<u>Lec. 20</u> <u>(12/01/2022)</u>

(The Last!!)

We assume that ψ is differentiable & strictly convex, $\psi : \text{dom}(\psi) \mapsto \mathbb{R}$, such that the constraint set satisfies:

- $\mathcal{X} \subset \text{closure}(\text{dom}(\psi))$,
- $\text{range}(\nabla \psi) = \mathbb{R}^n$,
- and $\|\nabla \psi\|_2 \to +\infty$ as $\underline{x} \to$ boundary of closure(dom(ψ))

Then we say $\psi(\cdot)$ is a <u>mirror map</u>.

Conceptually, what's going on?

Take $\underline{x}_k$ $\xrightarrow{\text{map to}}$ dual space by taking the gradient/subgradient

element of the primal space

do

update in gradient/dual space

come back

primal space to obtain $\underline{x}_{k+1}$

$\boxed{\text{via projection w.r.t. } D_\psi}$ (Bregman projection)

Bregman divergence:

Given a mirror map ψ, the associated Bregman divergence $D_\psi : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}_{\geq 0}$ is defined as:

$$D_\psi(\underline{x}, \underline{y}) := \psi(\underline{x}) - \psi(\underline{y}) - \langle \nabla\psi(\underline{y}), \underline{x} - \underline{y} \rangle$$

$$\forall\ \underline{x}, \underline{y} \in \mathcal{X}.$$

Very clean geometric interpretation: $D_\psi(\underline{x}, \underline{y})$ quantifies the amount by which a strictly convex function lies above its tangent hyperplane

Example:

- $\psi(\underline{x}) = \|\underline{x}\|_2^2$, $\mathcal{X}$ (closed) $\subset \mathbb{R}^n$

$$\begin{aligned} D_\psi(\underline{x}, \underline{y}) &= \|\underline{x}\|_2^2 - \|\underline{y}\|_2^2 - \langle 2\underline{y}, \underline{x} - \underline{y} \rangle \\ &= \|\underline{x}\|_2^2 - \|\underline{y}\|_2^2 + 2\|\underline{y}\|_2^2 - \langle 2\underline{y}, \underline{x} \rangle \\ &= \|\underline{x}\|_2^2 + \|\underline{y}\|_2^2 - 2\langle \underline{y}, \underline{x} \rangle \\ &= \|\underline{x} - \underline{y}\|_2^2 \end{aligned}$$

- $\psi(\underline{x}) = \sum_{i=1}^{n} x_i \log x_i$, $\mathcal{X} \equiv \Delta^{n-1} = \{\underline{x} \in \mathbb{R}^n_{\geq 0} \mid \sum_{i=1}^{n} x_i = 1\}$

$$D_\psi(\underline{x}, \underline{y}) = \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right),$$ Kullback-Leibler divergence

In general, we think about $D_\psi(\underline{x}, \underline{y})$ as some kind of distance squared

But $D_\psi(\underline{x}, \underline{y})$ may not be a metric, in general

e.g., $\underbrace{D_\psi(\underline{x}, \underline{y}) \neq D_\psi(\underline{y}, \underline{x})}_{\text{maybe non-symmetric}}$

e.g., $D_\psi(\underline{x}, \underline{y})$ may not satisfy the triangle inequality.

- When $\psi(\cdot) = \|\cdot\|_2^2$, Then $D_\psi(\underline{x}, \underline{y}) = \|\underline{x} - \underline{y}\|_2^2$

and mirror descent $\xrightarrow[\text{to}]{\text{reduces}}$ projected grad. descent

- When $\psi(\cdot) = \sum_{i=1}^{n} x_i \log x_i$ (neg. entropy)

$$\mathcal{X} = \Delta^{n-1}$$

$$\underbrace{D_\psi(\underline{x}, \underline{y}) = \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right)}_{\text{Kullback-Leibler divergence.}}$$

Then the mirror descent algorithm becomes:

$$\begin{cases} \underline{y}_{k+1} = \underline{x}_k \odot \exp\left(-\eta_k \nabla f(x_k)\right) \\ \left.\underline{x}_{k+1} = \dfrac{\underline{y}_{k+1}}{\mathbb{1}^T \underline{y}_{k+1}}\right\} \text{Kullback-Leibler projection} \end{cases}$$

(see Lec. 16, p. 10-15)

where $\odot$ denotes elementwise vector-vector multiplication