Proximal-point like algorithms for abstract convex minimisation problems

R. Díaz Millán J. Ugon

AustMS Meeting 2025

Introduction

Generalising the gradient descent method

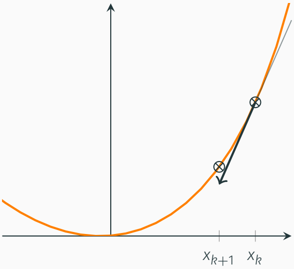
$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

We want to generalise this method to *globally* minimise some classes of *nonconvex* functions.

Let f be convex. The (sub)gradient method step is:

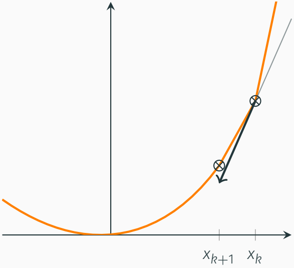
$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$



Let f be convex. The (sub)gradient method step is:

$$x_{k+1} = x_k - \alpha_k g_k$$

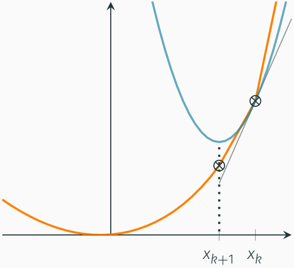
where $g_k \in \partial f(x_k)$.



Let f be convex. The (sub)gradient method step is:

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k g_k \\ &= \operatorname{argmin}\left(\langle g_k, x\rangle + \frac{1}{2c_k}\|x - x_k\|^2\right) \end{aligned}$$

where $g_k \in \partial f(x_k)$.



The subgradient method is a linearisation of the proximal point method:

$$x_{k+1} = \mathrm{prox}_{c_k, f}(x_k)$$

$$\text{where } \mathrm{prox}_{c_k, f}(x_k) = \operatorname*{argmin}_{x \in X} f(x) + \frac{1}{2c_k} \|x - x_k\|^2$$

The subgradient method is a linearisation of the proximal point method:

$$x_{k+1} = \mathsf{prox}_{c_k, f}(x_k)$$

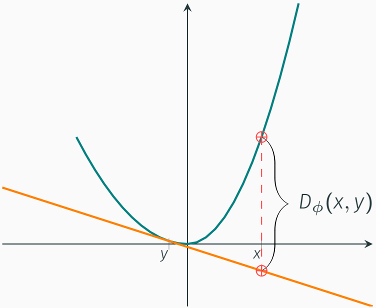
$$\text{where } \mathsf{prox}_{c_k, f}(x_k) = \operatorname*{argmin}_{x \in X} f(x) + \frac{1}{2c_k}\|x - x_k\|^2$$

Bregman-based methods

Let ϕ be a continuously differentiable strongly convex function.

Definition (Bregman Divergence (Brègman 1965))
The *Bregman divergence* is the function

$$D_\phi(x, y) = \phi(x) - \phi(y) - \langle \nabla\phi(y), x - y \rangle.$$



Bregman Proximal Point Method (Bregman 1967)

$$x_{k+1} = \operatorname*{argmin}_{x \in X} f(x) + \frac{1}{c_k} D_\phi(x, x_k)$$

Mirror Descent Method (Nemirovsky and Yudin 1983)

$$
\begin{aligned}
u_k &\in \partial f(x_k) \\
x_{k+1} &= \operatorname*{argmin}_{x \in X} \langle u_k, x \rangle + \frac{1}{c_k} D_\phi(x, x_k),
\end{aligned}
$$

Let ϕ be a strongly convex function.

Definition (Generalised Bregman Divergence (Kiwiel 1997a))
The *generalised Bregman divergence* is a function

$$D_\phi^\lambda(x, y) = \phi(x) - \phi(y) - \langle \lambda, x - y \rangle.$$

where $\lambda \in \partial\phi(y)$.

$D_\varphi^\lambda(x, y) \geq 0$ for any x, y with equality iff $x = y$.

