

# Regret bounds

*and their applications to online learning,  
optimization and games*

Joon Kwon

January 24, 2024

# Contents

<b>1</b>	<b>Convexity tools</b>	<b>4</b>
1.1	Preliminaries . . . . .	4
1.2	Convexity . . . . .	4
1.3	Subgradients . . . . .	6
1.4	Legendre–Fenchel transform . . . . .	7
1.5	Bregman divergences . . . . .	8
1.6	Strong convexity and smoothness . . . . .	9

# Foreword

As of 2024, this document contains lecture notes from a course given in Master 2 in *Université Paris–Saclay*. These are highly incomplete and constantly updated as the lectures are given.

## Acknowledgements

These notes highly benefited from discussions with Sylvain Sorin, Jaouad Mourtada, and the encouragements from Liliane Bel.

# Introduction

# Chapter 1

## Convexity tools

### 1.1 Preliminaries

Let  $d \geq 1$ . Throughout the chapter, we consider Euclidean space  $\mathbb{R}^d$ . For a set  $A \subset \mathbb{R}^d$ ,  $\text{int } A$  and  $\text{cl } A$  denote its interior and closure, respectively.

**Definition 1.1.1** (Domain of a function). The *domain* of a function  $f : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  is the set

$$\left\{x \in \mathbb{R}^d, f(x) < +\infty\right\}.$$

$f$  is said to be *proper* if its domain is nonempty.

**Definition 1.1.2** (Dual norm). Let  $\|\cdot\|$  be a norm in  $\mathbb{R}^d$ . Its *dual norm* is defined as

$$\|y\|_* = \max_{\|x\| \leq 1} \langle y, x \rangle, \quad y \in \mathbb{R}^d.$$

**Proposition 1.1.3.** Let  $\|\cdot\|$  be a norm in  $\mathbb{R}^d$ . Then,  $\|\cdot\|_{**} = \|\cdot\|$ .

**Example 1.1.4** (Common dual norms). In  $\mathbb{R}^d$ ,  $\ell_2$  is its own dual norm,  $\ell_p$  and  $\ell_q$  (with  $p, q \geq 1$  such that  $1/p + 1/q = 1$ ) are dual of each other, and  $\ell_1$  and  $\ell_\infty$  are dual of each other. If  $A$  is a positive definite matrix, the dual norm of the associated Mahalanobis norm  $x \mapsto \sqrt{\langle x, Ax \rangle}$  is the Mahalanobis norm associated with  $A^{-1}$ .

*Remark 1.1.5.* It follows from the definition of the dual norm that for all  $x, y \in \mathbb{R}^d$ ,  $\langle y, x \rangle \leq \|y\|_* \|x\|$ , which, together with the above examples recovers Cauchy-Schwarz and Hölder's inequalities.

### 1.2 Convexity

**Definition 1.2.1.** A set  $\mathcal{X} \subset \mathbb{R}^d$  is convex if for all  $x, x' \in \mathcal{X}$  and  $\lambda \in [0, 1]$ ,  $\lambda x + (1 - \lambda)x' \in \mathcal{X}$ .

**Example 1.2.2** (Simplex). Denote  $\Delta_d$  the simplex in  $\mathbb{R}^d$ :

$$\Delta_d = \left\{ x \in \mathbb{R}_+^d, \sum_{i=1}^d x_i = 1 \right\},$$

which is a closed convex set.

**Definition 1.2.3** (Convex functions). A function  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is *convex* if for all  $x, x' \in \mathbb{R}^d$  and  $\lambda \in [0, 1]$ ,

$$f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x').$$

$f$  is *strictly convex* if the above inequality is strict for  $\lambda \in (0, 1)$ .

*Remark 1.2.4.* The convexity of a function  $f$  is closely related to the convexity sets, as the former can be equivalently defined as all sublevel sets  $\{x \in \mathbb{R}^d, f(x) \leq a\}$  ( $a \in \mathbb{R}$ ) being convex.

**Example 1.2.5** (Norms). Let  $a \geq 1$ . Any norm to the power  $a$  ( $x \mapsto \|x\|^a$ ) is a convex function. In particular, convex functions  $x \mapsto \frac{1}{2} \|\cdot\|_2^2$  and more generally  $x \mapsto \frac{1}{2} \|\cdot\|_p^2$  ( $p \geq 1$ ), as well as squared Mahalanobis norms  $x \mapsto \frac{1}{2} \langle x, Ax \rangle$  (where  $A$  is a positive definite  $d \times d$  matrix) will be used in future chapters.

**Proposition 1.2.6** (First and second order characterizations of convexity). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  a function with open domain.

