Lecture 3

W. Hebisch

October 26, 2021

1 Convex sets

Recall that points of form

$$tx + (1-t)y$$

where $t \in [0,1]$ form a (closed) line segment joining x and y.

Set S is convex if and only if for each two points $x, y \in S$ line segment joining x and y is contained in S.

Lemma 1.1 *Let Δ be an arbitrary family of convex subsets of $\mathbb{R}^n$. Then intersection*

$$K = \bigcap_{S \in \Delta} S$$

is a convex set.

Now, let F be an arbitrary subset of $\mathbb{R}^n$ and let Δ be family of all convex subsets of $\mathbb{R}^n$ containing F. Then $K = \bigcap_{S \in \Delta} S$ is a convex set containing F and clearly K is smallest such set. K is called *convex hull* of F and denoted by $\mathrm{conv}(K)$.

We can give more constructive description of *convex hull*. When $x_1, \dots, x_m \in \mathbb{R}^n$, $t_1, \dots, t_m \in [0,1]$ and $\sum_{i=1}^m t_i = 1$, then

$$\sum_{i=1}^m t_i x_i$$

is called convex combination of x_i.

Lemma 1.2 *K is convex if and only if each convex combination of elements of K belongs to K.*

Lemma 1.3 *If $z_1, \dots, z_j$ are convex combinations of $x_1, \dots x_m$, then any convex combination of z_i is a convex combination of $x_1, \dots x_m$.*

So, convex hull $\mathrm{conv}(F)$ is just set of all convex combinations of elements of F.

If $F \subset \mathbb{R}^n$, and $z \in \text{conv}(F)$ then there exists $x_1, \dots, x_{n+1}$ such that z is a convex combination of x_i.

If $F \subset \mathbb{R}^n$ and F is compact, than $\text{conv}(F)$ is compact.

Separating hyperplane:

Lemma 1.4 *If $A, B \subset \mathbb{R}^n$ are convex sets, $A \cap B = \emptyset$, then there exists linear function ϕ and $b \in \mathbb{R}$ such that $\phi(x) \leq b$ for $x \in A$ and $\phi(x) \geq b$ for $x \in B$. If A is open, then $\phi(x) < b$ for $x \in A$. If A is closed and B is compact, then we can choose ϕ and b so that $\phi(x) \leq b$ for $x \in A$ and $\phi(x) > b$ for $x \in B$.*

Corollary: open (closed) convex set C can be written as intersection of open (closed) halfspaces. Namely, for each point $y \notin C$ take $H = \{x : \phi(x) < b\}$. H open halfspace such that $C \subset H$ and $y \notin H$. Single point set is compact, so by second part of the lemma for closed convex C halfspace $H = \{x : \phi(x) \leq b\}$ contains C and $y \notin H$.

Remark: Typically intersection above must be infinite. Set which can be written as finite intersection of closed halfspaces is called convex polyhedral set.

Supporting plane:

Lemma 1.5 *If A is convex and x belongs to boundary of A, then there exists linear function ϕ such that for all $y \in A$*

$$\phi(y) \leq \phi(x)$$

Remark: If F is open or closed and satisfies one of properties above, then F is convex.

1.1 Convex functions

Function $f : S \to \mathbb{R} \cup \infty$ is convex if and only if for all $x, y \in S$ and all $t \in [0, 1]$ point $tx + (1-t)y$ belongs to S and

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Function f is concave if and only if $-f$ is convex.

Function f is strictly convex if in the inequality above equality holds only for $x = y$ or $t \in \{0, 1\}$.

Suppose f is finite and convex on interval $[a, b]$. We can write $x \in [a, b]$ as convex combination of a, b:

$$x = \frac{b-x}{b-a}a + \frac{x-a}{b-a}b$$

so by convexity of f

$$f(x) \leq \frac{b-x}{b-a}f(a) + \frac{x-a}{b-a}f(b).$$

Subtracting $f(a)$ from both sides we get

$$f(x) - f(a) \leq \frac{x-a}{b-a}(f(b) - f(a))$$

which for $x \neq a$ gives

$$\frac{f(x) - f(a)}{x-a} \leq \frac{f(b) - f(a)}{b-a}$$

Similarly, for $x \neq b$ we have

$$\frac{f(b) - f(a)}{b-a} \leq \frac{f(b) - f(x)}{b-x}$$

For differentiable f when $x = a + h$ inequality

$$\frac{f(x) - f(a)}{x-a} \leq \frac{f(b) - f(a)}{b-a}$$

in the limit gives

$$f'(a) \leq \frac{f(b) - f(a)}{b-a}.$$

Similarly

$$\frac{f(b) - f(a)}{b-a} \leq f'(b).$$

Norms give important example of convex functions. Function usually denoted as $\|x\|$ on $\mathbb{R}^n$ is a norm if and only if it satisfies the following properties

- for all $t \in \mathbb{R}$ $\|tx\| = |t|\|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$
- $\|x\| = 0$ implies $x = 0$

Norm is convex:

$$\|tx + (1-t)y\| \leq \|tx\| + \|(1-t)y\| = t\|x\| + (1-t)\|y\|.$$

Example: for $p \geq 1$ l^p norm

$$\|x\|_{l^p} = \left(\sum_{i=1}^{n} |x|^p\right)^{1/p}.$$

In particular for $p = 2$ we get usual euclidean norm.

