[SD-TSIA 211] Optimization for Machine Learning
“TD-04 Corrections”

Olivier FERCOQ — Iyad WALWIL

January 13, 2023

Recalls

Theorem 1. *(Fermat's rule)*

$$x \in \arg\min f \iff 0 \in \partial f(x)$$

Proposition 1. *Let $f\colon \mathcal{X} \to (-\infty, +\infty]$ be a convex function, $g\colon \mathcal{Y} \to \mathbb{R}$ a convex differentiable function and $M\colon \mathcal{X} \to \mathcal{Y}$ a linear operator*

$$\forall x \in \mathcal{X},\ \partial(f + g \circ M)(x) = \partial f(x) + \{M^* \nabla g(Mx)\}$$

Definition 1. *(Convex cone)*
We say that a subset $\mathcal{C} \subset \mathcal{X}$ is a cone if $\forall x \in \mathcal{C}, \alpha \in \mathbb{R}_+$, the product $\alpha x \in \mathcal{C}$

Exercise 1. *(Optimisation with explicit constraints)*

We consider the following optimization problem

$$\min_{x \in \mathcal{C}} f(x) \tag{1}$$

where $\mathcal{C} \subset \mathbb{R}^d$ is a convex set, and $f\colon \mathbb{R}^d \to \mathbb{R}$ is differentiable and convex.

1. We define the convex indicator function of the set $\mathcal{C}$ as:

$$\imath_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{if } x \notin \mathcal{C} \end{cases}$$

   Show that (1) is equivalent to:

$$\min_{x \in \mathbb{R}^d} f(x) + \imath_{\mathcal{C}}(x) \tag{2}$$

2. Show that for all $x \in \mathcal{C}, \partial \imath_{\mathcal{C}}(x) = \{q \in \mathbb{R}^d : \forall y \in \mathcal{C}, \langle q, y - x \rangle \leq 0\}$ and that $\partial \imath_{\mathcal{C}}(x)$ is a cone (it is called the normal cone to $\mathcal{C}$ at x). Show that for all $x \notin \mathcal{C}, \partial \imath_{\mathcal{C}}(x) = \varnothing$.

3. Show that $x^\star$ is a solution to (2) if, and only if, $-\nabla f(x^\star) \in \partial \imath_{\mathcal{C}}(x^\star)$.

4. Denote $\mathcal{H}_{w,b} = \{x \in \mathbb{R}^d : \langle w, x \rangle + b = 0\}$. Compute $\partial \imath_{\mathcal{H}_{w,b}}(x)$ for all $x \in \mathbb{R}^d$.

5. Prove that the distance of a point z to $\mathcal{H}$ is equal to:

$$d(z, \mathcal{H}_{w,b}) = \min_{x \in \mathcal{H}_{w,b}} \|x - z\|_2 = \frac{|\langle w, z \rangle + b|}{\|w\|_2}$$

Answers

1.

$$\begin{aligned}
(2) = \min_{x \in \mathbb{R}^d} f(x) + \imath_{\mathcal{C}}(x) &= \min_{x \in \mathbb{R}^d} f(x) + \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{if } x \notin \mathcal{C} \end{cases} \\
&= \min_{x \in \mathbb{R}^d} \begin{cases} f(x) & \text{if } x \in \mathcal{C} \\ +\infty & \text{if } x \notin \mathcal{C} \end{cases} \\
&= \min_{x \in \mathcal{C}} f(x) = (1)
\end{aligned}$$

2. We have three thing to be proven here:

- For $x \in \mathcal{C}$ (i.e. $\imath_{\mathcal{C}}(x) = 0$),

$$\begin{aligned}
q \in \partial \imath_{\mathcal{C}}(x) &\iff \forall y \in \mathbb{R}^d,\ \imath_{\mathcal{C}}(y) \geq \imath_{\mathcal{C}}(x) + \langle q, y - x \rangle \\
&\iff \forall y \in \mathbb{R}^d,\ \imath_{\mathcal{C}}(y) \geq \langle q, y - x \rangle \\
&\quad (i) \text{ if } y \notin \mathcal{C},\ +\infty \geq \langle q, y - x \rangle \Rightarrow q \in \mathbb{R}^d \\
&\quad (ii) \text{ if } y \in \mathcal{C},\ 0 \geq \langle q, y - x \rangle
\end{aligned}$$

