ECE 273 Convex Optimization and Applications — Instructor: Jun-Kun Wang
Scribe: Jonah Soong, Tony Guan, Younus Ahmad — April 16, 2024
Editor/TA: Marialena Sfyraki

Lecture 5: Convex Analysis II: Subgradient and Optimality Conditions of Constrained Optimization

1 Subgradients and Subdifferentials

Definition 1. *(**Subgradient**) For a function $f(\cdot)$, we say $g_{\mathbf{x}}$ is a subgradient of $f(\cdot)$ at $\mathbf{x} \in$ dom f, if for **all** $\mathbf{y}$*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle g_{\mathbf{x}}, \mathbf{y} - \mathbf{x} \rangle$$

Fact 1: If $f(\cdot)$ is convex, a subgradient at any $\mathbf{x} \in$ dom f exists.
Fact 2: When f is convex and differentiable, $g_{\mathbf{x}} = \nabla f(\mathbf{x})$.

Example: $f(x) = |x|$. Letting g_0 be the subgradient at x = 0. we have $g_0 \in [-1, 1]$.

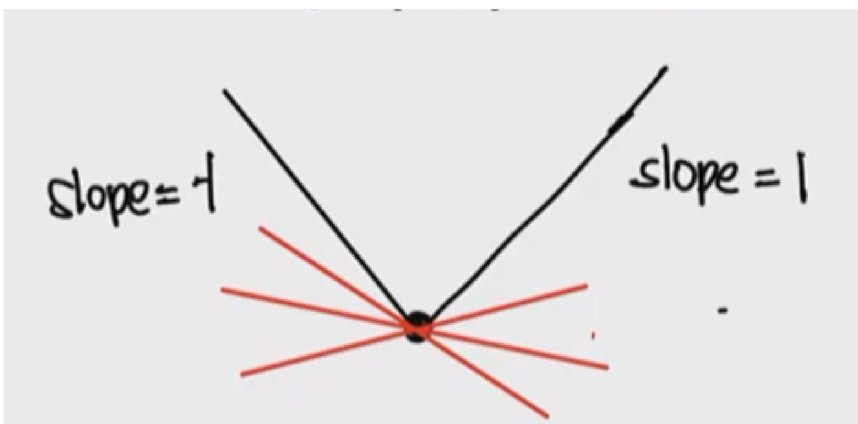

Figure 1: Example Subgradients for $f(x) = |x|$.

Definition 2. *(**Subdifferential**) The subdiffenential at a point $\mathbf{x} \in$ dom f is the set of subgradient at $\mathbf{x} \in$ dom f:*

$$\partial f(\mathbf{x}) := \{g_{\mathbf{x}} : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle g_{\mathbf{x}}, \mathbf{y} - \mathbf{x} \rangle, \textit{ for all } \mathbf{y}\}$$

Properties of subdifferential $\partial f(\mathbf{x})$:

- When $f(\cdot)$ is convex, $\partial f(\mathbf{x})$ is nonempty.

- $\partial f(\mathbf{x})$ is closed and convex. Informally, a closed set is one with a boundary. For example, [0, 1] is a closed set but (0,1) is not.
- When f is convex and differentiable, $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$, i.e., has unique element.

Proof. **(Subdifferential is a convex set)**

$$\partial f(\mathbf{x}) := \{g_{\mathbf{x}} : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle g_{\mathbf{x}}, \mathbf{y} - \mathbf{x} \rangle, \text{ for all } \mathbf{y}\}$$

$$= \cap_{\mathbf{y}} \{g_{\mathbf{x}} : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle g_{\mathbf{x}}, \mathbf{y} - \mathbf{x} \rangle\}.$$

Because it is the intersection of sets of values $g_{\mathbf{x}}$ that satisfy the inequality for a given $\mathbf{y}$, the subdifferential set can be represented as an intersection of half spaces. In other words, the intersection gives us values $g_{\mathbf{x}}$ that satisfy the inequality for all values $\mathbf{y}$. We know an intersection of half spaces is convex, so the subdifferential is a convex set. □

2 Convex Sets

Definition 3. ***(Convex Sets)****: A set $C \subseteq \mathbb{R}^d$ is called convex if for any $\mathbf{x}, \mathbf{y} \in C$ and any $\alpha \in [0, 1]$, we have*

$$\alpha \mathbf{x} + (1 - \alpha)\mathbf{y} \in C.$$

This is to say that the for any $\mathbf{x}, \mathbf{y} \in C$ all convex combinations of these two points are within the set. Geometrically we can visualize this as a line connecting the two points, for any points in C. If all points along this line are within the set C then the set is considered to be convex.