If function ϕ satisfies $\phi(tx) = |t|x$ and set

$$B = \{x : \phi(x) < 1\}$$

is a bounded convex set, then ϕ is a norm.

Important property: function f is (strictly) convex if and only if its restriction to any line is (strictly) convex.

Lemma 1.6 *Function f is convex if and only if its epigraph*

$$\mathrm{epi}(f) = \{(x,y) \in S \times \mathbb{R} : f(x) \leq y\}$$

is a convex set.

Example: If K is a convex set than indicator $I_k(x) = 0$ for $x \in K$ and $I_k(x) = \infty$ for $x \notin K$ is convex.

For convex f and all $t \in \mathbb{R}$ sublevel set $\{x : f(x) \leq t\}$ is convex.

1.2 Using properties convex functions

Example: We can use properties above to show that for $p \geq 1$ l^p norm defined above is really a norm. Namely, it is easy to see that for $t \in \mathbb{R}$ we have

$$\|tx\|_{l^p} = |t|\|x\|_{l^p}$$

so by properties above it enough to show that

$$B = \{x : \|x\|_{l^p} < 1\}$$

is convex. But we have

$$B = \{x : \|x\|_{l^p}^p < 1\}$$

which is a sublevel set of $\|x\|_{l^p}^p$. So it is enough to show that $\|x\|_{l^p}^p$ is convex.

But

$$\|x\|_{l^p}^p = \sum_{i=1}^{n} |x_i|^p.$$

Sum of convex functions is convex, so it is enough to show that $f(y) = |y|^p$ is convex on real line. But this follows from first order criterion of convexity (next subsection) since for $p > 1$ out f is differentiable and increasing. For $p = 1$ we can easily check convexity directly from definition.

Remark: In general, when f^p is convex and $p > 1$ there is no reason for f to be convex. However, in case of l^p norm (or more general when $f(tx) = |t|f(x)$) we can use B to prove that f is a norm (hence convex).

1.3 Back to convex functions

First order criterion of convexity: differentiable f is convex if and only if domain S of f is convex and derivative of f is nondecreasing, that is for all $x, y \in S$ we have

$$(f'(y) - f'(x))(y - x) \geq 0$$

Using gradient we can write this as

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0.$$

If above for $x \neq y$ we have strict inequality, then f is strictly convex.

Proof: Both convexity and property above depend only on restriction of f to lines. So it is enough to prove then for functions of single variable. We already proved that for convex differentiable f derivative is nondecreasing. To prove the converse assume that derivative of f is nondecreasing. Differentiating we get

$$\partial_t(f(x+t) - (f(x) + f'(x)t)) = f'(x+t) - f'(x)$$

so nondecreasing derivative implies

$$\partial_t(f(x+t) - (f(x) + f'(x)t)) \geq 0$$

for $t \geq 0$, hence

$$f(x+t) \geq f(x) + f'(x)t$$

for $t \geq 0$. But the same argument works also for $t \leq 0$, so the inequality above is valid for all t. Now, let $x = ty + (1-t)z$ be convex combination of y and z. We have $y = x + \alpha$, $z = x + \beta$ and $t\alpha + (1-t)\beta = 0$. Also

$$f(y) = f(x+\alpha) \geq f(x) + f'(x)\alpha,$$

$$f(z) = f(x+\beta) \geq f(x) + f'(x)\beta$$

so

$$tf(y) + (1-t)f(z) \geq tf(x) + f'(x)t\alpha + (1-t)f(x) + f'(x)(1-t)\beta$$

$$= f(x) + f'(x)(t\alpha + (1-t)\beta) = f(x)$$

so indeed f is convex. □

As a corollary we get another first order criterion of convexity: differentiable f is convex if and only if domain S of f is convex and for all $x, y \in S$ we have

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

Last criterion can be weakened in following way: Assume that f is taking only finite values, domain S of f is convex, f is continuous and set of points such that f is differentiable and condition above holds is dense in S. Then f is convex.

On the other hand if S is open and f is convex, then f has property above.

Related properties: if f is convex and $f'(x) = 0$ then x is global minimum of f. Local minimum of convex function is also global minimum. If f is strictly convex, then there is at most one local minimum.