  Thus, $\partial \imath_{\mathcal{C}}(x) = \{q \in \mathbb{R}^d : \forall y \in \mathcal{C},\ \langle q, y - x \rangle \leq 0\}$ $\square$

- To prove that for $x \in \mathcal{C}, \partial \imath_{\mathcal{C}}(x)$ is a cone, let $\lambda \in \mathbb{R}_+$ and $q \in \partial \imath_{\mathcal{C}}(x)$, then

$$\langle \lambda q, y - x\rangle = \underbrace{\lambda}_{\geq 0} \underbrace{\langle q, y - x\rangle}_{\leq 0} \leq 0 \Rightarrow \lambda q \in \partial \imath_{\mathcal{C}}(x) \qquad \square$$

- For $x \notin \mathcal{C}$ (i.e. $\imath_{\mathcal{C}}(x) = +\infty$),

$$\begin{aligned} q \in \partial \imath_{\mathcal{C}}(x) &\iff \forall y \in \mathbb{R}^d,\ \imath_{\mathcal{C}}(y) \geq \imath_{\mathcal{C}}(x) + \langle q, y - x\rangle \\ &\iff \forall y \in \mathbb{R}^d,\ \imath_{\mathcal{C}}(y) \geq +\infty \\ &\text{if } y \in \mathcal{C},\ 0 \geq +\infty \Rightarrow \partial \imath_{\mathcal{C}}(x) = \varnothing \qquad \square \end{aligned}$$

3. Using Fermat's Rule:

$$\begin{aligned} x^\star = \arg\min_{x \in \mathbb{R}^d} f(x) + \imath_{\mathcal{C}}(x) &\iff 0 \in \partial(f(.) + \imath_{\mathcal{C}}(.))(x^\star) \\ &\overset{(1)}{\iff} 0 \in \nabla f(x^\star) + \partial \imath_{\mathcal{C}}(x^\star) \\ &\iff -\nabla f(x^\star) \in \partial \imath_{\mathcal{C}}(x^\star) \qquad \square \end{aligned}$$

4. $\mathcal{H}_{w,b}$ is a convex set (hyperplane), then we can use the result from question 2.

$$\partial \imath_{\mathcal{H}_{w,b}}(x) = \begin{cases} \{q \in \mathbb{R}^d : \forall y \in \mathcal{H}_{w,b}, \langle q, y - x\rangle \leq 0\} & \text{If } x \in \mathcal{H}_{w,b} \\ \varnothing & \text{Otherwise} \end{cases}$$

Let $x \in \mathcal{H}_{w,b}$ (i.e. $\langle w, x\rangle + b = 0$), then:

$$\begin{aligned} \partial \imath_{\mathcal{H}_{w,b}}(x) &= \{q \in \mathbb{R}^d : \forall y \in \mathcal{H}_{w,b}, \langle q, y - x\rangle \leq 0\} \\ &= \{q \in \mathbb{R}^d : \forall y \in \mathbb{R}^d, \langle w, y\rangle + b = 0, \langle q, y - x\rangle \leq 0\} \end{aligned}$$

- Assume $\exists q' \in \mathbb{R}^d$ such that

$$\forall y \in \mathbb{R}^d, \langle w, y\rangle + b = 0 \ \&\ \langle q', y - x\rangle < 0$$

Then, $y' = 2x - y$ is such that $\langle w, y'\rangle + b = 0$, so:

$$\begin{aligned} \langle q', y' - x\rangle < 0 &\Rightarrow \langle q', x - y\rangle < 0 \\ &\Rightarrow \langle q', y - x\rangle > 0 \quad \text{contradiction} \end{aligned}$$

Thus, for $x \in \mathcal{H}_{w,b}$

$$\begin{aligned}\partial \imath_{\mathcal{H}_{w,b}}(x) &= \{q \in \mathbb{R}^d : \forall y \in \mathbb{R}^d, \langle w, y\rangle + b = 0, \langle q, y - x\rangle = 0\} \\ &= \mathrm{Span}(w) \end{aligned} \tag{3}$$