Figure 2: Examples of Convex Set (Left) vs Non-Convex Set (Right)

2.1 Examples of Convex Sets

Example 1: The vector space $\mathbb{R}^d$.

Vector spaces are closed under vector addition and scalar multiplication. Because of this, the convex combination of two vectors is still a vector in $\mathbb{R}^d$.

Example 2: Hyper-planes

$$\{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{x}, \mathbf{a} \rangle = c\},\ \mathbf{a} \in \mathbb{R}^d,\ c \in \mathbb{R}.$$

From the first example, we know every convex combination of $\mathbf{x}$ is a vector in $\mathbb{R}^d$. We will show a convex combination of elements in the hyper-plane set is also in that set. Consider two elements $\mathbf{x}$, $\mathbf{y}$ such that $\langle \mathbf{x}, \mathbf{a} \rangle = c$ and $\langle \mathbf{y}, \mathbf{a} \rangle = c$

$$\begin{aligned}
&\langle \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \rangle \\
&= \langle \alpha\mathbf{x}, \mathbf{a} \rangle + \langle \mathbf{y}, a \rangle - \langle \alpha\mathbf{y}, \mathbf{a} \rangle \\
&= \alpha c + c - \alpha c \\
&= c
\end{aligned}$$

The above condition for a convex combination of elements is still satisfied and are in the hyper-plane set, thus hyper-planes are considered convex sets.

Example 3: Half-spaces

$$\{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{x}, \mathbf{a} \rangle \leq c\},\ \mathbf{a} \in \mathbb{R}^d,\ c \in \mathbb{R}$$

From the first example, we know every convex combination of $\mathbf{x}$ is a vector in $\mathbb{R}^d$. We will show a convex combination of elements in the hyper-plane set is also in that set. Consider two elements $\mathbf{x}$, $\mathbf{y}$ such that $\langle \mathbf{x}, \mathbf{a} \rangle \leq c$ and $\langle \mathbf{y}, \mathbf{a} \rangle \leq c$

$$\begin{aligned}
&\langle \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \rangle \\
&= \langle \alpha\mathbf{x}, \mathbf{a} \rangle + \langle \mathbf{y}, \mathbf{a} \rangle - \langle \alpha\mathbf{y}, \mathbf{a} \rangle \\
&\leq \alpha c + c - \alpha c \\
&\leq c
\end{aligned}$$

The above condition for a convex combination of elements is still satisfied and are in the half space, thus half spaces are considered convex sets.

Example 4: A norm ball with a radius $r \geq 0$

$$\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq r\}.$$

Applying the triangular inequality:

$$
\begin{aligned}
||\alpha\mathbf{x} + (1-\alpha)\mathbf{y}|| &\leq \alpha||\mathbf{x}|| + (1-\alpha)||\mathbf{y}|| \\
&\leq \alpha r + (1-\alpha) r \\
&\leq r
\end{aligned}
$$

Thus the convex combination of $\mathbf{x}, \mathbf{y}$ is also in the norm ball making this a convex set.

Example 5: (Convex Hull) A convex hull of a set C, denoted as **conv** C, is the set of **all** convex combinations of the points in C.

$$
\mathbf{conv} C := \left\{ \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_n \mathbf{x}_n | \text{ each } \mathbf{x}_i \in C, \alpha_i \geq 0, \sum_i \alpha_i = 1 \right\}.
$$

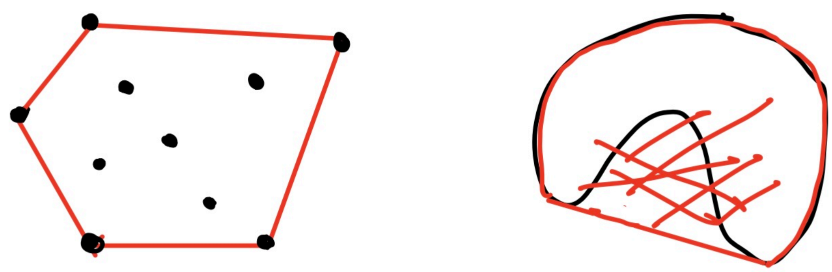

Figure 3: Example Convex Hulls

Note: The convex hull of any (non-convex) set is a convex set, like shown in Figure 3.

