10725/36725 Optimization
Recitation 2, Sep 11

Aaditya Ramdas
aramdas@cs.cmu.edu

1 Agenda

(Strong) Convexity, Lipschitzness and differentiability of functions/gradients Convex vs strongly convex, lipschitz function vs lipschitz gradient, first and second order definitions of strong convexity and lipschitz gradients in appropriate norms, etc.

Geometric intuition for operations preserving convexity of sets/functions Via the epigraph, max, sums, integrals, intersections, etc. Log-convex, quasi-convex, etc.

2 Random Properties and Thoughts

Max of convex functions is convex (also strictly convex) Its epigraph is the intersection of the epigraphs of the individual functions, and since that intersection is convex, it must mean that the function is convex. Imp example is piecewise linear functions $\max_i \{a_i^\top x + b_i\}$. Hence, for any C, $f(x) = \sup_{y \in C} \|x - y\|$ is convex. $f(X) = \lambda_{max}(X) = \sup\{y^\top X y \,|\, \|y\| = 1\}$ is convex. $f(X) = \sigma_{max}(X) = \|X\|_2 = \sup\{u^\top X v \,|\, \|u\| = \|v\| = 1\}$ is convex. If C is convex, $\min_{y \in C} \|x - y\|$ is convex.

Non-neg weighted sums of convex functions (and integrals!) is convex If $w(y) \geq 0$ for all $y \in A$, and $f(x, y)$ is convex in x for all $y \in A$, then $\int_A w(y) f(x, y) dy$ is convex.

Convex functions have to be continuous except for endpoints. Convex functions are Lipschitz continuous on any closed subinterval.

Strictly convex functions can have a countable number of non-differentiable points. Eg: $f(x) = e^x$ if $x < 0$ and $f(x) = 2e^x - 1$ if $x \geq 0$. So $\max\{e^x, e^{-x}\}$ is strictly convex and not differentiable at 0.

f is Lipschitz in $\|.\|$ iff its subgradient is bounded in the dual norm $\|.\|_*$. Assuming f is Lipschitz, we have by convexity $\partial f_x^\top (y - x) \leq f(y) - f(x) \leq L\|y - x\|$, and since this is true for all y, $\max_y \partial_x^\top \left(\frac{y-x}{\|y-x\|}\right) \leq L$ or in other words $\|\partial_x\|_* \leq L$. Assuming $\|\partial_x\|_* \leq L$, we have by convexity $f(x) - f(y) \leq \partial_x^\top (x - y) \leq L\|x - y\|$ by Holder's inequality.

The sum of a (possibly non-differentiable) convex function with a (possibly non-differentiable) strongly convex function is strongly convex. Let $h = f + g$ where f is strongly convex and g is convex, then h is strongly convex. $h(tx + (1-t)y) = f(tx + (1-t)y) + g(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\lambda}{2}t(1-t)\|x - y\|^2 + tg(x) + (1-t)g(y) \leq th(x) + (1-t)h(y) - \frac{\lambda}{2}t(1-t)\|x - y\|^2$. Hence $x^2 + |x|$ is strongly convex.

Implications of strongly convex functions $f(y) \geq f(x) + \partial_x^\top (y - x) + \frac{\lambda}{2}\|y - x\|^2$ and also $f(y) \leq f(x) + \partial_x^\top (y - x) + \frac{1}{2\lambda}\|\partial_y - \partial_x\|^2$ (how?). Just as for convexity (by adding two equations of definition of convexity) $(\partial_y - \partial_x)^\top (y - x) \geq 0$, for strong convexity we have $(\partial_y - \partial_x)^\top (y - x) \geq \lambda\|y - x\|^2$ and also $(\partial_y - \partial_x)^\top (y - x) \leq \frac{1}{\lambda}\|\partial_y - \partial_x\|^2$ (by adding two equations of the second kind).

A differentiable function f is said to have a lipschitz continuous gradient in norm $\|.\|$ if $\|\nabla_y - \nabla_x\| \leq L\|y - x\|$ Also, if it is twice differentiable, $LI \succeq \nabla_x^2$ (while strong convexity says $\nabla_x^2 \succeq \lambda I$).

Implications of Lipschitz continuous functions $f(y) \leq f(x) + \nabla_x^\top (y - x) + \frac{L}{2}\|y - x\|^2$ and also $f(y) \geq f(x) + \nabla_x^\top (y - x) + \frac{1}{2L}\|\nabla_y - \nabla_x\|^2$ (how?). Similarly, adding two such equations gives $(\nabla_y - \nabla_x)^\top (y - x) \leq L\|y - x\|^2$ and also $(\nabla_y - \nabla_x)^\top (y - x) \geq \frac{1}{L}\|\nabla_y - \nabla_x\|^2$.

$\|Ax - b\|^2$ is not necessarily strongly convex, but $\|Ax - b\|^2 + \lambda\|x\|^2$ always is The hessian $A^\top A$ is $n \times n$ where $x \in \mathbb{R}^n$, but could be rank $m < n$ if the system $Ax = b$ is under-constrained with only $m < n$ constraints. However, $A^\top A + \lambda I$ is always full rank, has lowest eigenvalue of $\geq \lambda$ and hence is $\geq \lambda$-strongly convex. Similarly for non-squared-loss and non-differentiable loss functions.

Strong Convexity depends on the norm $f(x) = \sum_i x_i \log x_i$ defined on the probability simplex $\Delta_n \to \mathbb{R}$ is 1-strongly convex with respect to the 1-norm.