Generalised Bregman Proximal Point Method (Kiwiel 1997b)

$$\begin{cases} x_{k+1} = \operatorname*{argmin}_{x \in X} f(x) + \dfrac{1}{c_k} D_\phi^{\lambda_k}(x, x_k) \\ \lambda_{k+1} = c_k u_k - \lambda_k, & \text{where } u_k \in \partial f(x_k). \end{cases}$$

Mirror Descent Method (Díaz Millán and U. 2025)

$$\begin{cases} u_k \in \partial f(x_k) \\ x_{k+1} = \operatorname*{argmin}_{x \in X} \langle u_k, x \rangle + \dfrac{1}{c_k} D_\phi^{\lambda_k}(x, x_k), \\ \lambda_{k+1} = c_k u_k - \lambda_k \end{cases}$$

- ϕ strongly convex:

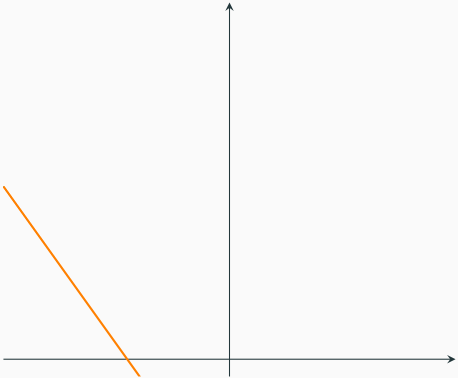
$$\phi(y) \geq \phi(x) + \langle u, x \rangle + \sigma \|y - x\|^2, \forall x, y \in X, u \in \partial\phi(x)$$

- f convex and continuous.
- $\|u\|_*$ bounded for all $u \in \partial f(x)$, $x \in X$.

Abstract Convexity

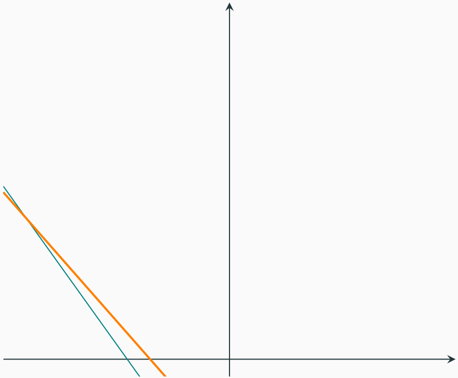
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



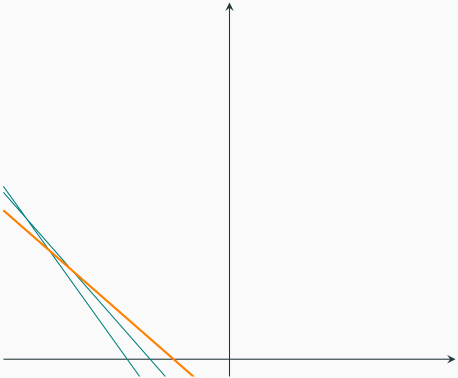
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b) \in U} \langle a, x \rangle + b$$



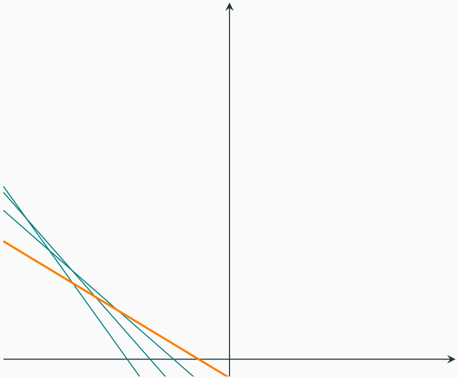
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



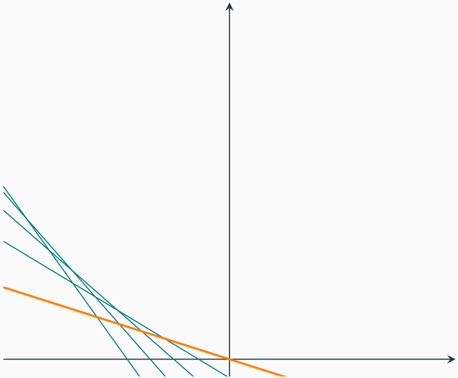
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



