CPSC 665 : An Algorithmist's Toolkit Lecture 1 : 12 Jan 2015

Lecture title

Lecturer: Sushant Sachdeva *Scribe: Anup Rao*

1. Introduction

In this lecture we talked about convexity, and some inequalities that follows as a result of it. We started off with an elegant proof of Caucy-Schwarz inequality. We then defined convexity for general functions, and then studied various equivalent characterizations for special classes (continuos functions, differentiable functions etc...) of functions and proved a fundamental inequality satisfied by convex functions called Jensen's inequality. Finally, we used Jensen's inequality to prove Holder's inequality.

2. Cauchy-Schwarz Inequality

Theorem 2.1 (Caucy-Schwarz Inequality). *Suppose* $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$*, then*

$$\boldsymbol{a}^T \boldsymbol{b} \leq \|\boldsymbol{a}\| \|\boldsymbol{b}\| .$$

Proof. We first observe that it is sufficient to prove the above inequality when the vectors have unit norm $\|a\| = 1 = \|b\|$. This is because if $\boldsymbol{a}, \boldsymbol{b}$ don't satisfy the unit norm condition, then we can apply the above inequality to the vectors $\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}, \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}$.

We therefore assume without loss of generality that $\|\boldsymbol{a}\| = 1 = \|\boldsymbol{b}\|$. Now we note the following simple inequality

$$xy \leq \frac{x^2 + y^2}{2}, \forall x, y \in \mathbb{R}.$$

We start from the left hand side

$$\begin{aligned}
\boldsymbol{a}^T \boldsymbol{b} &= \sum_i a_i b_i \\
&\leq \sum_i \frac{a_i^2 + b_i^2}{2} \\
&= 1/2 \sum_i a_i^2 + 1/2 \sum_i b_i^2 \\
&= 1.
\end{aligned}$$

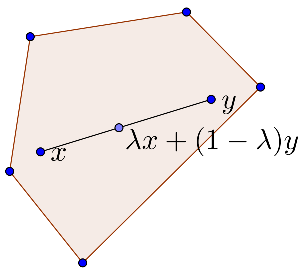
The last inequality follows from the unit norm assumption.

3. Convexity and Jensen's Inequality

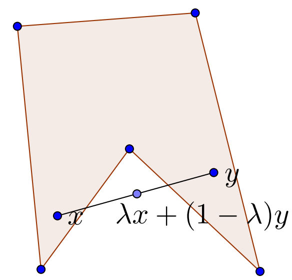
Definition 3.1 (Convexity of a set). *A set* $K \subset \mathbb{R}^n$ *is said to be a convex set if* $\forall \boldsymbol{x}, \boldsymbol{y} \in K, \lambda \in [0, 1]$

$$\lambda \boldsymbol{x} + (1 - \lambda) \boldsymbol{y} \in K.$$

In words, a set K is convex if for any two points $\boldsymbol{x}, \boldsymbol{y} \in K$, the line segment joining $\boldsymbol{x}$ and $\boldsymbol{y}$ is also contained in K.



Convex set



Not a convex set

Now we define what it means to say that a function is convex.

Definition 3.2 (Convexity of a function). *A function $f : K \to \mathbb{R}$ is said to be convex function if $dom(f) = K$ is a convex set, and $\forall \boldsymbol{x}, \boldsymbol{y} \in K, \lambda \in [0, 1]$*

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}).$$

When $f : \mathbb{R} \to \mathbb{R}$, this means that the line segment joining $f(x)$ and $f(y)$ is above the function f. Therefore, in this case, we have for all $x, y \in \mathbb{R}$ and $z \in [x, y]$

$$f(z) \leq f(x) + \frac{f(y) - f(x)}{y - x}(z - x).$$

We now give some equivalent characterizations of convexity which are easier to verify.

Continuous functions: If $f(x)$ is a continuous function, then $f(\boldsymbol{x})$ is convex iff for all $\boldsymbol{x}, \boldsymbol{y} \in K$

$$f(\frac{\boldsymbol{x}+\boldsymbol{y}}{2}) \leq \frac{f(\boldsymbol{x}) + f(\boldsymbol{y})}{2}.$$

That is, if a function is continuous, then to check if it is convex, we can take $\lambda = 1/2$ in the definition of convex functions.