Fact: Intersection preserves the convexity. If C_1 and C_2 are convex sets, then so is $C_1 \cap C_2$. Let

$$
\begin{aligned}
\mathbf{x} &\in C_1 \cap C_2, \\
\mathbf{y} &\in C_1 \cap C_2,
\end{aligned}
$$

where $C_1 \cap C_2 \subseteq C_1$ and $C_1 \cap C_2 \subseteq C_2$. Then,

$$
\begin{aligned}
\alpha\mathbf{x} + (1-\alpha)\mathbf{y} &\in C_1, \\
\alpha\mathbf{x} + (1-\alpha)\mathbf{y} &\in C_2.
\end{aligned}
$$

Therefore,

$$\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C_1 \cap C_2.$$

Thus, convexity is preserved under intersection.

3 Optimality conditions of Constrained Convex Optimization

Theorem 1. *Assume f is a convex function and* ***differentiable****, then*

$$\mathbf{x}_* \in \arg\min_{\mathbf{x}\in C} f(\mathbf{x}) \tag{1}$$

iff *for any $\mathbf{y} \in C$*

$$\langle \nabla f(\mathbf{x}_*), \mathbf{y} - \mathbf{x}_* \rangle \geq 0 \tag{2}$$

Futhermore, if

$$C \equiv \mathbb{R}^d,$$

then

$$\begin{aligned} &\langle \nabla f(\mathbf{x}_*), \mathbf{y} - \mathbf{x}_* \rangle \geq 0, \forall \mathbf{y} \in \mathbb{R}^d \\ &\equiv \nabla f(\mathbf{x}_*) = 0 \end{aligned}$$

Proof (Theorem 1). First, we prove $2 \to 1$. By convexity, we know

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{x}_*) + \langle \nabla f(\mathbf{x}_*), \mathbf{y} - \mathbf{x}_* \rangle, \\ &\geq f(\mathbf{x}_*), \quad \forall \mathbf{y} \in C \quad \text{, by condition 2} \\ &\Rightarrow \mathbf{x}_* \in \arg\min_{\mathbf{x}\in C} f(\mathbf{x}) \end{aligned}$$

Now, we prove $1 \to 2$. This is equivalent to showing the contrapositive, i.e. when $\langle \nabla f(\mathbf{x}_*), \mathbf{y} - \mathbf{x}_* \rangle < 0$ for some $\mathbf{y} \in C$, then there exists $\mathbf{z} \in C$ such that

$$f(\mathbf{z}) < f(\mathbf{x}_*), \text{i.e. } \mathbf{x}_* \notin \arg\min_{\mathbf{x}\in C} f(\mathbf{x}).$$

Denote $\mathbf{x}_\alpha := \alpha\mathbf{y} + (1-\alpha)\mathbf{x}_*$ and denote $h(\alpha) := f(\mathbf{x}_\alpha)$, where

$$h : \mathbb{R} \to \mathbb{R},$$

$$f : \mathbb{R}^d \to \mathbb{R}.$$

We have that

$$h'(\alpha) = \langle \nabla f(\mathbf{x}_\alpha), \mathbf{y} - \mathbf{x}_* \rangle.$$

Observe that

$$\alpha = 0 \Rightarrow \mathbf{x}_\alpha = \mathbf{x}_*.$$

Then,

$$\lim_{\delta \to 0} \frac{h(0+\delta) - h(0)}{\delta} = h^{'}(0) = \langle \nabla f(\mathbf{x}_*), \mathbf{y} - \mathbf{x}_* \rangle < 0.$$

But $h(0+\delta) - h(0) = f(\mathbf{x}_\delta) - f(\mathbf{x}_*)$. Thus, we have that $f(\mathbf{x}_\delta) < f(\mathbf{x}_*)$, which contradicts the definition of the $\mathbf{x}_*$, completing our proof. □

