

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

Joon Kwon

January 27, 2024

Contents

1	Convexity tools	4
1.1	Preliminaries	4
1.2	Convexity	5
1.3	Subgradients	6
1.4	Legendre–Fenchel transform	7
1.5	Bregman divergences	9
1.6	Strong convexity and smoothness	10

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.

The course proposes a unified presentation of regret minimization for adversarial online learning problems, and its application to various problems such as Blackwell’s approachability, optimization algorithms (GD, Nesterov, SGD, AdaGrad), variational inequalities with monotone operators (Extra-gradient, Mirror-Prox, Dual Extrapolation), fixed-point iterations (Krasnoselskii–Mann), and games. The presentation aims at being modular, so that introduced tools and techniques could easily be used to define and analyze new algorithms.

The central notion of this presentation is the *regret*, which will be analyzed using the Legendre–Fenchel transform and Bregman divergences. An excellent recent monograph on the topic of online learning is the following:

- Francesco Orabona. A modern introduction to online learning. *arXiv:1912.13213*, 2023.

Additional notable references on the topic include:

- H. Brendan McMahan. A survey of algorithms and analysis for adaptive online learning. *The Journal of Machine Learning Research*, 18(1):3117–3166, 2017,
- Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016,
- Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011,
- Sébastien Bubeck. *Introduction to Online Optimization: Lecture Notes*. Princeton University, 2011,
- Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Regarding convex analysis, we refer to the following classical book:

-
- R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

Acknowledgements

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

Chapter 1

Convexity tools

We present the basic convexity notions and tools that will be used in the subsequent chapters. Most of them are classical and are given without proof.

1.1 Preliminaries

Let $d \geq 1$. Throughout the chapter, we consider Euclidean space \mathbb{R}^d equipped with its canonical inner product. 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

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

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

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

Remark 1.1.3. The above maximum is indeed attained because for a given $y \in \mathbb{R}^d$, function $x \mapsto \langle y, x \rangle$ is continuous on the closed unit ball, which is compact. Besides, one can check that the dual norm is indeed a norm.

Proposition 1.1.4. Let $\|\cdot\|$ be a norm in \mathbb{R}^d . Then, $\|\cdot\|_{**} = \|\cdot\|$.

Example 1.1.5 (Common dual norms). In \mathbb{R}^d , ℓ_2 is its own dual norm, ℓ_p and ℓ_q (with $p, q \geq 1$ such that $1/p + 1/q = 1$) are dual of each other, and ℓ_1 and ℓ_∞ 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 1.1.6. 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 (Unit balls). For all norms, as an immediate consequence of the triangle inequality, the unit ball is convex.

Example 1.2.3 (Simplex). Denote Δ_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. Note that it is contained in a hyperplane and therefore has empty interior.

Definition 1.2.4 (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 1.2.5. The convexity of a function f is closely related to the convexity of sets, as the former can be equivalently defined as the epigraph

$$\text{epi } f = \left\{ (a, x) \in \mathbb{R} \times \mathbb{R}^d, a \geq f(x) \right\}$$

being convex.

Example 1.2.6. The following functions are convex: linear functions, quadratic functions of the form $x \mapsto \langle x, Ax \rangle$ where A is a positive semi-definite matrix, the exponential, the negative logarithm, convex combinations of convex functions, the point-wise supremum of convex functions. Let $\|\cdot\|$ be a norm in \mathbb{R}^d and $a \geq 1$. Function $x \mapsto \|x\|^a$ is convex.

Proposition 1.2.7 (Jensen's inequality). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex function, and X a random variable with values in $\text{dom } f$ so that $\mathbb{E}[X]$ exists. Then, f is measurable, $\mathbb{E}[f(X)]$ exists in $\mathbb{R} \cup \{+\infty\}$ and*

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

In particular, for $n \geq 1$, $x_1, \dots, x_n \in \mathbb{R}^d$ and $\lambda_1, \dots, \lambda_n \geq 0$ such that $\lambda_1 + \dots + \lambda_n = 1$,

$$f\left(\frac{\sum_{i=1}^n \lambda_i x_i}{\sum_{i=1}^n \lambda_i}\right) \leq \frac{\sum_{i=1}^n \lambda_i f(x_i)}{\sum_{i=1}^n \lambda_i}.$$

