Math 484: Nonlinear Programming[1] | Mikhail Lavrov

Chapter 5, Lecture 3: Separating and Supporting Inequalities

March 11, 2019 | *University of Illinois at Urbana-Champaign*

1 Two applications of Bolzano–Weierstrass

Last time, we discussed the Bolzano–Weierstrass theorem. Today, we're going to see some results it's good for. For reasons of time, we skip the proof; if you want to see it, it's in the lecture notes for the previous class.

Theorem 1.1 (Bolzano–Weierstrass). *Let $S \subseteq \mathbb{R}^n$ be a closed and bounded set, and let $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ be any sequence of points in S.*

Then we can pick out a subsequence of these points that has a limit $\mathbf{x} \in S$. Formally, we can find indices $i_1 < i_2 < i_3 < \dots$ such that the sequence $\mathbf{x}^{(i_1)}, \mathbf{x}^{(i_2)}, \mathbf{x}^{(i_3)}, \dots$ converges to $\mathbf{x} \in S$.

Corollary 1.1 (Extreme value theorem). *If $S \subseteq \mathbb{R}^n$ is a closed and bounded set, and $f : S \to \mathbb{R}$ is a continuous function, then f has a global maximizer on S.*

(By applying this theorem to $-f$, we conclude that f also has a global minimizer.)

Proof. Our first step is to show (by contradiction) that f is bounded above on S.

Suppose not; then we can find points where f has arbitrarily large values. Let $\mathbf{x}^{(1)} \in S$ be a point with $f(\mathbf{x}^{(1)}) > 1$; let $\mathbf{x}^{(2)} \in S$ be a point with $f(\mathbf{x}^{(2)}) > 2$, and so on, with $f(\mathbf{x}^{(k)}) > k$.

Then the sequence $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ has a subsequence $\mathbf{x}^{(i_1)}, \mathbf{x}^{(i_2)}, \dots$ which converges to some $\mathbf{x}^* \in S$.

Along this subsequence, we know that

- $\lim_{k\to\infty} f(\mathbf{x}^{(i_k)}) = +\infty$, because $f(\mathbf{x}^{(i_k)}) > i_k$ and $i_k \to \infty$ as $k \to \infty$,
- but also that $\lim_{k\to\infty} f(\mathbf{x}^{(i_k)}) = f(\mathbf{x}^*)$, because f is continuous.

This is a contradiction. So f must be bounded above.

Let M be the least upper bound on the set of values $\{f(\mathbf{x}) : \mathbf{x} \in S\}$. Our second step is to show that there is some $\mathbf{x}^*$ with $f(\mathbf{x}^*) = M$.

For all k, we know that $M - \frac{1}{k}$ is *not* an upper bound on f, so we can find some $\mathbf{x}^{(k)} \in S$ with $f(\mathbf{x}^{(k)}) > M - \frac{1}{k}$. This gives us a sequence $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots$; it has a subsequence $\mathbf{x}^{(i_1)}, \mathbf{x}^{(i_2)}, \mathbf{x}^{(i_3)}, \dots$ converging to $\mathbf{x}^* \in S$.

Since we have $M - \frac{1}{i_k} < f(\mathbf{x}^{(i_k)}) \le M$, we can take the limit as $k \to \infty$ and get the inequality $M \le f(\mathbf{x}^*) \le M$. So $f(\mathbf{x}^*) = M$ and $\mathbf{x}^*$ is the global maximizer we wanted. □

[1] This document comes from the Math 484 course webpage: https://faculty.math.illinois.edu/~mlavrov/courses/484-spring-2019.html

Corollary 1.2 (Support theorem; Theorem 5.1.9)**.** *Let $C \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{z} \in \mathrm{bd}(C)$. Then there is an inequality supporting C at $\mathbf{z}$: we can choose $\mathbf{u} \in \mathbb{R}^n$ with $\|\mathbf{u}\| = 1$ such that*

$$\mathbf{u} \cdot \mathbf{x} \le \mathbf{u} \cdot \mathbf{z}$$

for all $\mathbf{x} \in C$.

Proof. Since $\mathbf{z} \in \mathrm{bd}(C)$, we can find a sequence $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots \notin C$ converging to $\mathbf{z}$. For concreteness, here is one way to do this. For every k, the ball of radius $\frac{1}{k}$ around $\mathbf{z}$ contains a point outside C, so let $\mathbf{y}^{(k)}$ be one such point: it will satisfy $\|\mathbf{y}^{(k)} - \mathbf{z}\| \le \frac{1}{k}$.