(i) If  $f$  is differentiable on its domain,  $f$  is convex if, and only if, for all  $x, x' \in \text{dom } f$ ,

$$f(x') \geq f(x) + \langle \nabla f(x), x' - x \rangle.$$

(ii) If  $f$  is twice differentiable on its domain,  $f$  is convex if, and only if, for all  $x \in \text{dom } f$ ,  $\nabla^2 f(x)$  is positive semi-definite.

**Proposition 1.2.7.** Let  $\mathcal{X} \subset \mathbb{R}^d$  be a nonempty convex set, and  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a convex function, differentiable on an open set containing  $\mathcal{X}$ . Then,  $x_* \in \mathcal{X}$  is a minimizer of  $f$  on  $\mathcal{X}$  if, and only if,

$$\forall x \in \mathcal{X}, \langle \nabla f(x_*), x - x_* \rangle \geq 0.$$

**Definition 1.2.8** (Lower semi-continuity). A convex function  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is lower-semicontinuous if for all  $a \in \mathbb{R}$ , the set  $\{x \in \mathbb{R}^d, f(x) \leq a\}$  is closed.

**Example 1.2.9.** Let  $\mathcal{X} \subset \mathbb{R}^d$  be a convex set. The *convex indicator* of  $\mathcal{X}$  is the convex function defined as

$$I_{\mathcal{X}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{X} \\ +\infty & \text{otherwise.} \end{cases}$$

If  $\mathcal{X}$  is closed, then  $I_{\mathcal{X}}$  is also lower-semicontinuous.

**Example 1.2.10** (Negative entropy on the simplex). The function  $h_{\text{ent}} : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  defined as

$$h_{\text{ent}}(x) = \begin{cases} \sum_{i=1}^d x_i \log x_i & \text{if } x \in \Delta_d \\ +\infty & \text{otherwise,} \end{cases}$$

with convention  $0 \log 0 = 0$ , is a lower-semicontinuous convex function.

### 1.3 Subgradients

**Definition 1.3.1** (Subgradients). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $x, y \in \mathbb{R}^d$ .  $y$  is a *subgradient* of  $f$  at  $x$  if for all  $x' \in \mathbb{R}^d$ ,

$$f(x') \geq f(x) + \langle y, x' - x \rangle.$$

The set of all subgradients of  $f$  at  $x$  is called the *subdifferential* of  $f$  at  $x$  and is denoted as  $\partial f(x)$ .

**Example 1.3.2** (Absolute value). For  $f : x \mapsto |x|$  defined on  $\mathbb{R}$ , the subdifferential is given by

$$\partial f(x) = \begin{cases} \{-1\} & \text{if } x < 0 \\ [0, 1] & \text{if } x = 0 \\ \{1\} & \text{if } x > 0. \end{cases}$$

**Proposition 1.3.3.** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ . Then  $x_* \in \mathbb{R}^d$  is a global minimizer of  $f$  if, and only if  $0 \in \partial f(x_*)$ .

*Proof.*  $x_*$  being a global minimizer can be written

$$\forall x \in \mathbb{R}^d, \quad f(x) \geq f(x_*) + \langle 0, x - x_* \rangle,$$

in other words  $0 \in \partial f(x_*)$  □

**Proposition 1.3.4.** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a convex function and  $x \in \text{int dom } f$ . Then,  $f$  is differentiable in  $x$  if, and only if,  $\partial f(x) = \{\nabla f(x)\}$ .

*Remark 1.3.5.* Even in the case of a point in the domain of a convex function, the may be empty. Consider for instance

**Proposition 1.3.6.** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper convex function,  $\|\cdot\|$  a norm on  $\mathbb{R}^d$ , and  $L > 0$ . Then,  $f$  is  $L$ -Lipschitz in  $\text{int dom } f$  with respect to  $\|\cdot\|$  if, and only if:

$$\forall x \in \text{int dom } f, \quad \forall y \in \partial f(x), \quad \|y\|_* \leq L.$$

*Proof.* □

## 1.4 Legendre–Fenchel transform

**Definition 1.4.1.** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper function. The Legendre–Fenchel transform of  $f$  is a function  $f^* : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  defined as

$$f^*(y) = \sup_{x \in \mathbb{R}^d} \{\langle y, x \rangle - f(x)\}, \quad y \in \mathbb{R}^d.$$

**Proposition 1.4.2** (Fenchel’s inequality). *For all  $x, y \in \mathbb{R}^d$ ,  $\langle y, x \rangle \leq f(x) + f^*(y)$ .*