Proposition 1.2.8 (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 1.2.9. *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 1.2.10 (Lower semi-continuity). *A function $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower-semicontinuous if for all $a \in \mathbb{R}^d$, the set $\{x \in \mathbb{R}^d, f(x) \leq a\}$ is closed.*

Example 1.2.11. Let $\mathcal{X} \subset \mathbb{R}^d$ be a convex set. The *convex indicator* of \mathcal{X} is the convex function defined as

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

which is lower-semicontinuous if, and only if \mathcal{X} is closed.

Example 1.2.12 (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. It will also be called the *entropic regularizer*.

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 $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$. A point $x_* \in \mathbb{R}^d$ 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 $x \in \text{int dom } f$. f is differentiable in x if, and only if, $\partial f(x)$ is a singleton. When this is the case, $\partial f(x) = \{\nabla f(x)\}$.

Remark 1.3.5. Even in the case of a point in the domain of a convex function, the subdifferential may be empty. Consider for instance $f : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ defined as:

$$f(x) = \begin{cases} -\sqrt{x} & \text{if } x \geq 0 \\ +\infty & \text{if } x < 0, \end{cases}$$

which is a proper lower-semicontinuous convex function. 0 belongs to the domain of f and yet, $\partial f(0) = \emptyset$.

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:

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

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 (or convex conjugate) of f is a function $f^* : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ defined as

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

Remark 1.4.2. In the above definition, for a given $y \in \mathbb{R}^d$, because f is assumed proper, quantity $\langle y, x \rangle - f(x)$ is not $-\infty$ for at least some point $x \in \mathbb{R}^d$, and therefore, the supremum is indeed a value in $\mathbb{R} \cup \{+\infty\}$.

Remark 1.4.3 (Fenchel’s inequality). It follows from the above definition that 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 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:

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

which can be proved to be an alternative definition of the Legendre–Fenchel transform in this special case (although the proof is somewhat involved). As the primitive of an increasing function, f^* is also convex. Because of the inverse relation between the derivatives, this transformation $f \mapsto f^*$ is involutorial up a constant: $f^{**} = f + a$, ($a \in \mathbb{R}$). Then, it can be proved that $a = 0$, which can also be noticed graphically. 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 builds on the usual duality between points and hyperplanes. The Legendre–Fenchel transform indeed generalizes the above to higher dimensions, and for nondifferentiable functions. The neatest properties of e.g. Propositions 1.4.5 and 1.4.6 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 *correspondence*.

Proposition 1.4.4. *If $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper function, f^* is lower-semicontinuous and convex.*

Theorem 1.4.5 (Fenchel–Moreau). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper function. Then, f is lower-semicontinuous and convex if, and only if $f = f^{**}$. In this case, f^* is proper.*

Proposition 1.4.6. *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 1.4.7 (Norms and squared 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 $I_B^* = \|\cdot\|_*$. Therefore, the involutonal property of the Legendre-Fenchel transform is an extension of the involutonal property for dual norms. Besides, if $f : x \mapsto \frac{1}{2} \|x\|^2$, then for $y \in \mathbb{R}^d$, $f^*(y) = \frac{1}{2} \|y\|_*^2$.

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. We present the classical definition which involves a gradient.

Definition 1.5.1. Let $\mathcal{X} \subset \mathbb{R}^d$, $f : \mathcal{X} \rightarrow \mathbb{R}$, $x \in \text{int } \mathcal{X}$ and $x' \in \mathcal{X}$ 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 first 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, we also consider the following generalization (proposed in [JKM23]) which involves a subgradient instead of a gradient.

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. The above generalized Bregman divergence may not exist even when x belongs to the domain of f , as the subdifferential may be empty.

When it exists, because $x \in \text{dom } f$, it belongs to $\mathbb{R} \cup \{+\infty\}$ and is nonnegative when f is convex. When f is convex and differentiable at point

x , the only subgradient at x is $\nabla f(x)$ according to Proposition 1.3.4, and the two previous definitions coincide.