Second order criterion: f having two derivatives is convex if and only if domain S of f is convex and second derivative of f is positive definite, that is for each $x \in S$ and $h \in \mathbb{R}^n$ we have

$$f''(x)(h,h) \geq 0$$

Again, if we have strict inequality for $h \neq 0$, then f is strictly convex.

Remark: Second order criterion of convexity is a consequence of first order criterion and calculus characterization of nondecreasing functions on real line.

Example: most our previous examples can be easily proved convex using second derivative criterion.

Example: quadratic function $(Qx, x) + (b, x) + c$ is convex if and only if Q is positive definite.

1.4 Operations on convex sets

- image and counterimage of convex set by affine function is convex
- in particular scaling and translation of convex set is convex
- cartesian product of convex sets is convex
- intersection of convex sets is convex
- complex sum $A + B = \{x + y : x \in A, y \in B\}$ is convex for convex A and B

Example: A rectangle, i.e., a set of the form $\{x \in \mathbb{R}^n : \alpha_i \leq x_i \leq \beta_i, \quad i = 1, \ldots, n\}$ is a convex set. Namely, a rectangle is an intersection of halfspaces $L_i = \{x : \alpha_i \leq x_i\}$ and $U_i = \{x : x_i \leq \beta_i\}$. A rectangle is sometimes called a hyperrectangle when $n > 2$.

Alternatively, we can treat a rectangle as a cartesian product of intervals.

Example: When A is a fixed linear operator and a is a fixed vector than

$$\{x : Ax = a\}$$

is a convex set as counterimage of one point set.

Example: Set of $x \in \mathbb{R}^n$ such that $x \geq 0$ is a convex set as cartesian product of convex sets $\{x_i : x_i \geq 0\} \subset \mathbb{R}$.

Example: For fixed $b \in \mathbb{R}^n$ set

$$\{x : x \geq b\}$$

is convex as translation of set from previous example.

Example: When B is a linear operator, and b is a fixed vector, then set

$$\{x : Bx \geq b\}$$

is a convex set as counterimage of convex set from previous example.

Example: Solution set of general linear programming problem is a convex set as intersection of set above and set from example about linear equalities.

1.5 Operations on convex functions

- nonnegative linear combination of convex functions is convex,
- supremum of family of convex functions is convex
- parametric minimum of convex function: if $f(x, y)$ is convex function and K is convex set then
$$g(x) = \inf_{y \in K} f(x, y)$$
is a convex function

Remark: Nonnegative linear combination means all coefficients are nonnegative. Functions may be negative.

Remark: Sum of convex functions is linear combination with all coefficients equal to 1, so is covered as nonnegative linear combination.

Example: Assume domain of f is open and f only takes finite values. f is convex if and only if domain of f is convex and f is supremum of family of affine functions.

Example: Distance to convex set K:

$$d(x, K) = \inf_{y \in K} \|x - y\|.$$

This follows from parametric minimum property and would be tricky in other ways.

1.6 Jensen inequality

If $K \subset \mathbb{R}^n$ is a convex set, f is convex function defined on K and μ is probability measure on K, then
$$f(\int x d\mu) \leq \int f(x) d\mu.$$
Note: x is vector valued function, we integrate such functions integrating all coordinates separately.

Alternatively, if X is random variable with values in K, then

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

In particular, for finite number of points x_i, $i = 1, ..., m$ and weights $t_i \geq 0$ such that $\sum_{i=1}^m t_i = 1$ we have

$$f(\sum_{i=1}^m t_i x_i) \leq \sum_{i=1}^m t_i f(x_i).$$

Example: $-\log(x)$ is convex on $(0, \infty)$. Using Jensen inequality for $p_i, x_i > 0$ such that $\sum p_i = 1$ we get

$$\sum p_i \log(x_i) \leq \log(\sum p_i x_i).$$

Passing to exponentials we get

$$\prod x_i^{p_i} = \exp(\sum p_i \log(x_i)) \le \sum p_i x_i$$

which is called generalized arithmetic-geometric mean inequality.

1.7 Convex optimization

When $g_i : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ are convex functions, then $S = \{x : g_i(x) \le 0\}$ is a convex set. When f is defined and convex on S then problem

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \le 0 \end{array}$$

is called (constrained) convex optimization problem. For linear (or more generally affine) function g_i we can use equality constraint $g_i(x) = 0$, namely we write equality as conjunction of two inequalities $g_i(x) \le 0$ and $-g_i(x) \ge 0$ (for affine g_i both g_i and $-g_i$ is convex, otherwise w could not do this).

Example: linear programming problem, LASSO.

Definition. We say that x if a local minimum of f if there is a neighborhood U of x such that

$$x = \text{argmin}_{x \in U} f(x),$$

that is x is a minimum of f restricted to U.

Similarly we define local maximum.