Theorem 2. *Assume f is a convex function, then*

$$\mathbf{x}_* \in \arg\min_{\mathbf{x} \in C} f(\mathbf{x}) \tag{3}$$

iff *there exists a subgradient $g_{\mathbf{x}_*}$ such that for any $\mathbf{y} \in C$*

$$\langle g_{\mathbf{x}_*}, \mathbf{y} - \mathbf{x}_* \rangle \geq 0 \tag{4}$$

Proof (Theorem 2). First, we prove $4 \to 3$. Let $\mathbf{y} \in C$. Then,

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{x}_*) + \langle g_{\mathbf{x}_*}, \mathbf{y} - \mathbf{x}_* \rangle \\ &\geq f(\mathbf{x}_*), \quad \forall \mathbf{y} \in C \quad, \text{ by condition 4} \end{aligned}$$

□

In order to prove $3 \to 4$, we will need more toolikt. First, we will revisit the definition of directional derivative and its relation to subdifferential $\partial f(\mathbf{x})$.

Definition 4. *(**Directional Derivative**) For convex f, at any point $\mathbf{x} \in dom\ f$, and any $\mathbf{u} \in \mathbb{R}^d$, the directional derivative $f^{'}(\mathbf{x}; \mathbf{u})$ exists and is*

$$f^{'}(\mathbf{x}; \mathbf{u}) := \inf_{\alpha > 0} \frac{f(\mathbf{x} + \alpha \mathbf{u}) - f(\mathbf{x})}{\alpha}.$$

Let us define $h(\alpha) : \alpha \to \frac{f(\mathbf{x}+\alpha\mathbf{u}) - f(\mathbf{x})}{\alpha}$. We will use the following lemma:

Lemma 1. *When $f(\cdot)$ is convex, $h(\alpha)$ is increasing w.r.t. α.*

Proof (Lemma 1). Consider $0 \leq \alpha_1 \leq \alpha_2$

$$\frac{f(\mathbf{x} + \alpha_1 \mathbf{u}) - f(\mathbf{x})}{\alpha_1} = \frac{\alpha_2}{\alpha_1} \frac{f(\mathbf{x} + \alpha_2(\alpha_1/\alpha_2)\mathbf{u}) - f(\mathbf{x})}{\alpha_2} \tag{5}$$

$$= \frac{\alpha_2}{\alpha_1} \frac{f\left((1 - \alpha_1/\alpha_2)\mathbf{x} + (\alpha_1/\alpha_2)\left(\mathbf{x} + \alpha_2 \mathbf{u}\right)\right) - f(\mathbf{x})}{\alpha_2} \tag{6}$$

$$\leq \frac{\alpha_2}{\alpha_1} \frac{(1-\alpha_1/\alpha_2) f(\mathbf{x}) + (\alpha_1/\alpha_2) f(\mathbf{x}+\alpha_2\mathbf{u}) - f(\mathbf{x})}{\alpha_2} \tag{7}$$

$$= \frac{f(\mathbf{x}+\alpha_2\mathbf{u}) - f(\mathbf{x})}{\alpha_2} \tag{8}$$

□

Implication. *If $f(\cdot)$ is convex, then every local minimum is also a global minimum.*

Recall, that by definition of the directional derivative we have

$$f'(\mathbf{x};\mathbf{u}) := \inf_{\alpha>0} \frac{f(\mathbf{x}+\alpha\mathbf{u}) - f(\mathbf{x})}{\alpha}.$$

The following theorem demonstrates the connection between subdifferential $\partial f(\mathbf{x})$ and directional derivative.

