Lecture 10 - scribbles - Frank-Wolfe

$$\frac{1}{2}$$

$$\frac{1}$$

Teaching Page 1

$$\int \int du - succh's de august $\int (x_{c} + \delta_{c}(s_{c} - x_{c}))$

$$\int \int du - succh's de august $\int f(x_{c} + \delta_{c}(s_{c} - x_{c}))$

$$\int \int f(x_{c}) - hot f(x_{c}) \leq O(\frac{1}{2})$$

$$if = \int f(x_{c}) - f(x_{c}) + \int f(x_{c}) - \int f(x_{c}) + \int f(x_{c}$$$$$$

$$\begin{aligned} & \mathcal{L}_{\mathrm{G}} \mathbb{N}^{-1} \qquad \mathcal{L}_{\mathrm{G}} \mathbb{N}^{-1} \qquad \mathcal{L}_{\mathrm{G}} \mathbb{N}_{\mathrm{G}} \mathbb{N}_{\mathrm{G}} \\ & (\operatorname{larine}^{-1} \operatorname{Loppose}^{-1} \operatorname{Inn} \mathbb{F} \mathbb{W} \text{ and } \mathbb{S}^{+} \mathbb{K}^{-1} \text{ for a gent } \mathbb{S}^{+} \\ & (\operatorname{larine}^{-1} \operatorname{Loppose}^{-1} \operatorname{Inn} \mathbb{F} \mathbb{W} \text{ and } \mathbb{S}^{+} \mathbb{K}^{-1} \text{ (power life brocks}^{-1}) \\ & (\operatorname{larine}^{-1} \operatorname{Loppose}^{-1} \operatorname{Inn} \mathbb{F} \mathbb{W} \text{ and } \mathbb{S}^{+} \mathbb{K}^{-1} \\ & (\operatorname{larine}^{-1} \operatorname{Loppose}^{-1} \operatorname{Inn} \mathbb{F} \mathbb{W} \text{ and } \mathbb{S}^{+} \mathbb{K}^{-1} \\ & (\operatorname{larine}^{-1} \operatorname{Loppose}^{-1} \operatorname{Lop}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \operatorname{Lop}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \operatorname{Lop}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \operatorname{Lop}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \operatorname{Lop}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \operatorname{Lop}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \operatorname{Lop}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \mathbb{K}^{-} \\ & (\operatorname{larine}^{-1} \mathbb{K}^{-} \\ \\ \\ & (\operatorname{larine}^{-1} \mathbb{K$$

Sup LIISZIIT
$$(G \in L_{HM} \cdot diam_{H/I}(M)^{2})$$

by definition, affine unionizative version of doscent demonse
 $f(x_{X}) \leq f(x) + \delta(trg(x), s-x) + \chi^{2} \leq \forall X \in [0, 1]$
Fill is
multia this as machine possible
 $= -9t \; ft$ Euclidition
for FW step: $f(x_{0}) \leq f(x) - \chi \; g_{t} + \chi^{2} \leq optimal \; step \; size$
with adaptive step size $\chi_{t} = g_{t}$
 $f(x_{X}) \leq f(x_{0}) \leq f(x) - \chi \; g_{t} + \chi^{2} \leq optimal \; step \; size$
 $(i) th adaptive step \; size \; \chi_{t} = g_{t}$
 $f(x_{X}) \leq f(x_{0}) \leq f(x) - \chi \; g_{t} + \chi^{2} \leq optimal \; step \; size$
 $f(x_{X}) \leq f(x) - \chi \; g_{t} + \chi^{2} \leq optimal \; step \; size$
 $f(x_{X}) \leq f(x) - \chi \; g_{t} + \chi^{2} \leq optimal \; step \; size$
 $f(x_{X}) \leq f(x) - \chi \; g_{t} + \chi^{2} \leq optimal \; step \; size$
 $f(x_{X}) \leq f(x) - \chi \; g_{t} + \chi \; g_{t} = g_{t}$

- Good modern overview of Frank-Wolfe algorithm with applications in machine learning:
 - Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization, M. Jaggi, ICML 2013 <u>http://m8j.net/math/revisited-FW.pdf</u>
 - ICML 2014 tutorial: https://sites.google.com/site/frankwolfegreedytutorial/
- Modern survey of away-step Frank-Wolfe and other variants:

 (I might cover these later in the course in more details):
 On the Global Linear Convergence of Frank-Wolfe Optimization Variants, S.
 Lacoste-Julien and M. Jaggi, NIPS 2015
 http://arxiv.org/abs/1511.05932 | NIPS link
 - a more efficient version of the away-step recently revisited: Linear-Memory and Decomposition-Invariant Linearly Convergent Conditional Gradient Algorithm for Structured Polytopes, D. Garber & O. Meshi, NIPS 2016 <u>NIPS link</u>