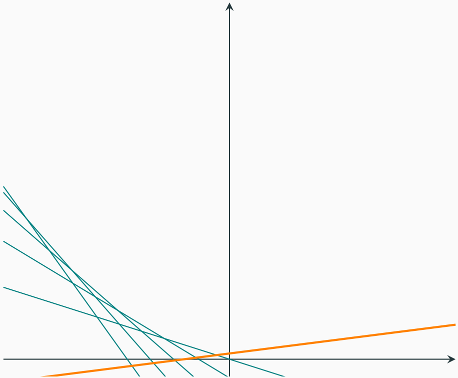
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



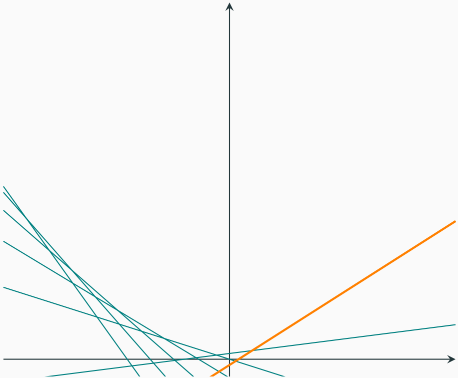
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



A lower semi-continuous convex function can be written as a supremum of affine functions:

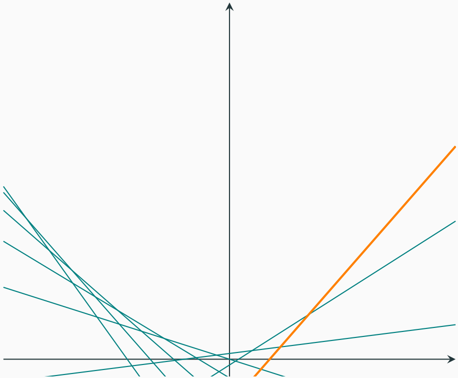
$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



Convexity

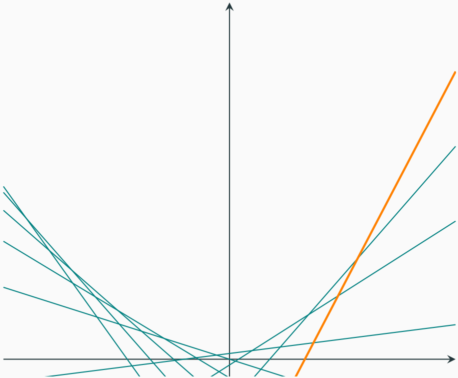
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



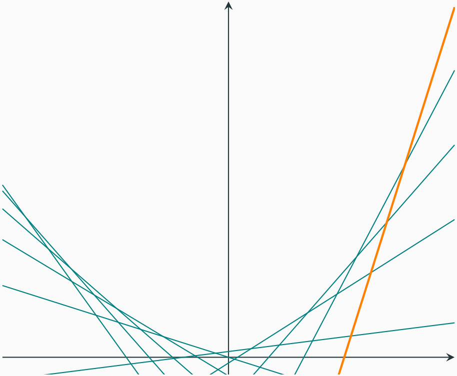
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



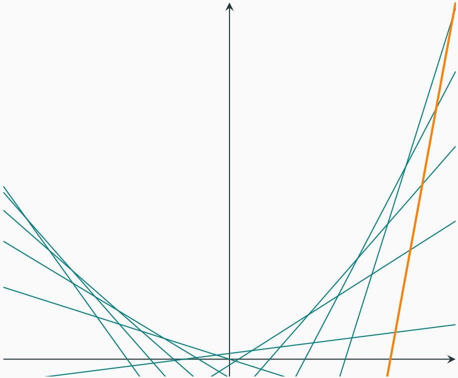
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