5. If $z \in \mathcal{H}_{w,b}$, then $x = z$, and $d(z, \mathcal{H}_{w,b}) = 0$. Otherwise, (i.e. $z \notin \mathcal{H}_{w,b}$):

$$\begin{aligned}
x^\star &= \arg\min_{x \in \mathcal{H}_{w,b}} \|x - z\|_2 \\
&\overset{(2)}{=} \arg\min_{x \in \mathbb{R}^d} \|x - z\|_2 + \imath_{\mathcal{H}_{w,b}}(x) \\
&\iff 0 \in \partial(\|. - z\|_2 + \imath_{\mathcal{H}_{w,b}}(.))(x^\star) \\
&\iff 0 \in \frac{x^\star - z}{\|x^\star - z\|_2} + \partial \imath_{\mathcal{H}_{w,b}}(x^\star) \\
&\overset{(3)}{\iff} \frac{z - x^\star}{\|x^\star - z\|_2} \in \mathrm{Span}(w) \\
&\iff \frac{z - x^\star}{\|x^\star - z\|_2} = mw && \text{for some } m \in \mathbb{R} \\
&\iff x^\star = z - m'w && m' = \|x^\star - z\|_2 m
\end{aligned} \tag{4}$$

Moreover, we know that $x^\star \in \mathcal{H}_{w,b}$, thus:

$$\begin{aligned}
\langle w, x^\star\rangle + b = 0 &\iff \langle w, z - m'w\rangle + b = 0 \\
&\iff \langle w, z\rangle + bm'\|w\|^2 \\
&\iff m' = \frac{\langle w, z\rangle + b}{\|w\|^2}
\end{aligned} \tag{5}$$

Hence,

$$\begin{aligned}
d(z, \mathcal{H}_{w,b}) = \min_{x \in \mathcal{H}_{w,b}} \|x - z\|_2 &= \|x^\star - z\|_2 \\
&\overset{(4)}{=} \|z - m'w - z\|_2 = |m'|\|w\| \\
&\overset{(5)}{=} \frac{|\langle w, z\rangle + b|}{\|w\|^2}\|w\| \\
&= \frac{|\langle w, z\rangle + b|}{\|w\|} \qquad \square
\end{aligned}$$

Exercise 2. *(Projected stochastic gradient)*

We consider the following optimization problem:

$$\min_{x\in\mathcal{C}} \sum_{i=1}^{n} f_i(x) \tag{6}$$

where $\mathcal{C} = [0,1]^d$ for all i, $f_i\colon \mathbb{R}^d \to \mathbb{R}$ is differentiable.

1. Show that (6) is equivalent to:

$$\min_{x\in\mathbb{R}^d} \sum_{i=1}^{n} f_i(x) + \imath_{\mathcal{C}}(x)$$

2. Compute the proximal operator of $\imath_{\mathcal{C}}$.

3. Write the proximal stochastic gradient method for the resolution of (6).

Answers:

1. The same as in the previous exercise.

2. By the definition of the proximal operator:

$$\begin{aligned}
\mathrm{Prox}_{\imath_{\mathcal{C}}}(x) &= \arg\min_{y\in\mathbb{R}^d} \imath_{\mathcal{C}}(y) + \frac{1}{2}\|y-x\|^2 \\
&\overset{(2)}{=} \arg\min_{y\in\mathcal{C}} \frac{1}{2}\|y-x\|^2 \\
&= \mathrm{Proj}_{\mathcal{C}}(x)
\end{aligned}$$

3. Let $I \sim \mathcal{U}(\{1,\dots,n\})$, then $\sum_{i=1}^{n} f_i(x) = \mathbb{E}[nf_I(x)]$. Hence,

$$\begin{aligned}
(6) &\equiv \min_{x\in\mathcal{C}} \sum_{i=1}^{n} f_i(x) \\
&\equiv \min_{x\in\mathcal{C}} \mathbb{E}[nf_I(x)] \\
&\equiv \min_{x\in\mathbb{R}^d} \mathbb{E}[nf_I(x)] + \imath_{\mathcal{C}}(x)
\end{aligned}$$

Thus, the proximal stochastic gradient method for the resolution of (6) is:

$$\begin{cases} \text{Generate } I_{k+1} \sim \mathcal{U}(\{1, \dots, n\}) \\ x_{k+1} = \mathrm{Proj}_{\mathcal{C}}(x_k - \gamma_k \nabla n f_{I_{k+1}}(x_k)) \end{cases}$$