For each $\mathbf{y}^{(k)}$, there is an inequality separating it from $\mathrm{cl}(C)$: some $\mathbf{a}^{(k)}$ such that

$$\mathbf{a}^{(k)} \cdot \mathbf{x} < \mathbf{a}^{(k)} \cdot \mathbf{y}^{(k)} \text{ for all } \mathbf{x} \in C.$$

We saw this as a corollary of the obtuse angle criterion in the previous lecture. Dividing through by $\|\mathbf{a}^{(k)}\|$, we get

$$\frac{\mathbf{a}^{(k)}}{\|\mathbf{a}^{(k)}\|} \cdot \mathbf{x} < \frac{\mathbf{a}^{(k)}}{\|\mathbf{a}^{(k)}\|} \cdot \mathbf{y}^{(k)} \text{ for all } \mathbf{x} \in C,$$

so $\mathbf{u}^{(k)} = \frac{\mathbf{a}^{(k)}}{\|\mathbf{a}^{(k)}\|}$ is a vector with $\|\mathbf{u}\| = 1$ that satisfies $\mathbf{u}^{(k)} \cdot \mathbf{x} < \mathbf{u}^{(k)} \cdot \mathbf{y}^{(k)}$ for all $\mathbf{x} \in C$.

The points $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots$ are a sequence of points in the closed and bounded set $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$. This means we can pick out a subsequence of them converging to some $\mathbf{u} \in \mathbb{R}^n$ with $\|\mathbf{u}\| = 1$.

Taking the limit of the inequality $\mathbf{u}^{(k)} \cdot \mathbf{x} < \mathbf{u}^{(k)} \cdot \mathbf{y}^{(k)}$ for this subsequence, we get $\mathbf{u} \cdot \mathbf{x} \le \mathbf{u} \cdot \mathbf{z}$ for all $\mathbf{x} \in C$, which was what we wanted. □

There are two ideas that show up in these proofs that are worth remembering. They show up again and again in applications of Bolzano–Weierstrass, not only in this subject, but in many others.

The first is the way we obtain the sequence we want. In both of our applications, we didn't start out with a sequence. Rather, we started out with a hypothesis that something is true for arbitrarily large, or arbitrarily small, values:

- If f is not bounded above, there are points $\mathbf{x}$ with arbitrarily large values of f.
- If M is a least upper bound on f, there are points $\mathbf{x}$ with $f(\mathbf{x})$ arbitrarily close to M.
- If $\mathbf{z}$ is a boundary point of C, then there are points $\mathbf{y} \notin C$ arbitrarily close to $\mathbf{z}$.

To use this hypothesis, we choose a sequence of points where the parameter in question (value of f, or difference $M - f(\mathbf{x})$, or distance $\|\mathbf{y} - \mathbf{z}\|$) gets arbitrarily close to its extreme. Then we apply Bolzano–Weierstrass.

The second idea shows up in the second proof. If we just took arbitrary inequalities $\mathbf{a} \cdot \mathbf{x} < \mathbf{a} \cdot \mathbf{y}$ at each step, we would have no control over $\mathbf{a}$: it can be any vector in $\mathbb{R}^n$ other than $\mathbf{0}$. However, we can normalize to enforce $\|\mathbf{a}\| = 1$, and then suddenly we are working over a closed and bounded set.

We can do this whenever we are looking at some property of nonzero vectors in $\mathbb{R}^n$ that doesn't change when we scale them: assuming that the norm is 1 lets us work over a very convenient closed and bounded set, and apply either Bolzano–Weierstrass or the extreme value theorem.

2 Subgradients of convex functions

Recall that we proved earlier in this course that for any convex function $f : \mathbb{R}^n \to \mathbb{R}$ with continuous first derivatives, and any $\mathbf{x}^* \in \mathbb{R}$, we have an inequality

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*).$$

That is, the graph of f always stays above its tangent lines or (in higher dimensions) tangent hyperplanes.

We are about to end up in a situation where our convex functions are unlikely to be differentiable everywhere. However, we can still make a similar claim.

Lemma 2.1. *If $C \subseteq \mathbb{R}^n$ is a convex set, $f : C \to \mathbb{R}$ is a convex function, and $\mathbf{x}^* \in \operatorname{int}(C)$, then there exists some vector $\mathbf{d} \in \mathbb{R}^n$ such that*

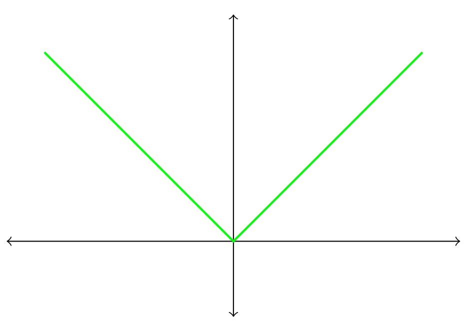
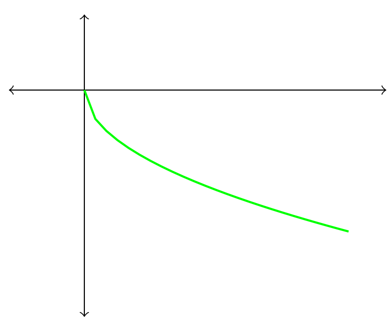
$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \mathbf{d} \cdot (\mathbf{x} - \mathbf{x}^*).$$

That is, at any point $\mathbf{x}^*$ in the interior of the domain of f, we can draw a hyperplane that *acts* like a tangent hyperplane at $\mathbf{x}^*$: just like the tangent hyperplane, it's always a lower bound.