Theorem 3. *Let $f(\cdot)$ be closed and convex. Then,*

$$f'(\mathbf{x};\mathbf{u}) = \sup_{g_{\mathbf{x}} \in \partial f(\mathbf{x})} \langle g_{\mathbf{x}}, \mathbf{u} \rangle.$$

Proof (Theorem 3). See Section 2.4 of [Duchi (2010)] for the proof. □

Let's try to make sense this relation. We have that for any subgradient $g_{\mathbf{x}} \in \partial f(\mathbf{x})$,

$$f(\mathbf{x}+\alpha\mathbf{u}) \geq f(\mathbf{x}) + \alpha\langle g_{\mathbf{x}}, \mathbf{u} \rangle.$$

Subtracting $f(\mathbf{x})$ from both sides and dividing by $\alpha > 0$ gives that

$$\frac{f(\mathbf{x}+\alpha\mathbf{u}) - f(\mathbf{x})}{\alpha} \geq \langle g_{\mathbf{x}}, \mathbf{u} \rangle.$$

Therefore,

$$\frac{f(\mathbf{x}+\alpha\mathbf{u}) - f(\mathbf{x})}{\alpha} \geq \sup_{g_{\mathbf{x}} \in \partial f(\mathbf{x})} \langle g_{\mathbf{x}}, \mathbf{u} \rangle.$$

Thus, for any $g_{\mathbf{x}} \in \partial f(\mathbf{x})$, we have $g_{\mathbf{x}} \in \{g_{\mathbf{x}} : \langle g_{\mathbf{x}}, \mathbf{u} \rangle \leq f'(\mathbf{x};\mathbf{u}), \forall \mathbf{u}\}$. Conversely, let $g_{\mathbf{x}} \in \{g_{\mathbf{x}} : \langle g_{\mathbf{x}}, \mathbf{u} \rangle \leq f'(\mathbf{x};\mathbf{u}), \forall \mathbf{u}\}$. By the increasing slopes condition, we have

$$\langle g_{\mathbf{x}}, \mathbf{u} \rangle \leq f'(\mathbf{x};\mathbf{u}) \leq \frac{f(\mathbf{x}+\alpha\mathbf{u}) - f(\mathbf{x})}{\alpha},$$

for all $\mathbf{u}$ and $\alpha > 0$. Taking $\alpha = 1$, we get the standard inequality

$$f(\mathbf{x} + \alpha\mathbf{u}) \geq f(\mathbf{x}) + \alpha\langle g_{\mathbf{x}}, \mathbf{u}\rangle.$$

Thus, if $g_{\mathbf{x}} \in \{g_{\mathbf{x}} : \langle g_{\mathbf{x}}, \mathbf{u}\rangle \leq f^{'}(\mathbf{x}; \mathbf{u}), \forall \mathbf{u}\}$, then $g_{\mathbf{x}} \in \partial f(\mathbf{x})$. Therefore, an equivalent characterization of the subdifferential is

$$\partial f(\mathbf{x}) = \{g_{\mathbf{x}} : \langle g_{\mathbf{x}}, \mathbf{u}\rangle \leq f^{'}(\mathbf{x}; \mathbf{u}), \forall \mathbf{u}\}.$$

(Continue) Proof (Theorem 2). Now, we prove $3 \Rightarrow 4$ using the definition of directional derivative. It suffices to show that for any $g_{\mathbf{x}_*} \in \partial f(\mathbf{x}_*)$, if there exists a $\mathbf{y} \in C$ such that

$$\langle g_{\mathbf{x}_*}, \mathbf{y} - \mathbf{x}_*\rangle < 0,$$

then we have a contradiction, i.e. $\mathbf{x}_* \notin \underset{\mathbf{x}\in C}{\arg\min}\ f(\mathbf{x})$. By

$$f^{'}(\mathbf{x}_*; \mathbf{y} - \mathbf{x}_*) = \sup_{g_{\mathbf{x}_*} \in \partial f(\mathbf{x}_*)} \langle g_{\mathbf{x}_*}, \mathbf{y} - \mathbf{x}_*\rangle,$$

the condition implies that

$$\langle g_{\mathbf{x}_*}, \mathbf{y} - \mathbf{x}_*\rangle \leq f^{'}(\mathbf{x}_*; \mathbf{y} - \mathbf{x}_*).$$