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 1.4.6. \square

Example 1.5.6 (Squared Euclidean norm). Consider the squared Euclidean norm $h_2 : x \mapsto \frac{1}{2} \|x\|_2^2$, which is differentiable in \mathbb{R}^d . Then for all $x, x' \in \mathbb{R}^d$, $D_h(x', x) = \frac{1}{2} \|x' - x\|_2^2$.

1.6 Strong convexity and smoothness

We now introduce strongly convexity and smoothness which intuitively correspond to the curvature of a function being respectively bounded from below (by a positive number), and bounded from above in absolute value.

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 $\lambda \in [0, 1]$,

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

Definition 1.6.2 (Smoothness). Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $x, x' \in \mathbb{R}^d$, $|D_f(x', x)| \leq \frac{L}{2} \|x' - x\|^2$, in other words,

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

Remark 1.6.3. If f is convex, the above definition reduces to $D_f(x', x) \leq \frac{L}{2} \|x' - x\|_2^2$ ($x, x' \in \mathbb{R}^d$).

Proposition 1.6.4 (Duality between strong convexity and smoothness). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, lower-semicontinuous 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) For all $x, x', y, y' \in \mathbb{R}^d$ such that $y \in \partial f(x)$ and $y' \in \partial f(x')$,

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

(v) For all $x, x', y, y' \in \mathbb{R}^d$ such that $y \in \partial f(x)$ and $y' \in \partial f(x')$, $D_f(x', x; y) \leq \frac{1}{2K} \|y' - y\|_*^2$, in other words

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

(vi) f^* is differentiable on \mathbb{R}^d and $1/K$ -smooth with respect to $\|\cdot\|_*$.

Proposition 1.6.5 (Second order characterization of strong convexity). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a twice differentiable function, $\|\cdot\|$ a norm in \mathbb{R}^d and $K > 0$. Then, f is K -strongly convex with respect to $\|\cdot\|$ if, and only if,*

$$\forall x \in \mathbb{R}^d, \forall u \in \mathbb{R}^d, \quad \langle u, \nabla^2 f(x) u \rangle \geq K \|u\|^2.$$

Proposition 1.6.6 (First and second order characterizations of smoothness). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a differentiable function, $\|\cdot\|$ a norm in \mathbb{R}^d and $L > 0$.*

(i) f is L -smooth if, and only if, ∇f is L -Lipschitz with respect to $\|\cdot\|$.

(ii) Moreover, if f is twice differentiable, f is L -smooth with respect to $\|\cdot\|$ if, and only if,

$$\forall x \in \mathbb{R}^d, \forall u \in \mathbb{R}^d, \quad |\langle u, \nabla^2 f(x) u \rangle| \leq L \|u\|^2.$$

Corollary 1.6.7. *The squared Euclidean norm $h_2 : x \mapsto \frac{1}{2} \|x\|_2^2$ is 1-strongly convex and 1-smooth with respect to $\|\cdot\|_2$.*

Proposition 1.6.8. *For $p \in (1, 2)$, the squared ℓ_p norm $h_p : x \mapsto \frac{1}{2} \|x\|_p^2$ is $(p-1)$ -strongly convex with respect to ℓ_p .*

Proposition 1.6.9. *The negative entropy h_{ent} is 1-strongly convex with respect to ℓ_1 .*

Bibliography

- [Bub11] Sébastien Bubeck. *Introduction to Online Optimization: Lecture Notes*. Princeton University, 2011.
- [CBL06] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
- [Haz16] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.
- [JKM23] Anatoli Juditsky, Joon Kwon, and Éric Moulines. Unifying mirror descent and dual averaging. *Mathematical Programming*, (199):793–830, 2023.
- [McM17] H. Brendan McMahan. A survey of algorithms and analysis for adaptive online learning. *The Journal of Machine Learning Research*, 18(1):3117–3166, 2017.
- [Ora23] Francesco Orabona. A modern introduction to online learning. *arXiv:1912.13213*, 2023.
- [Roc70] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [SS11] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.