Glossary of optimization in machine learning `Ver.April 23, 2023`

Andersen Ang homepage:`angms.science`

What is this A collection of things I have seen 999 times $\oplus$ my set of notations.

1 Symbol and set

- $a \coloneqq b$: a is defined by b
- $a =: b$: b is defined by a
- k : iteration counter, $k \in \mathbb{N}$
- $[n] = \{1, 2, \cdots, n\}$
- $[a, b]$: close interval of a, b, with a, b included
- $]a, b[$: open interval of a, b, with a, b excluded
- $\mathbb{R}, \bar{\mathbb{R}}, \mathbb{R}_+, \mathbb{R}^n$: reals, extended reals (including $\pm\infty$), nonnegative reals, n-dimensional reals
- $\text{supp}(\boldsymbol{x}) = \big\{ i \in [n] \mid x_i \neq 0 \big\}$, for a given vector $\boldsymbol{x} \in \mathbb{R}^n$
- Id : identity operator $\quad \text{Id}\boldsymbol{x} \equiv \boldsymbol{x}$
- If $\boldsymbol{x} \in \mathbb{R}^n$, then
  - $\boldsymbol{x} \geqq \mathbf{0}$ means $\boldsymbol{x} > \mathbf{0}$ or $\boldsymbol{x} = \mathbf{0}$, element-wise
  - $\boldsymbol{x} \geq \mathbf{0}$ means $\boldsymbol{x} \neq \mathbf{0}$ (not all x_i are zero)

Given a function $f : \mathbb{R}^n \to \mathbb{R}$

- $\text{dom} f$: the effective domain of f $\quad \text{dom} f \coloneqq \{x \in X \ : \ f(\boldsymbol{x}) < +\infty\}$ for $f : X \to \mathbb{R}$
- $\text{epi}\, f$: the epigraph of f $\quad \text{epi}\, f \coloneqq \Big\{ (\boldsymbol{x}, \alpha) \in \text{dom} f \times \mathbb{R} \ : \ \alpha \geq f(\boldsymbol{x}) \Big\}$
- ∂f: the subdifferential of f $\quad \partial f \coloneqq \Big\{ \boldsymbol{v} \in \mathbb{R}^n \ : \ f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \langle \boldsymbol{v}, \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle \Big\}$
- $\text{cl} f$: the lower closure of f
- $\min f$: the minimum value of f $\quad \min f$
- $\inf f$: the infimal value of f
- $\text{argmin}\ f$: the set of minimizer of f $\quad \text{argmin}\ f \coloneqq \Big\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ f(\boldsymbol{x}) = \min f \Big\}$
- $\boldsymbol{x}^*$: an optimizer, an means it is possible for $\boldsymbol{x}^*$ to be non-unique $\quad \boldsymbol{x}^* \in \text{argmin}\ f$
- f^* : optimal function value $\quad f^* \coloneqq f(\boldsymbol{x}^*)$

Given a set $\mathcal{S}$,

- $d_{\mathcal{C}}(\boldsymbol{x})$: distance of a point $\boldsymbol{x}$ to $\mathcal{S}$ $\quad d_{\mathcal{C}}(\boldsymbol{x}) = \min\limits_{\boldsymbol{s} \in \mathcal{S}} \|\boldsymbol{x} - \boldsymbol{s}\|$
- $\text{proj}_{\mathcal{S}}(\boldsymbol{x})$: projection of a point $\boldsymbol{x}$ onto $\mathcal{S}$ $\quad d_{\mathcal{C}}(\boldsymbol{x}) = \operatorname*{argmin}\limits_{\boldsymbol{s} \in \mathcal{S}} \|\boldsymbol{x} - \boldsymbol{s}\|$
- $\text{conv}\mathcal{S}$: convex hull of $\mathcal{S}$
- $\text{int}\mathcal{S}$: interior of $\mathcal{S}$
- $\iota_{\mathcal{C}}(\boldsymbol{x})$: indicator function of $\mathcal{S}$ at $\boldsymbol{x}$ $\quad \iota_{\mathcal{C}}(\boldsymbol{x}) = \begin{cases} 0 & \boldsymbol{x} \in \mathcal{S} \\ \infty & \boldsymbol{x} \notin \mathcal{S} \end{cases}$
- $\mathcal{N}_{\mathcal{C}}(\boldsymbol{x})$: normal cone of $\mathcal{S}$ at $\boldsymbol{x}$

2 Function

For a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$,

- $f \in \mathcal{C}^0$: f is continuous
- $f \in \mathcal{C}^1$: f and gradient of f are continuous
- $f \in \mathcal{C}^2$: f, gradient of f and Hessian of f are continuous
- $f \in \mathcal{C}^1_L$: $f \in \mathcal{C}^1$ and gradient of f is L-Lipschitz, or f is L-smooth
- $f \in \mathcal{F}^k_L$: f is $\mathcal{C}^k_L$ and convex, or f is convex and L-smooth
- $f \in \mathcal{S}^k_{M,L}$: f is $\mathcal{F}^k_L$ and M-strongly convex
- f is L-Lipschitz if $|f(x) - f(y)| \leq L\|x - y\|$, $L \geq 0$
- f is nonexpansive if f is 1-Lipschitz
- f^* convex conjugate of f
- f is KL : f fulfills Kurdyka-Lojasiewicz inequality
- f is L-smooth: f is differentiable and $\nabla f(\boldsymbol{x})$ is L-Lipschitz