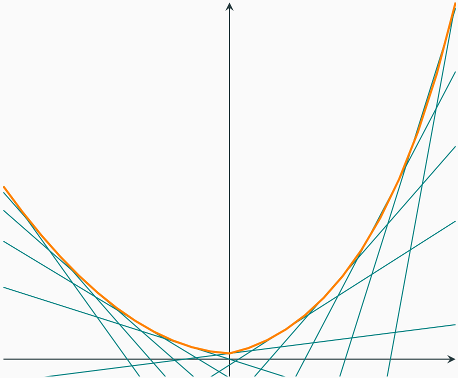
A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x \rangle + b$$



A lower semi-continuous convex function can be written as a supremum of affine functions:

$$f(x) := \sup_{(a,b)\in U} \langle a, x\rangle + b$$



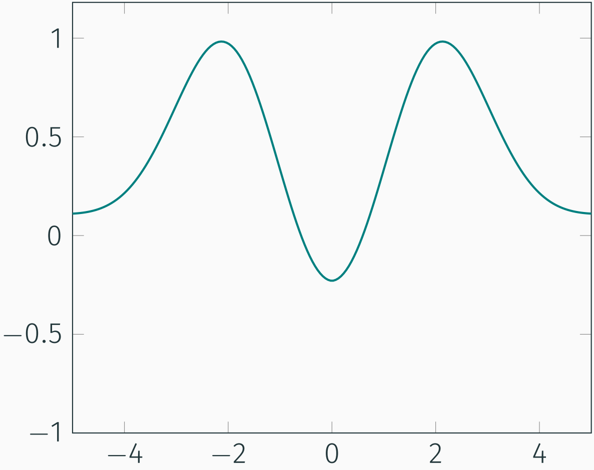
An abstract convex function can be written as a supremum of abstract affine functions.

An abstract convex function can be written as a supremum of abstract affine functions.

Definition (abstract convex function (Kutateladze and Rubinov 1972, 1976))
Let L be a family of functions defined on X. The function $f : X \to \mathbb{R}$ is called *(L,X)-convex* if there exists $U \subset L$ such that

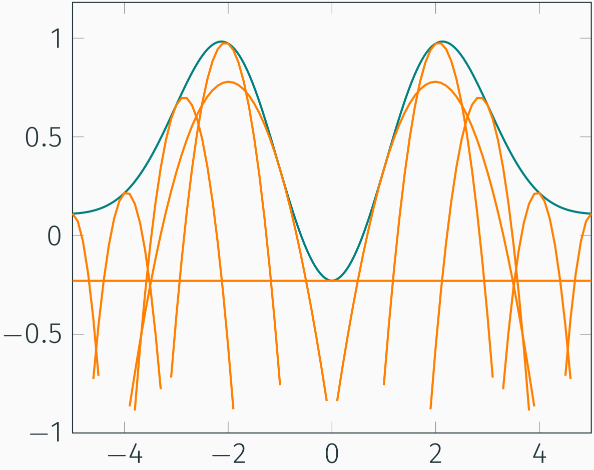
$$f(x) := \sup_{u \in U} u(x)$$

$$L = \{ax^2 + bx : a, b \in \mathbb{R}\}$$



Abstract convex function: example

$$L = \{ax^2 + bx : a, b \in \mathbb{R}\}$$



Let L be a family of functions and f a (L, X)-convex function.

Definition (Abstract Subdifferential (Rubinov 2000))

$$\partial_L f(x) = \{u \in L : f(y) \geq f(x) + u(y) - u(x), \forall y \in X\}$$

1. Cutting Angle Method (Andramonov, Rubinov, and Glover 1999): Lipschitz continuous functions.
2. Proximal Point Method (Bednarczuk, Lorenz, and Tran 2025): Weakly convex functions.

Abstract Bregman proximal methods

Definition (Abstract Bregman Divergence (Grasmair 2010))

Let ϕ be a (L, X)-function and $y \in X$, and let $\lambda \in \partial_L \phi(y)$. The L-Bregman divergence from y with respect to λ as

$$D_\phi^\lambda(x, y) = \phi(x) - \phi(y) - (\lambda(x) - \lambda(y))$$

Abstract Bregman Proximal Point Method (Díaz Millán and U. 2025)

$$\begin{cases} x_{k+1} \in \operatorname*{Argmin}_{x \in X} f(x) + \dfrac{1}{c_k} D_{\phi}^{\lambda_k}(x, x_k) \\ \lambda_{k+1} = c_k u_k - \lambda_k, & \text{where } u_k \in \partial_L f(x_k). \end{cases}$$

1. L is a linear space.
2. Sum Rule: $\partial_L \phi(x) + \partial_L f(x) = \partial_L(\phi + f)(x)$.
3. $\sum_{k=1}^{\infty} c_k = \infty$.

