neumann in 1935: *As far as I can see, there could be no theory of games ... without that theorem ... I thought there was nothing worth publishing until the Minimax Theorem was proved.*
- Extended to quasi-convex/concave functions by Sion in 1958.

Non-convex minimax theorem

Finally, we present a result on general minimax problems:

Theorem

Let $f : X \times Y \mapsto \mathbb{R}$ be a saddle point problem, where X and Y are some subsets of $\mathcal{X}$ and $\mathcal{X}^*$, then

$$\min_{x \in X} \max_{y \in Y} f(x, y) \geq \max_{y \in Y} \min_{x \in X} f(x, y).$$

- Interpretation: It matters who plays first in games.
- Proof: First note that we have

$$f(x, y) \geq \min_{\xi} f(\xi, y), \quad \forall (x, y) \in X \times Y$$

- Taking the maximum wrt y on both sides

$$\max_{y} f(x, y) \geq \max_{y} \min_{\xi} f(\xi, y), \quad \forall x \in X$$

- Finally, taking the minimum wrt x on both sides yields the desired result.

Example: Dual formulation of the ROF model

Recall that the primal ROF model is given by

$$\min_u \|\mathrm{D}u\|_{p,1} + \frac{1}{2}\|u - d\|^2$$

The dual ROF model is given by

$$\max_{\mathbf{p}} -\delta_{\{\|\cdot\|_{q,\infty}\leq\lambda\}}(\mathbf{p}) - \frac{1}{2}\|\mathrm{D}^*\mathbf{p}\|^2 + \langle \mathrm{D}^*\mathbf{p},\, f\rangle\,,$$

where $u = d - \mathrm{D}^*\mathbf{p}$. The primal-dual gap is given by

$$\mathcal{G}(u,\mathbf{p}) = \|\mathrm{D}u\|_{p,1} + \delta_{\{\|\cdot\|_{q,\infty}\leq\lambda\}}(\mathbf{p}) - \langle \mathbf{p},\, \mathrm{D}u\rangle + \frac{1}{2}\|d - \mathrm{D}^*\mathbf{p} - u\|^2 \geq 0$$

It turns out that for u^* being the minimizer of the primal ROF model,

$$\mathcal{G}(u,\mathbf{p}) \geq \frac{1}{2}\|u - u^*\|^2 + \frac{1}{2}\|d - \mathrm{D}^*\mathbf{p} - u^*\|^2\,.$$