Additionally, since $\partial f(\mathbf{x}_*)$ is closed and bounded (hence compact), we have that $\sup_{g_{\mathbf{x}_*} \in \partial f(\mathbf{x}_*)} \langle g_{\mathbf{x}_*}, \mathbf{y} - \mathbf{x}_*\rangle$ is attained and thus

$$f^{'}(\mathbf{x}_*; \mathbf{y} - \mathbf{x}_*) < 0,$$

which is a contradiction. For the proof of compactness of $\partial f(\mathbf{x}_*)$, see Section 2.4 of [Duchi (2010)].
On the other hand, since $\mathbf{x}_* \in \underset{\mathbf{x}\in C}{\arg\min}\ f(\mathbf{x})$, we have for any $\alpha \geq 0$

$$0 \leq \frac{f(\mathbf{x}_* + \alpha(\mathbf{y} - \mathbf{x}_*)) - f(\mathbf{x}_*)}{\alpha} \tag{9}$$

Taking the limit as $\alpha \to 0$, we have that $f^{'}(\mathbf{x}_*; \mathbf{y} - \mathbf{x}_*) \geq 0, \forall \mathbf{y} \in C$.

□

Example Application - Projections

Definition 5. $\text{Proj}_C(y) := \underset{\mathbf{x}\in C}{\arg\min} \|y - x\|_2$ *is the projection operator onto a constraint set* C.

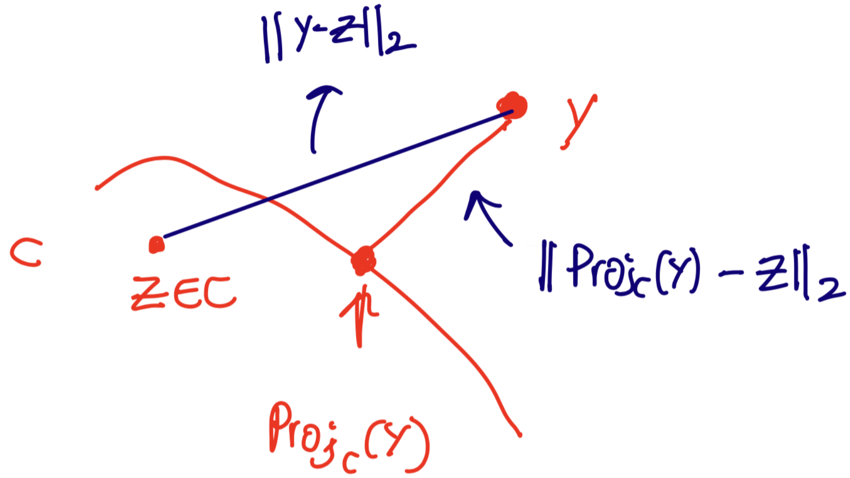

Figure 4: Illustration of projection operator

Theorem 4. *Projection onto any convex set C are non-expansive, i.e.,*

$$\|\text{Proj}_C(y) - z\|_2 \leq \|y - z\|_2,$$

for any z in the convex set C.

Proof. Recall that

$$\text{Proj}_C(y) := \arg\min_{x \in C} \|y - x\|_2 = \arg\min_{x \in C} \|y - x\|_2^2.$$

Denoting

$$\min_{x \in C} f(x), \text{where} f(x) \triangleq \|x - y\|_2^2$$

and using the optimality condition, we get

$$\langle \nabla f(x_*), z - x_* \rangle \geq 0, \forall z \in C \tag{10}$$

$$\Leftrightarrow \langle x_* - y, z - x_* \rangle \geq 0, \forall z \in C \tag{11}$$

Substituting in $\text{Proj}_C(y)$ for x_*, we get:

$$0 \leq \langle \text{Proj}_C(y) - y, z - \text{Proj}_C(y) \rangle, \forall z \in C. \tag{12}$$

$$= \langle \text{Proj}_C(y) - z + z - y, z - \text{Proj}_C(y) \rangle \tag{13}$$

$$= -\|\text{Proj}_C(y) - z\|_2^2 + \langle z - y, z - \text{Proj}_C(y) \rangle. \tag{14}$$

□

Bibliographic notes

More information can be found in Chapter 2 of [Duchi (2010)], Chapter 3 and 4.2 of [Sidford (2024)] and Chapter 1 of [Drusvyatskiy (2020)]

References

[Drusvyatskiy (2020)] Dmitriy Drusvyatskiy. Convex Analysis and Nonsmooth Optimization. 2020.

[Sidford (2024)] Aaron Sidford. Optimization Algorithms. 2024.

[Duchi (2010)] John Duchi. Introductory Lectures on Stochastic Optimization. 2010.