Let ϕ and f be L-convex functions. We want to minimise $f(x)$.

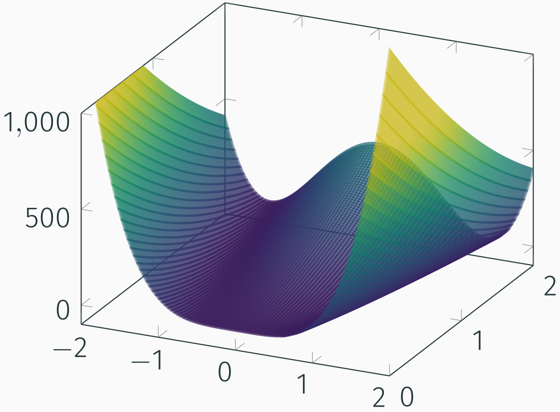
$$\begin{cases} u_k \in \partial_L f(x_k) \\ x_{k+1} \in \operatorname*{Argmin}_{x \in X} u_k(x) + \frac{1}{c_k} D_\phi^{\lambda_k}(x, x_k), \\ \lambda_{k+1} = c_k u_k - \lambda_k \end{cases}$$

Assumptions

1. $L - L = L$.
2. $u(x) - u(y) \leq D_\phi^\lambda(x, y) + \phi^*(u)$.
   Classical case: $\langle u, x - y\rangle \leq 1/2\|x - y\|^2 + 1/2\|u\|_*^2$.
3. $\exists\alpha > 1 : \phi^(cu) \leq c^\alpha\phi^*(u)$ for any $c \geq 0$.
   Classical case: ϕ is strongly convex implies $\alpha = 2$.
4. $\phi^*(u)$ bounded for any $u \in \partial_L f(x)$, any $x \in X$.
   Classical case: $1/2\|u\|_*^2$ is bounded.
5. $\sum_{i=1}^\infty c_k = \infty$ and $\sum_{i=1}^\infty c_k^\alpha < \infty$.
6. Very minimal assumption on L.

Numerical Experiments

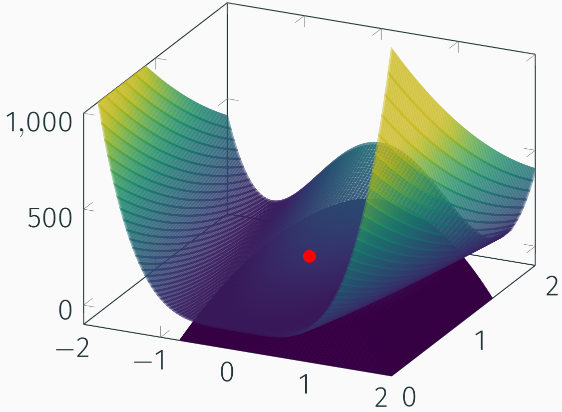
$$L = \{a\|x\|^2 + \langle b, x \rangle : (a, b) \in \mathbb{R} \times \mathbb{R}^n\}$$
$$\phi(x) = \|x\|^3$$
$$f(x) = (1 - x)^2 + a(y - x^2)^2, a = 100.$$



