Epigraphical Projections in Nonsmooth Optimization

Ariel Goodwin

Supervisor: Tim Hoheisel

University of McGill

ariel.goodwin@mail.mcgill.ca

August 28th, 2020

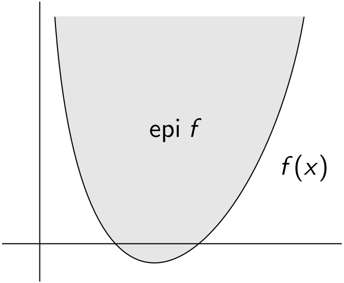
Convexity

Definition 1

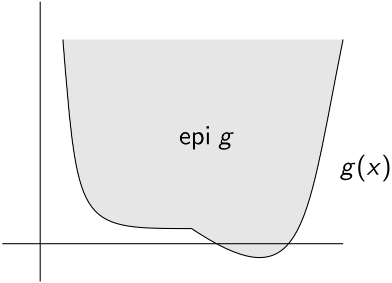
A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex if
epi $f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \alpha\}$ (**epigraph** of f) is a convex set.

Equivalently, for all $x, y \in$ dom $f := \{x \in \mathbb{R}^n : f(x) < +\infty\}, \lambda \in [0, 1]$ we have:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \tag{1}$$



Convex function



Non-convex function

The function class Γ_0

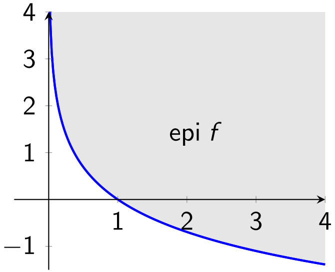
$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called:

- **closed** if epi f is a closed set in $\mathbb{R}^n \times \mathbb{R}$
- **proper** if dom $f \neq \emptyset$

<u>Example</u>:

The function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$

$$f(x) = \begin{cases} -\log x, & x > 0, \\ +\infty, & x \leq 0, \end{cases}$$



Set $\Gamma_0 := \{f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \mid f \text{ is convex, proper, and closed}\}$

Proximal Operators

$$\text{prox}_f(x) := \underset{u \in \mathbb{R}^n}{\text{argmin}} \; f(u) + \frac{1}{2}||u - x||^2$$

- Ubiquitous in convex optimization algorithms
- Exists uniquely if $f \in \Gamma_0$

Example: The *indicator function* for a set $C \subseteq \mathbb{R}^n$ is

$$\delta_C(x) := \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C, \end{cases}$$

$\text{prox}_{\delta_C}(x) = \underset{y \in C}{\text{argmin}}||x - y||^2 =: P_C(x)$ – the **projection operator**

Epigraphical Projection via Prox Operator

Given $(\bar{x}, \bar{\alpha}) \in \mathbb{R}^n \times \mathbb{R}$ and $f \in \Gamma_0$, consider projecting $(\bar{x}, \bar{\alpha})$ onto epi f.

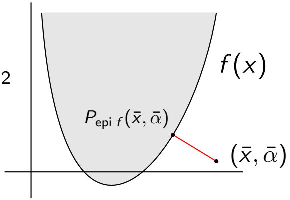
Theorem 1

$$P_{\text{epi } f}(\bar{x}, \bar{\alpha}) = (\text{prox}_{\bar{\lambda} f}(\bar{x}), \bar{\alpha} + \bar{\lambda}) \tag{2}$$

where $\bar{\lambda} > 0$ is the unique minimizer of the (strongly convex) optimization problem

$$\min_{\lambda \geq 0} \ \theta_{\text{epi}}(\lambda) := \frac{1}{2}\lambda^2 + \bar{\alpha}\lambda + \bar{\phi}_f^{\bar{x}}(\lambda) \tag{3}$$

- We focus on the case $(\bar{x}, \bar{\alpha}) \notin$ epi f
- $\bar{\phi}_f^{\bar{x}}(\lambda) := -\lambda f(\text{prox}_{\lambda f}(\bar{x})) - \frac{1}{2}||\bar{x} - \text{prox}_{\lambda f}(\bar{x})||^2$ for $\lambda > 0$

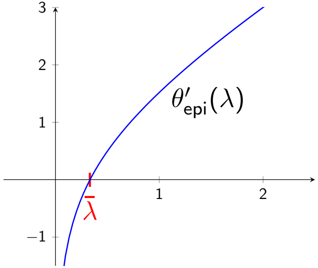


We propose a variant of Newton's method, based on [4, Algorithm 3.1].

$$\lambda_{k+1} = \lambda_k + t_k P_{[-\lambda_k,\infty]}\left(-\frac{\theta'_{\text{epi}}(\lambda_k)}{g_k}\right)$$

- g_k are generalized gradients, by Clarke [1]
- Armijo line search to choose t_k
- If $-\theta'_{\text{epi}}$ has some convexity we can set $t_k = 1$

Example: The function θ'_{epi} when projecting $(-1,1) \in \mathbb{R} \times \mathbb{R}$ onto $\text{epi}(-\log(\cdot))$



Level Set Projections

We can similarly project onto **level sets** of $f \in \Gamma_0$

$$\text{Lev}(f,\alpha) := \{x \in \mathbb{R}^n : f(x) \leq \alpha\} \; (\alpha \in \mathbb{R})$$

Note that $\theta_{\text{lev}}(\lambda) = \theta_{\text{epi}}(\lambda) - \frac{1}{2}\lambda^2$

Example: Let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = ||x||_1 = \sum_{i=1}^n |x_i|$, the l_1-norm. Then

$$\text{Lev}(f,1) = B_{||\cdot||_1}[0,1]$$

is the l_1-unit ball.

- Applications to machine learning and image problems
- Promotes sparse solutions

Projection onto the l_1-ball

We tested the Newton method against two competitive algorithms described in [2] and [3]: the proposed algorithm of Condat and Improved Bisection (IBIS) of Liu and Ye.

Table: Time (seconds) for projecting vectors onto the l_1-unit ball in dimension N with coordinates chosen using a Gaussian distribution with $\sigma = 0.1$

N	Warm Newton	Condat	IBIS
20	1.44×10^{-6}	1.53×10^{-6}	1.83×10^{-6}
10^3	1.83×10^{-5}	2.11×10^{-5}	3.65×10^{-5}
10^6	1.38×10^{-2}	1.44×10^{-2}	2.89×10^{-2}
10^7	1.51×10^{-1}	1.43×10^{-1}	2.85×10^{-1}

Using a warm start implementation, our algorithm performs better than or roughly on par with the competitors.

References

Frank H. Clarke (1983)
Optimization and Nonsmooth Analysis
John Wiley & Sons, New York

Laurent Condat (2016)
Fast projection onto the simplex and l_1 ball
Mathematical Programming, Series A, Springer
158 (1), pp. 575 - 585

Jun Liu and Jieping Ye (2009)
Efficient Euclidean projections in linear time
Proceedings of the 26th Annual International Conference on Machine Learning
pp. 657 - 664

Jong-Shi Pang and Liqun Qi (1995)
A globally convergent Newton method for convex SC^1 minimization problems
Journal of Optimization Theory and Applications
85(3), pp. 633 - 648