One page note from EE263C-Optimization Methods for Large-Scale Systems Prof. L. Vandenberghe, UCLA

(Update 25-FEB-2020, Konstantin Burlachenko, burlachenkok@gmail.com)

Subgradient optimization methods

• Handles general nondifferentiable convex problem, Often leads to very simple algorithms

• Convergence can be very slow, no good stopping criterion

• Theoretical complexity: $O\left(\frac{1}{e^2}\right)$ iterations to find eps-suboptimal point

• An "optimal" first-order method: $O\left(\frac{1}{e^2}\right)$ bound cannot be improved

• So rate of optimal value decreasing is $O\left(\frac{1}{\sqrt{k}}\right)$

• $O\left(\frac{G|x_0 - x_*|_2}{\sqrt{k+1}}\right)$ bound cannot be improved for subgradient classical model

Gradient descend method

• Due to Y.Nesterov for any first-order method and any iterative algorithm that selects xk+1 in the set x0 with span covered by gradients calculated so far.

• There is a bad function for which convergence will be bound by $O\left(\frac{1}{k^2}\right)$

• This idea suggests 1/k rate for gradient method is not optimal in fact

• More recent accelerated gradient methods have $O\left(\frac{1}{k^2}\right)$ convergence

Proximal gradient method

• Minimizes sums of differentiable g(x) and maybe non-differentiable convex functions h(x)

• Useful when non-differentiable term h(x) is simple (has inexpensive prox-operator)

• $prox_h(x) = argmin_u\left(h(u) + |u - x|_2^2\right)$

• Convergence properties are similar to standard gradient method

• Less general but faster than sub gradient method, Each proximal gradient iteration is a descent step

• Sub-optimality after k iterations is O(1/k)

• Gradient Descend and Projected sub-gradient method can be considered as custom case of Proximal Gradient method.

Proximal Mapping

• The proximal mapping of a closed convex function f is $prox_f(x) = argmin_u\left(f(u) + |u - x|_2^2\right)$

• The proximal mapping defines a function because for each **x** there is a unique **u.**

- Subgradient characterization $u = prox_f(x) <=> x - u \in \partial f(u)$

Nesterov momentum rate of convergence $f\left(x_k\right) - f^* \leq \frac{2L}{(k+1)^2} \left| x_0 - x^* \right|_2^2$