Differentiable functions: If $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then f is convex iff for all $x, y \in \mathbb{R}$,

$$f(x) \geq f(y) + f'(y)(x - y).$$

Proof. In one direction, we have from the convexity of f

$$f(x) - f(y) \geq \frac{f(\lambda x + (1-\lambda)y) - f(y)}{\lambda}.$$

Now letting $\lambda \to 0$ we get that $f(x) - f(y) \geq f'(x)(x - y)$. For the other direction, let $z = \lambda x + (1-\lambda)y$. We then have

$$f(x) \geq f(z) + f'(z)(x - z)$$
$$f(y) \geq f(z) + f'(z)(y - z).$$

Multiplying the first inequality by λ, the second one by $1 - \lambda$ and adding the two gives

$$\lambda f(x) + (1-\lambda) f(y) \geq f(z) = f(\lambda x + (1-\lambda)y).$$

$\square$

Twice Differentiable functions: If a function f is twice differentiable, then f is convex iff for all x in the domain of f

$$f''(x) \geq 0.$$

This immediately implies that $f(x) = e^x$ and $f(x) = -\log x$ are convex functions.

Remark 3.3. *Even when a convex function is not differentiable (e.g.. $f(x) = |x|$), there is a relaxation of the notion of the derivative(s) for a convex function. At any point x in the domain, we can assign a non empty convex set of subderivatives, denoted by $\partial f(x)$, such that any $c \in \partial f(x)$ satisfies*

$$f(x) \geq f(y) + c(x - y).$$

We now state some inequalities for $f(x) = e^x$ that we can now prove that are often very useful. Since e^x is a convex function which is differentiable, we have that for all x, y,

$$e^y \geq e^x + e^x(y - x).$$

Taking $x = 0$, we immediately get that for all $y \in \mathbb{R}$

$$1 + y \leq e^y. \tag{1}$$

Next, using the definition of convexity, we have for all $x, y \in \mathbb{R}$ and $z \in [x, y]$

$$e^z \leq e^x + \frac{e^y - e^x}{y - x}(z - x).$$

Letting $x = 0, y = 1$, we have for all $z \in [0, 1]$

$$e^z \leq 1 + (e - 1)z. \tag{2}$$

The above results readily generalizes to multidimensional case. Before we state the results, let us first introduce some notations. For a function $f : \mathbb{R}^n \to \mathbb{R}$, we define the gradient of f at $\boldsymbol{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ to be

$$\nabla f(\boldsymbol{x}) := (\frac{\partial f}{\partial x_1}, .., \frac{\partial f}{\partial x_n}).$$

This is the generalization of the notion of derivative for functions defined on $\mathbb{R}$. The multivariate notion of second derivative is called the Hessian of f. Hessian of f at $\boldsymbol{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ is the the $n \times n$ matrix

$$\nabla^2 f(\boldsymbol{x}) := \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{ij}.$$

We are now ready to state the results in the multidimensional case.

Differentiable functions: If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then f is convex iff for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \nabla f(y)^T(\boldsymbol{x} - \boldsymbol{y}).$$

Note that as in the remark 3.3 for the single variable case, even if a convex function is not differentiable, at any point $\boldsymbol{y}$ in the domain, we assign a non-empty convex subset $\partial f(\boldsymbol{y}) \subset \mathbb{R}^n$) of subgradients. Furthermore, for any $\boldsymbol{v} \in \partial f(\boldsymbol{y})$ the following holds

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \boldsymbol{v}^T(\boldsymbol{x} - \boldsymbol{y}).$$

Twice Differentiable functions: If a function f is twice differentiable, then f is convex iff for all $\boldsymbol{x}$ in the domain of f and all $\boldsymbol{\zeta} \in \mathbb{R}^n$

$$\boldsymbol{\zeta}^T(\nabla^2 f(\boldsymbol{x}))\boldsymbol{\zeta} \geq 0.$$

Jensen's inequality. Now we can state and prove a fundamental inequality satisfied by convex functions, from which most other inequalities can be derived.