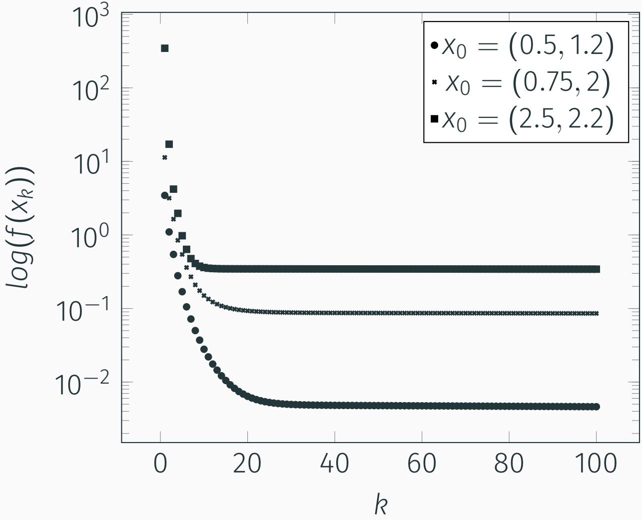
$$L = \{a\|x\|^2 + \langle b, x\rangle : (a, b) \in \mathbb{R} \times \mathbb{R}^n\}$$

$$\phi(x) = \|x\|^3$$

$$f(x) = (1-x)^2 + a(y-x^2)^2, a = 100.$$



Iterations



10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
$log(f(x_k))$
$x_0 = (0.5, 1.2)$
$x_0 = (0.75, 2)$
$x_0 = (2.5, 2.2)$
0
20
40
60
80
100
k

Example (Bui et al. 2021)

Let $L = \{ax^2 + b : a, b \in \mathbb{R}\}$ and let $f(x) = f_1(x) + f_2(x) + f_3(x)$ where

$$
\begin{aligned}
f_1(x) &= x^4 - x^2 \\
f_2(x) &= 1 - 2|x| \\
f_3(x) &= \begin{cases} 1 - 2|x| & \text{if } -\frac{1}{2} \leq x \leq \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

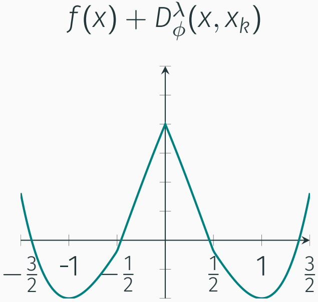
$$\phi(x) = -|x|.$$

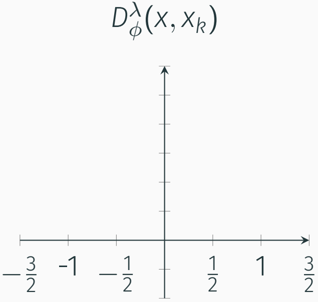
The Bregman divergence is then:

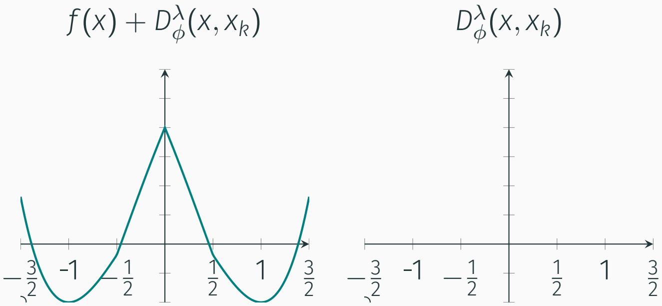
$$D_\phi^\lambda(x, y) = -|x| + |y| + \frac{x^2 - y^2}{2|y|}$$

- Well defined for any x and any $y \neq 0$.
- $D_\phi^\lambda(x, y) = D_\phi^\lambda(x, -y)$ for any x and $y \neq 0$.

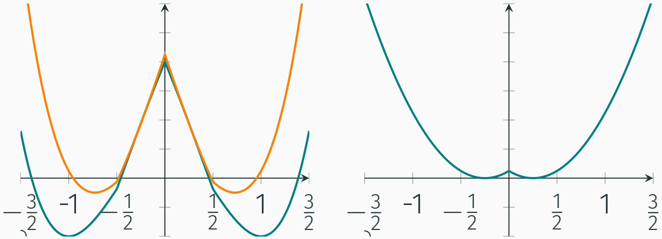
Numerical Experiments

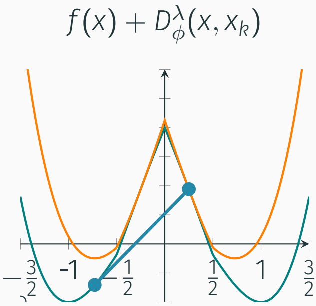


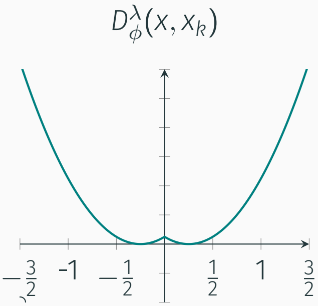



$f(x) + D_\phi^\lambda(x, x_k)$
$D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$