Lemma 1.7 *In convex optimization problem local minimum is also global minimum.*

Proof: If $f(y) < f(x)$, then for all $t \in (0,1)$ by convexity we have

$$f((1-t)x + ty) \le (1-t)f(x) + tf(y) < (1-t)f(x) + tf(x) = f(x)$$

But when U is a neighborhood of x, then for $t > 0$ and small enough we have

$$(1-t)x + ty \in U$$

which by previous inequality means that x is not a local minimum. Consequently if x is local minimum then for all y in domain of f we have $f(y) \ge f(x)$, so f is global minimum. □

We will look more at convex optimization, but previous lemma shows that convex optimization problems are easier then general optimization problems.

1.8 Examples of convex problems

1.8.1 Support vector machines (SVM)

Example: Given sequence of points $x_i \in \mathbb{R}^n$, $i = 1, \ldots, m$ with classification $y_i \in \{-1, 1\}$ we want to find optimal separating plane. Optimal means maximizing oriented distance to both classes. That is, we want to find maximal M such that there is β_0 and vector $\beta = (\beta_1, \ldots, \beta_n)$ such that $\|\beta\|_2 = 1$ and for $i = 1, \ldots, m$

$$y_i(\langle x_i, \beta\rangle + \beta_0) \geq M.$$

As above, this contains non-convex constraint $\|\beta\|_2 = 1$. However, for positive M dividing β by M we get new problem: minimize $\|\beta\|^2$ under constraints

$$y_i(\langle x_i, \beta\rangle + \beta_0) \geq 1$$

which is quadratic problem. When M is non-positive the transformed problem above is infeasible.

For $M = 0$ we get linear problem, for negative M we get problem of maximizing $\|\beta\|_2$ with constraints like above, but with 1 on the right hand side replaced by -1. This is non-convex problem, so usually we replace it by problem with relative slack variables that is for for $i = 1, \ldots, m$ we have

$$y_i(\langle x_i, \beta\rangle + \beta_0) \geq M(1 - \zeta_i)$$

where $\zeta_i \geq 0$ and $\sum \zeta_i$ is bounded by some constant. Now, dividing by M we get problem of minimizing $\|\beta\|^2$ under constraints

$$y_i(\langle x_i, \beta\rangle + \beta_0) \geq (1 - \zeta_i),$$

and $\zeta_i \geq 0$, $\sum \zeta_i \leq C$.

1.8.2 Basis pursuit

Example: basis pursuit as linear programming problem. Consider problem of minimizing $\|x\|_1 = \sum_{i=1}^n |x_i|$ under condition $Ax = b$. Writing $x_i = t_i - s_i$ we see that instead we can minimize $\sum_{i=1}^n (t_i + s_i)$ under condition $A(t - s) = b$ which is linear programming problem.

In similar way for LASSO we get quadratic optimization problem.

1.9 Further reading

Stephen Boyd, Lieven Vandenberghe, Convex Optimization, Chapters 2 and 3.

2 Appendix, pathologies and regularity

In some places we had assumptions that convex sets are open or close and some extra assumptions about regularity of convex functions. Below we show few examples showing that without regularity assumptions convex sets or convex functions may show quite bad behaviour. Those examples are mostly theoretical: functions coming from real problems typically are regular. However, they play role in logical structure of theory, showing that we need to be careful formulating our results. We also give some extra results about regularity.

Bad example: Let $B = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ be unit disc on the plane. Let A be arbitrary subset of unit circle $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Then $A \cup B$ is a convex set. This example shows to convex sets may behave very badly at boundary.

Bad example: Let B and C be as in previous example. Let h be arbitrary function on C such that $h \geq 0$. Put $f(x) = 0$ for $x \in B$ and $f(x) = h(x)$ for $x \in C$. Then f is convex on $B \cup C$, but f may be quite irregular on C.

Good property: When convex set S has nonempty interior, then S is contained in the closure of interior of S. This essentially means that all bad examples must be like our $A \cup B$ example: they are sum of interior and some subset of boundary.

Good property: If $S \subset \mathbb{R}^n$ is a convex set, f is convex and finite on S, then f is continuous in the interior $\text{Int}(S)$ of S. In other words, discontinuity is possible only at boundary of S. Next, f is Lipschitz continuous on compact subsets of $\text{Int}(S)$. f is almost everywhere differentiable on $\text{Int}(S)$. In particular, when $\text{Int}(S)$ is nonempty, then set of point at which f is differentiable is dense in S.

One can also define generalized derivatives. With assumptions as above f has two generalized derivatives in $\text{Int}(S)$. The second generalized derivative in general is a measure taking values in positive definite matrices.

2.1 Appendix, regular approximation

With notation as before, on $\text{Int}(S)$ convex and finite f is a pointwise limit of convex and smooth functions. This means that many properties when proven for convex and smooth functions must hold for arbitrary convex functions.