3 Special functions

- **Distance** The distance of a point $\boldsymbol{x} \in \mathbb{R}^n$ to a closed set $\mathcal{C} \subseteq \mathbb{R}^n$ is $\operatorname{dist}(\boldsymbol{x}, \mathcal{C}) \coloneqq \min\limits_{\boldsymbol{y} \in \mathcal{C}} \|\boldsymbol{y} - \boldsymbol{x}\|_2$
- **Proximal map** The proximal map of a point $\boldsymbol{x} \in \mathbb{R}^n$ under a proper and closed function h with parameter $\eta > 0$ is defined as $\operatorname{prox}_{\eta h}(\boldsymbol{x}) \coloneqq \operatorname*{argmin}\limits_{\boldsymbol{z}} H(\boldsymbol{z}) = h(\boldsymbol{z}) + \dfrac{1}{2\eta}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2$.
  - If h is convex, the map is a singleton because H is strongly convex.
  - If h is nonconvex, the map can be set-valued.
- Indicator function Given a set $\mathcal{C} \subseteq \mathbb{R}^n$, the indicator function of $\mathcal{C}$

$$i_{\mathcal{C}}(\boldsymbol{x}) = \begin{cases} 0 & \boldsymbol{x} \in \mathcal{C} \\ \infty & \boldsymbol{x} \notin \mathcal{C} \end{cases}$$

- $[\ \cdot\ ]_+ = \max\{\,\cdot\,, 0\,\}$.

4 Derivative of function

- Differentiation
- Total differential
- Derivative
- Gradient
- Directional derivative
- Subgradient an element in subdifferential ∂f
- Subdifferential $\partial f \coloneqq \Big\{\boldsymbol{v} \in \mathbb{R}^n \ : \ f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \langle \boldsymbol{v}, \boldsymbol{x} - \bar{\boldsymbol{x}}\rangle\Big\}$

5 Conjugate and Duality

- For a convex $f : \mathbb{R}^n \to \mathbb{R}$, its convex conjugate $f^* : \mathbb{R}^n \to \mathbb{R}$ is

$$f^*(\boldsymbol{x}_0) = \sup_{\boldsymbol{x}} \langle \boldsymbol{x}_0, \boldsymbol{x}\rangle - f(\boldsymbol{x})$$

6 Shorthands

- nnz : number of non-zeros
- FoC : First-order optimality condition / Fermat's rule
- FoM : First-order method (gradient-based method)
- NNLS : nonnegative least sqaures

7 Algorithm names

- GD : gradient descent $\boldsymbol{x}^+ = \boldsymbol{x} - \alpha \nabla f(\boldsymbol{x})$
- PGD / ProjGD : projected gradient descent $\boldsymbol{x}^+ = \operatorname{proj}_{\mathcal{S}}\Big(\boldsymbol{x} - \alpha \nabla f(\boldsymbol{x})\Big)$
- PGD / ProxGD : proximal gradient descent $\boldsymbol{x}^+ = \operatorname{prox}_{\lambda g}\Big(\boldsymbol{x} - \alpha \nabla f(\boldsymbol{x})\Big)$
- SGM / SubGM : subgradient method $\boldsymbol{x}^+ = \boldsymbol{x} - \alpha \boldsymbol{g}$
- SG : Stochastic gradient $\boldsymbol{x}^+ = \boldsymbol{x} - \alpha \nabla_i f_i(\boldsymbol{x})$
- HBM : Heavy ball method
- NAG : Nesterov's accelerated gradient $\boldsymbol{x}^+ = \boldsymbol{y} - \alpha \nabla f(\boldsymbol{y}), \quad \boldsymbol{y}^+ = \boldsymbol{x}^+ + \beta(\boldsymbol{x}^+ - \boldsymbol{x})$
- FW: Frank-Wolfe algorithm
- CPM : cutting plane method
- ADMM : Alternating direction method of multipliers
$$\begin{aligned} \boldsymbol{x}^+ &= \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}\,\boldsymbol{\lambda}) \\ \boldsymbol{y}^+ &= \operatorname*{argmin}_{\boldsymbol{y}} \mathcal{L}_\rho(\boldsymbol{x}^+, \boldsymbol{y}\,\boldsymbol{\lambda}) \\ \lambda^+ &= \lambda + \rho(\boldsymbol{A}\boldsymbol{x}^+ + \boldsymbol{B}\boldsymbol{y}^+ + \boldsymbol{c}) \end{aligned}$$
- IPM : Interior point method
- AS : Active Set
- AA : Anderson's Acceleration

8 Algorithm properties

- Monotone / Relaxation sequence : $f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k)$

9 Common setting

- Given a closed nonempty convex set $\mathcal{C} \subset \mathbb{R}^n$, the *distance* and *projection* of a point $\boldsymbol{x}_0$ onto $\mathcal{C}$ are given by
$$d_{\mathcal{C}}(\boldsymbol{x}_0) = \inf_{\boldsymbol{x}} \|\boldsymbol{x} - \boldsymbol{x}_0\|_2, \quad \operatorname{proj}_{\mathcal{C}}(\boldsymbol{x}_0) = \operatorname*{argmin}_{\boldsymbol{x}} \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 = \arg d_{\mathcal{C}}(\boldsymbol{x}_0).$$
- Given a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, the *domain* and *epigraph* of f are the sets
$$\operatorname{dom} f \coloneqq \Big\{\boldsymbol{x} \in \mathbb{R}^n \ : \ f(\boldsymbol{x}) \leq \infty\Big\}, \quad \operatorname{epi} f \coloneqq \Big\{(\boldsymbol{x}, r) \in \mathbb{R}^n \times \mathbb{R} \ : \ f(\boldsymbol{x}) \leq r\Big\},$$
- A function is closed if its epigraph is a closed set.
- A function is proper if it has nonempty domain and never take the value $-\infty$.
- Set of CCP function (convex, closed = LSC, proper)
$$\Gamma \coloneqq \Big\{ f : \mathbb{R}^n \to]-\infty, +\infty] \mid f \text{ is convex, lower semicontinuous, and proper} \Big\}$$