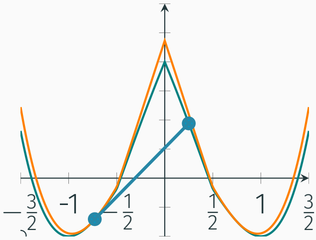
$f(x) + D_\phi^\lambda(x, x_k)$ $D_\phi^\lambda(x, x_k)$



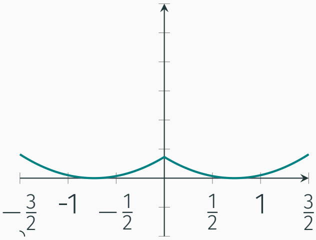

$f(x) + D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$

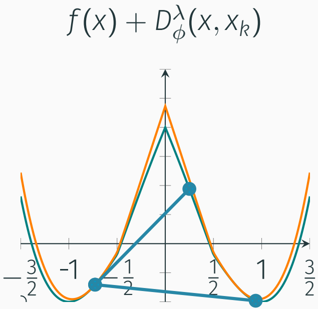

$D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$

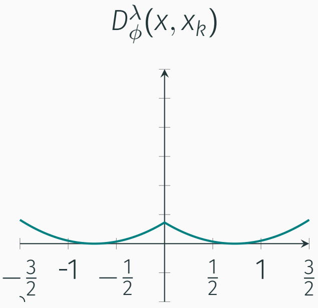
$f(x) + D^{\lambda}_{\phi}(x, x_k)$



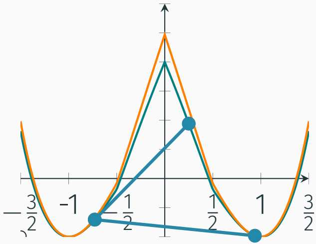
$D^{\lambda}_{\phi}(x, x_k)$



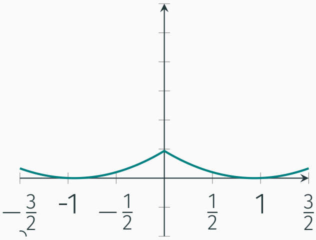

$f(x) + D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$

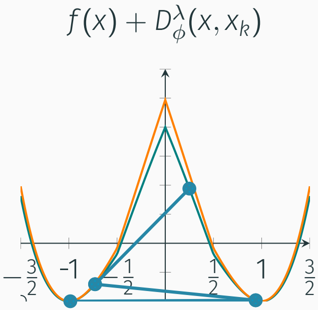

$D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$

$f(x) + D_\phi^\lambda(x, x_k)$

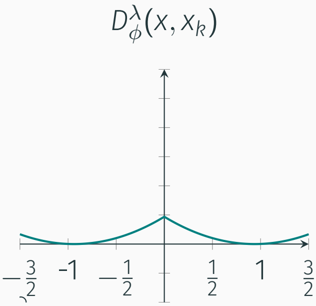


$D_\phi^\lambda(x, x_k)$

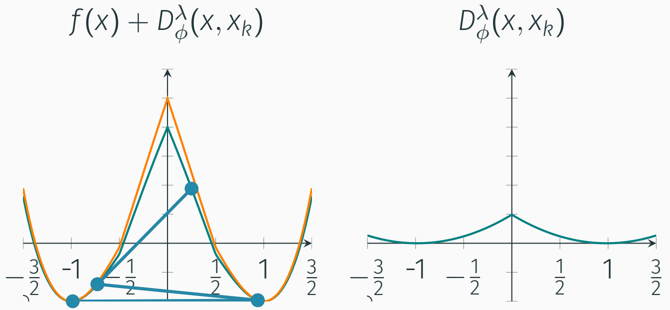


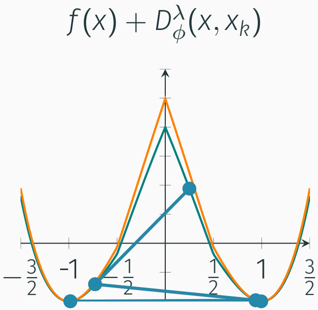


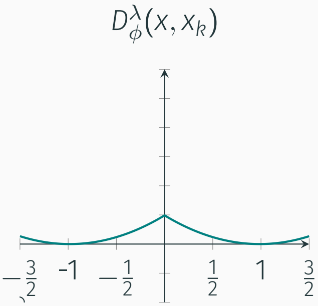
$f(x) + D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$



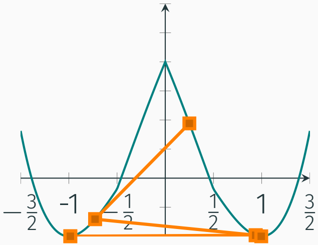
$D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$


$f(x) + D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$
$D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$

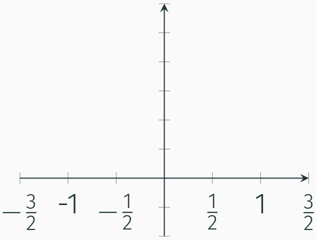

$f(x) + D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$


$D_\phi^\lambda(x, x_k)$
$-\frac{3}{2}$
-1
$-\frac{1}{2}$
$\frac{1}{2}$
1
$\frac{3}{2}$

$f(x) + D_\phi^\lambda(x, x_k)$



$D_\phi^\lambda(x, x_k)$



References

- Andramonov, M., A. Rubinov, and B. Glover (1999). **"Cutting angle methods in global optimization".** In: *Applied Mathematics Letters* 12.3, pp. 95–100.
- Bednarczuk, Ewa, Dirk Lorenz, and The Hung Tran (2025). **"Proximal Algorithms for a Class of Abstract Convex Functions".** In: *Set-Valued and Variational Analysis* 33.1.
