Geometrization I: Frank-Wolfe.

Wednesday, February 5, 2020 6:18 PM

Recall MD.

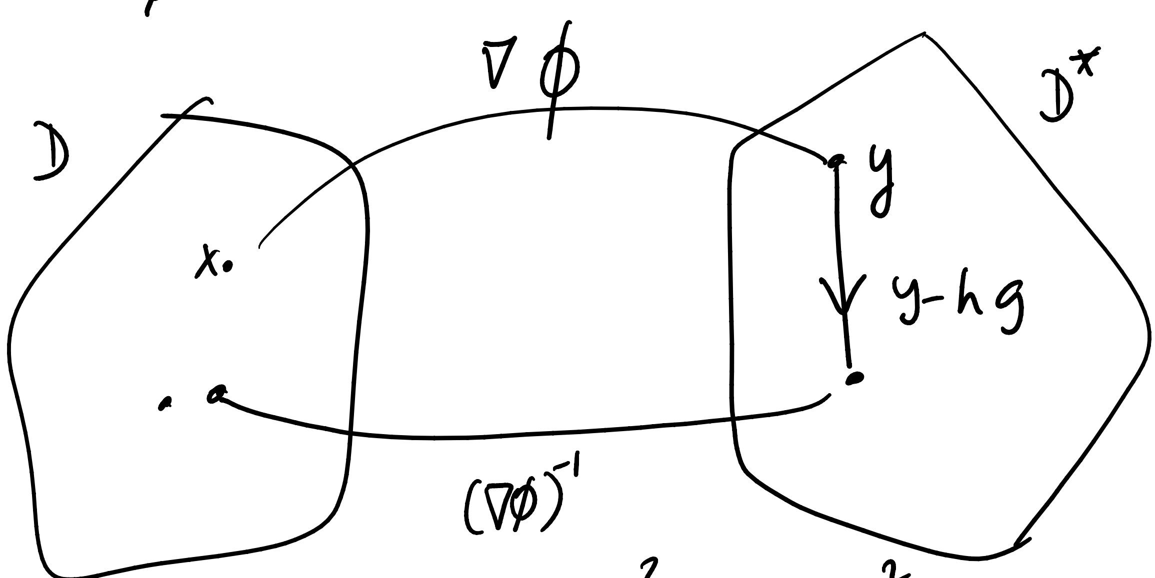
f. Lipshitz. convex

ϕ : ρ-strongly convex

$$R^2 = \sup_x \phi(x) - \phi(x^0).$$

$$D_\phi(x^*, x^0) = \phi(x^*) - \phi(x^0) \underbrace{- \langle \nabla\phi(x^0), \dots \rangle}_{=0}$$

$$D_\phi(y,x) = \phi(y) - \phi(x) - \langle \nabla\phi(x), y-x \rangle$$



$$f\left(\frac{1}{T}\sum_{k=0}^{T-1} x^{(k)}\right) - f(x^*) \le \frac{R^2}{hT} + \frac{hG^2}{2\rho} \le \frac{RG}{\sqrt{2\rho T}}.$$

<u>Algo:</u> start at $x^{(0)} = \arg\min_x \phi(x)$

$$y^{k+1}: \quad \nabla\phi(y^{k+1}) = \nabla\phi(x^k) - hg^k.$$

$$x^{k+1} = \arg\min_{x \in D} D_\phi(x, y^{k+1})$$

Note: $$x^{k+1} = \underset{x \in D}{\operatorname{argmin}} \ \phi(x) - \phi(y^{k+1}) - \langle \nabla\phi(y^{k+1}), y^{k+1} - x \rangle$$

$$= \underset{x \in D}{\operatorname{argmin}} \ \phi(x) - \langle \nabla\phi(y^{k+1}), x^k \rangle$$

$$= \underset{x \in D}{\operatorname{argmin}} \ \phi(x) - \langle \nabla\phi(x^k) - hg^k, x \rangle$$

$$= \underset{x \in D}{\operatorname{argmin}} \ \phi(x) - \phi(x^k) - \langle \nabla\phi(x^k), x - x^k \rangle + h \langle g^k, x \rangle$$

$$= \operatorname{argmin} \ h \langle g^k, x \rangle + D_\phi(x, x^k)$$

i.e. generalization of

$$x^{k+1} = \operatorname{argmin} \ h \langle g^k, x \rangle + \|x - x^k\|_2^2$$

Recall

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \ \|x - (x^k - hg^k)\|_2^2$$

$$= \underset{x}{\operatorname{argmin}} \ \|x - x^k\|^2 + 2h \langle g^k, x - x^k \rangle + h^2 \|g^k\|^2$$

$$= \underset{x}{\operatorname{argmin}} \ \|x - x^k\|^2 + h \langle g^k, x \rangle$$