Such a vector $\mathbf{d}$ is called a *subgradient* of f at $\mathbf{x}^*$. The vector $\mathbf{d}$ is not guaranteed to be unique (it is unique only in the special case where f is differentiable at $\mathbf{x}^*$ and we can set $\mathbf{d} = \nabla f(\mathbf{x}^*)$).

We write $\partial f(\mathbf{x}^*)$ for the set of all subgradients of f at $\mathbf{x}^*$. This set is called the *subdifferential* of f at $\mathbf{x}^*$.

Before we proceed with the proof, here are two representative examples that you should keep in mind.

The first example is $f(x) = |x|$: a convex function which is not differentiable. We have $|x| \geq dx$ for any $d \in [-1, 1]$; therefore, at $x^* = 0$, then subdifferential of f is the entire interval $[-1, 1]$. At other points, the only subgradient is the derivative.

The second example is $f(x) = -\sqrt{x}$: a convex function defined on $C = [0, \infty)$. It does not have a subgradient at $x^* = 0$, because the only possible tangent line is vertical. This shows that the hypothesis $x^* \in \operatorname{int}(C)$ is necessary.

Proof. The basic idea is to use the epigraph epi(f). Recall that for a function f defined on $C \subseteq \mathbb{R}^n$, we define

$$\operatorname{epi}(f) = \{(\mathbf{x}, y) \in C \times \mathbb{R} : y \geq f(\mathbf{x})\}.$$

When C is a convex set and f is a convex function on that set, the epigraph epi(f) is also convex (a convex subset of $\mathbb{R}^{n+1}$).

Pick an $\mathbf{x}^* \in \operatorname{int}(C)$. Then the point $(\mathbf{x}^*, f(\mathbf{x}^*))$ is a boundary point of C. By the support theorem, there is some vector $(\mathbf{a}, b)$ with norm 1 such that

$$(\mathbf{a}, b) \cdot (\mathbf{x}^*, f(\mathbf{x}^*)) \geq (\mathbf{a}, b) \cdot (\mathbf{x}, y)$$

for all $(\mathbf{x}, y) \in \operatorname{epi}(f)$. We can rewrite it as

$$\mathbf{a} \cdot \mathbf{x}^* + bf(\mathbf{x}^*) \geq \mathbf{a} \cdot \mathbf{x} + by.$$

In particular, since this is true for all $(\mathbf{x}, y) \in \operatorname{epi}(f)$, it's true for all points of the form $(\mathbf{x}, f(\mathbf{x}))$ with $\mathbf{x} \in C$, and we get the inequality

$$\mathbf{a} \cdot \mathbf{x}^* + bf(\mathbf{x}^*) \geq \mathbf{a} \cdot \mathbf{x} + bf(\mathbf{x}),$$

or

$$b \cdot f(\mathbf{x}) \leq b \cdot f(\mathbf{x}^*) - \mathbf{a} \cdot (\mathbf{x} - \mathbf{x}^*).$$

If we show that $b < 0$, then dividing by b will reverse the inequality, and we will get

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) - \frac{\mathbf{a}}{b} \cdot (\mathbf{x} - \mathbf{x}^*)$$

and so the vector $-\frac{\mathbf{a}}{b}$ is the subgradient we want.

There are two parts to this.

1. We have $b \leq 0$. To see this, take the inequality

   $$(\mathbf{a}, b) \cdot (\mathbf{x}^*, f(\mathbf{x}^*)) \geq (\mathbf{a}, b) \cdot (\mathbf{x}, y)$$

   with $\mathbf{x} = \mathbf{x}^*$ and $y = f(\mathbf{x}^*) + 1$. This gives us

   $$\mathbf{a} \cdot \mathbf{x}^* + b \cdot f(\mathbf{x}^*) \geq \mathbf{a} \cdot \mathbf{x}^* + b \cdot (f(\mathbf{x}^*) + 1)$$

   and cancelling like terms on both sides leaves us with $0 \geq b$.

2. We cannot have $b = 0$. Here is where we use the fact that $\mathbf{x}^*$ is an interior point of C.

   If we had $b = 0$, then our supporting inequality would simplify to $\mathbf{a} \cdot \mathbf{x}^* \geq \mathbf{a} \cdot \mathbf{x}$ for all $\mathbf{x} \in C$: a supporting inequality for $\mathbf{x}^*$ itself. We still have $\|\mathbf{a}\| = 1$ so in particular $\mathbf{a} \neq \mathbf{0}$.

   Such an inequality can only hold if $\mathbf{x}^*$ is a boundary point: after all, arbitrarily close to $\mathbf{x}^*$, we have points $\mathbf{y}$ where $\mathbf{a} \cdot \mathbf{y}$ is bigger than $\mathbf{a} \cdot \mathbf{x}^*$. This contradicts our assumption.

Therefore $b < 0$, and we get a subgradient. □