The above definition is somewhat abstract, and it may be insightful to informally examine a simple example in dimension 1 by decomposing the transformation into simpler steps. Consider  $f(x) = \frac{a}{2}(x - b)^2$  for some  $a, b \neq 0$ , which is a differentiable convex function with finite values on  $\mathbb{R}$ . Its derivative is given by  $f'(x) = a(x - b)$ , which is (strictly) increasing (as would be the case as soon as  $f$  is differentiable and strictly convex), and is therefore a bijection from its domain to its range (from  $\mathbb{R}$  to  $\mathbb{R}$  in this case). Then, the inverse  $(f')^{-1}$  is also an increasing function:  $(f')^{-1}(y) = y/a + b$ . We then consider the following primitive function:

$$f^*(y) = \int_0^y (f')^{-1} - \min_{x \in \mathbb{R}} f(x).$$

As the primitive of an increasing function,  $f^*$  is also convex. It is easy to see that the operation is involutorial:  $f^{**} = f$ . An intuition that appears in this example is that the more  $f$  is curved, the less  $f^*$  is so, and vice-versa. The derivatives being inverses of each other can be interpreted as follows:  $f$  has slope  $y$  at point  $x$  if, and only if,  $f^*$  has slope  $x$  at point  $y$ . In higher dimension, this would correspond to the usual duality between points and hyperplanes. The Legendre–Fenchel transform in fact generalizes the above for higher dimension, and for nondifferentiable functions. The neatest properties of e.g. Propositions 1.4.3 and 1.4.4 below are obtained for the class of proper lower-semicontinuous convex functions. The derivative, which is a function, is then replaced by the subdifferential, which is a *correspondance*.

**Proposition 1.4.3.** *Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper function.*

- (i)  $f^*$  is proper, lower-semicontinuous and convex.
- (ii) If  $f$  is proper, lower-semicontinuous and convex, then  $f^{**} = f$ .

**Proposition 1.4.4.** *Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower-semicontinuous convex function and  $x, y \in \mathbb{R}^d$ . The following statements are equivalent:*

- (i)  $x \in \partial f^*(y)$ ,
- (ii)  $y \in \partial f(x)$ ,



- (iii)  $\langle y, x \rangle = f(x) + f^*(y),$
- (iv)  $x \in \text{Arg max}_{x' \in \mathbb{R}^d} \{\langle y, x' \rangle - f(x')\},$
- (v)  $y \in \text{Arg max}_{y' \in \mathbb{R}^d} \{\langle y', x \rangle - f^*(y')\}.$

**Example 1.4.5** (Dual norms). Let  $\|\cdot\|$  be a norm in  $\mathbb{R}^d$  and denote  $B$  its closed unit ball. Then,  $I_B$  is then a proper lower-semicontinuous convex function and  $\|\cdot\|_* = I_B^*$ . Therefore, the involutorial property of the Legendre-Fenchel transform is an extension of the involutorial property for dual norms.

## 1.5 Bregman divergences

We now define a large class of similarity measures in  $\mathbb{R}^d$  called Bregman divergences, which in general are not distances because they may fail to be symmetric. They contain the squared Euclidean norm and the Kullback–Leibler divergence as special cases. Bregman divergences are used as an alternative geometry to the Euclidean one when it comes to e.g. defining and analyzing iterative algorithms.

**Definition 1.5.1.** Let  $\mathcal{X} \subset \mathbb{R}^d$ ,  $f : \mathcal{X} \rightarrow \mathbb{R}$  and  $x, x' \in \mathcal{X}$  such that  $f$  is differentiable in  $x$ . Then, the *Bregman divergence* from  $x$  to  $x'$  is defined as

$$D_f(x', x) = f(x') - f(x) - \langle \nabla f(x), x' - x \rangle.$$

*Remark 1.5.2.*  $D_f(x, x')$  is the remainder of the second order Taylor's expansion from  $x$  to  $x'$ , and is a measure of the curvature of  $f$  between those two points. The Bregman divergence is nonnegative as soon as  $f$  is convex. In the case of a linear function  $f$ , the Bregman divergence is zero, which corresponds to the linear function having no curvature.