<u>Example</u>. $D = \{ x : \|x\|_1 = 1, \ x \geq 0 \}$

$\phi = \sum x_i \log x_i \qquad x^{(0)} = \frac{1}{n} \mathbb{1}$.

$$\phi = \sum x_i \log x_i \qquad x^{(0)} = \frac{1}{n} \mathbb{1}.$$

$$D_\phi(y,x) = \sum y_i \log \frac{y_i}{x_i}$$

$$x^{k+1} = \underset{\substack{\sum x_i = 1 \\ x \geq 0}}{\operatorname{argmin}} \; h \langle g^k, x \rangle + \sum_i x_i \log \frac{x_i}{x_i^{(k)}}$$

i.e. $\quad h g_i^k + \left(1 + \log \frac{x_i}{x_i^k}\right) = \lambda$

$$x_i = x_i^k \cdot e^{-h g_i^k} \cdot e^{\lambda}$$

$$\text{s.t. } \sum x_i = 1.$$

Each x_i is updated by a multiplicative factor $e^{-h \cdot g_i^k}$ and normalized!

MD is not good when R^2 is large.

e.g. $\{x : \|x\|_\infty \leq 1\}$

Fact: $\forall$ 1-strongly convex ϕ,

$$\sup \phi(x) - \phi(0) = \frac{n}{2}.$$

Frank-Wolfe.

$$\begin{cases} y^{k+1} = \underset{y \in D}{\operatorname{argmin}} \quad \langle \nabla f(x^k), y \rangle \\ x^{k+1} = (1-h_k) x^k + h_k y^{k+1} \end{cases}$$

$$h_k = \frac{2}{k+2}.$$

Thm. $f(x^k) - f(x^*) \leq \frac{2 C_f}{k+2}$

where C_f:

$\forall x, y$:

$$f((1-h)x + hy) \leq f(x) + h \langle \nabla f(x), y-x \rangle + \frac{1}{2} C_f h^2.$$

∇f is "C_f - Lipschitz".

Usual condition:

$$f(x + h(y-x)) \leq f(x) + \langle \nabla f(x), h(y-x) \rangle + \frac{h^2}{2} (y-x)^T \nabla^2 f (y-x)$$

$$\leq \quad + \frac{1}{2} h^2 L \cdot \|y-x\|^2.$$

$$\leq \quad \text{———} \quad + \frac{1}{2} h^2 L \cdot \|y - x\|^2.$$

$$\boxed{C_f \leq L \cdot D^2}$$ in norm of interest.

Pf. $f(x^{k+1}) = f((1-h_k)x^k + h_k y^{k+1})$

$$\leq f(x^k) + h_k \langle \nabla f(x^k), y^{k+1} - x^k \rangle + \frac{1}{2} C_f h_k^2$$

On the other hand

$$f(x^*) \geq f(x^k) + \langle \nabla f(x^k), x^* - x^k \rangle$$

$$\geq f(x^k) + \langle \nabla f(x^k), y^{k+1} - x^k \rangle$$

since y^{k+1} minimizes $\langle \nabla f(x^k), y \rangle$

$$f(x^{k+1}) - f(x^*) \leq (1-h_k)(f(x^k) - f(x^*)) + \frac{1}{2} C_f h_k^2$$

$$e_{k+1} \leq (1-h_k) e_k + \frac{1}{2} C_f h_k^2.$$

$$f(x^{(2)}) \leq f(x^*) + 0 + \frac{1}{2} C_f \implies e_2 \leq \frac{C_f}{2}$$

$$e_{k+1} \leq \frac{2 C_f}{k+2}.$$ by induction