Theorem 3.4 (Jensen's Inequality)**.** *Suppose f is a convex function $f : \mathbb{R}^n \to \mathbb{R}$, and $X \in \mathbb{R}^n$ is a random variable. Then*

$$f(\mathbf{E}\, X) \leq \mathbf{E}\, f(X).$$

Proof. From the convexity of f, we have for all $\boldsymbol{y}$

$$f(X) \geq f(\mathbf{E}\, X) + \nabla f(\mathbf{E}\, X)^T (\boldsymbol{y} - \mathbf{E}\, X).$$

Since the above inequality holds deterministically for all $\boldsymbol{y}$, we can replace $\boldsymbol{y}$ with X to get

$$f(X) \geq f(\mathbf{E}\, X) + \nabla f(\mathbf{E}\, X)^T (X - \mathbf{E}\, X).$$

The above inequality should also hold in expectation. This gives

$$\begin{aligned}\mathbf{E}\, f(X) &\geq \mathbf{E}(f(\mathbf{E}\, X) + \nabla f(\mathbf{E}\, X)^T (\boldsymbol{y} - \mathbf{E}\, X)) \\ &= f(\mathbf{E}\, X).\end{aligned}$$

$\square$

Remark 3.5. *This proof goes through even for non differentiable functions by using subdgradients. We can replace $\nabla f(\mathbf{E}\, X)$ in the above proof by any subgradient $\boldsymbol{v} \in \partial f(\mathbf{E}\, X)$.*

4. Holder's Inequality

Now we use Jensen's inequality and ideas from Cauchy-Schwarz inequality to prove a generalization of the latter. For $\boldsymbol{x} \in \mathbb{R}^n$, we let $\|\boldsymbol{x}\|_p := (\sum_i |x_i|^p)^{1/p}$. We extend this to $p = \infty$ by letting $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$. Holder's inequality states the following

Holder's Inequality: Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$. Further $\infty > p \geq 1$ be any number and $q \geq 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\boldsymbol{a}^T \boldsymbol{b} \leq \|\boldsymbol{a}\|_p \|\boldsymbol{b}\|_q .$$

To prove this, we mimic the proof in the first section. One key additional inequality we will need is stated in the following lemma.

Lemma 4.1 (Young's Lemma)**.** *For scalars $x, y \geq 0$ and $p, q \in (0, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$,*

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Proof. Let us rewrite the above inequality in a way such that we can apply Jensen's inequality. Taking the logarithm on both sides, we see that the inequality is equivalent to

$$\log x + \log y \leq \log \left(\frac{x^p}{p} + \frac{y^q}{q} \right). \tag{3}$$

With the goal of writing left hand side as an expectation of a random variable, and since $\frac{1}{p} + \frac{1}{q} = 1$, we rewrite (3) as

$$\frac{1}{p} \log x^p + \frac{1}{q} \log y^q \leq \log \left(\frac{x^p}{p} + \frac{y^q}{q} \right).$$

Finally, multiplying throughout by -1, we have

$$-\frac{1}{p} \log x^p - \frac{1}{q} \log y^q \geq -\log \left(\frac{x^p}{p} + \frac{y^q}{q} \right).$$

This immediately follows by apply Jensen's inequality to the convex function $f(x) = -\log x$ and the random variable $X = \begin{cases} a^p & \text{w.p.} \quad \frac{1}{p} \\ b^q & \text{w.p.} \quad \frac{1}{q} \end{cases}$.

We are finally ready to prove Holder's inequality.

Theorem 4.2 (Holder's inequality)**.** *Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$. Further $\infty > p \geq 1$ be any number and $q \geq 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\boldsymbol{a}^T \boldsymbol{b} \leq \|\boldsymbol{a}\|_p \|\boldsymbol{b}\|_q .$$

Proof. As in the proof Cauchy-Schwarz inequality, we note that without loss of generality, we can assume $\|\boldsymbol{a}\|_p = 1 = \|\boldsymbol{b}\|_q$. Now,

$$\begin{aligned} \boldsymbol{a}^T \boldsymbol{b} &= \sum_i a_i b_i \\ &\leq \sum_i \frac{a_i^p}{p} + \frac{b_i^q}{q} \text{ (by Young's Lemma)} \\ &= 1/p \sum_i a_i^p + 1/q \sum_i b_i^q \\ &= 1/p + 1/q \text{ (since } \|\boldsymbol{a}\|_p = 1 = \|\boldsymbol{b}\|_q) \\ &= 1. \end{aligned}$$

□

HW 1 (Triange Inequality): Prove triangle inequality for p norms. That is, prove that for any $p \geq 1$ and $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$,

$$\|\boldsymbol{a} + \boldsymbol{b}\|_p \leq \|\boldsymbol{a}\|_p + \|\boldsymbol{b}\|_p .$$

HW 2 (Generalized Means Inequality): For all $a, b \geq 0$, prove that

$$M_p = \left(\frac{a^p + b^p}{2} \right)^{1/p}$$

is an increasing function of p. Here p is allowed to take all real values.