- Brègman, L. M. (1965). **"Finding the common point of convex sets by the method of successive projection".** In: *Doklady Akademii Nauk SSSR* 162 (3), pp. 487–490.

Bregman, L.M. (1967). **"The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming".** In: *USSR Computational Mathematics and Mathematical Physics* 7.3, pp. 200–217.

Bui, Hoa T. et al. (2021). **"Zero duality gap conditions via abstract convexity".** In: *Optimization*, pp. 1–37. arXiv: 1910.08156 [math.OC].

Díaz Millán, Reinier and J. (2025). **"Proximal-point-like algorithms for abstract convex minimisation problems".** In: arXiv: 2402.04281v1 [math.OC].

Grasmair, Markus (2010). **“Generalized Bregman distances and convergence rates for non-convex regularization methods”.** In: *Inverse Prob.* 26.11, p. 115014.

Kiwiel, Krzysztof C. (1997a). **“A Bregman-Projected Subgradient Method for Convex Constrained Nondifferentiable Minimization”.** In: *Operations Research Proceedings 1996*. Springer Berlin Heidelberg, pp. 26–30.

— (1997b). **“Proximal Minimization Methods with Generalized Bregman Functions”.** In: *SIAM Journal on Control and Optimization* 35.4, pp. 1142–1168.

Kutateladze, S S and A M Rubinov (1972). **“Minkowski Duality And Its Applications”.** In: *Russ. Math. Surv.* 27.3, pp. 137–191.

Kutateladze, S.S. and A.M. Rubinov (1976). ***Minkowski duality and its applications.*** Nauka Publishers.

Nemirovsky, A.S. and D.B. Yudin (1983). ***Problem Complexity and Method Efficiency in Optimization.*** Trans. by E.R. Dawson. Wiley Interscience. Wiley, New York.

Rubinov, Alexander (2000). ***Abstract Convexity and Global Optimization.*** Springer US.