The above definition which requires the differentiability at starting point  $x$  is the most common. For our purposes however,

**Definition 1.5.3** (Generalized Bregman divergences). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a function,  $x, x', y \in \mathbb{R}^d$  such that  $x \in \text{dom } f$  and  $y \in \partial f(x)$ . The *Bregman divergence* from  $x$  to  $x'$  with subgradient  $y$  is then defined as

$$D_f(x', x; y) = f(x') - f(x) - \langle y, x' - x \rangle.$$

*Remark 1.5.4.*

**Proposition 1.5.5.** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower-semicontinuous convex function and  $x, x', y, y' \in \mathbb{R}^d$  such that  $y \in \partial f(x)$  and  $y' \in \partial f(x')$ . Then,

$$D_f(x', x; y) = D_{f^*}(y, y'; x').$$

*Proof.*

$$\begin{aligned}
D_f(x', x; y) - D_{f^*}(y, y'; x') &= f(x') - f(x) - \langle y, x' - x \rangle \\
&\quad - f^*(y) + f^*(y') + \langle x', y - y' \rangle \\
&= \langle x', y' \rangle - \langle x, y \rangle - \langle y, x' - x \rangle + \langle x', y - y' \rangle \\
&= 0,
\end{aligned}$$

where for the second equality, we applied Fenchel's identity from property (iii) in Proposition 1.4.4.  $\square$

**Example 1.5.6** (Square Euclidean distance).

**Example 1.5.7** (Kullback–Leibler divergence).

## 1.6 Strong convexity and smoothness

**Definition 1.6.1** (Strong convexity). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $\|\cdot\|$  a norm in  $\mathbb{R}^d$  and  $K > 0$ .  $f$  is  $K$ -strongly convex with respect to  $\|\cdot\|$  if for all  $x, x' \in \mathbb{R}^d$  and  $\lambda \in [0, 1]$ ,

$$f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x') - \frac{\lambda(1 - \lambda)}{2} \|x' - x\|^2.$$

**Definition 1.6.2** (Smoothness). Let  $\Omega \subset \mathbb{R}^d$  be an open set and  $f : \Omega \rightarrow \mathbb{R}$  be a differentiable function,  $\|\cdot\|$  a norm in  $\mathbb{R}^d$  and  $L > 0$ .  $f$  is  $L$ -smooth with respect to  $\|\cdot\|$  if for all  $x, x' \in \Omega$ ,  $|D_f(x', x)| \leq \frac{K}{2} \|x' - x\|$ , in other words,

$$|f(x') - f(x) - \langle \nabla f(x), x' - x \rangle| \leq \frac{K}{2} \|x' - x\|^2.$$

**Proposition 1.6.3** (Characterization of smoothness). Let  $\Omega \subset \mathbb{R}^d$  be an open set,  $f : \Omega \rightarrow \mathbb{R}$  a differentiable function and  $L > 0$ .

- (i)  $f$  is  $L$ -smooth if, and only if,  $\nabla f$  is  $L$ -Lipschitz with respect to  $\|\cdot\|$ .
- (ii) Moreover, if  $f$  is twice differentiable,  $f$  is  $L$ -smooth if, and only if, for all  $u \in \mathbb{R}^d$ ,

$$|\langle u, \nabla^2 f(x)u \rangle| \leq L \|u\|^2.$$

**Proposition 1.6.4** (Duality between strong convexity and smoothness). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a convex function,  $\|\cdot\|$  a norm in  $\mathbb{R}^d$  and  $K > 0$ . The following statements are equivalent.

- (i)  $f$  is  $K$ -strongly convex with respect to  $\|\cdot\|$ .
- (ii) For all  $x, x', y \in \mathbb{R}^d$  such that  $y \in \partial f(x)$ ,  $D_f(x', x; y) \geq \frac{1}{2} \|x' - x\|^2$ , on other words

$$f(x') \geq f(x) + \langle y, x' - x \rangle + \frac{K}{2} \|x' - x\|^2.$$

(iii) For all  $x, x', y, y' \in \mathbb{R}^d$  such that  $y \in \partial f(x)$  and  $y' \in \partial f(x')$ ,

$$\langle y' - y, x' - x \rangle \geq K \|x' - x\|^2.$$

(iv)  $f^*$  is differentiable and  $\nabla f^*$  is  $1/K$ -Lipschitz with respect to  $\|\cdot\|_*$  and  $\|\cdot\|$ , in other words, for all  $y, y' \in \mathbb{R}^d$ ,

$$\|\nabla f^*(y') - \nabla f^*(y)\| \leq \frac{1}{K} \|y' - y\|_*.$$

(v) For all  $x, x', y, y' \in \mathbb{R}^d$  such that  $y \in \partial f(x)$  and  $y' \in \partial f(x')$ ,

$$\langle y' - y, x' - x \rangle \leq \frac{1}{K} \|y' - y\|_*^2.$$

(vi)  $f^*$  is differentiable and  $1/K$ -smooth with respect to  $\|\cdot\|_*$ .

*Proof.*

□