Module 6: Nonlinear Programs

CO 250: Introduction to Optimization

David Duan, 2019 Spring

Contents

Lecture 26. Convexity

26.1 Nonlinear Programs

Definition 26.1.1 A **nonlinear program** (NLP) is a program of the form

$$\begin{array}{ll} \min & f(x) \\ s.t. & g_i(x) \leq 0 \quad (i = 1, \ldots, k) \end{array} \qquad (P)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, k$.

Example 26.1.2

$$\begin{array}{ll} \min & x_2 \\ s.t. & -x_1^2 - x_2 + 2 \leq 0 \\ & x_2 - 3/2 \leq 0 \\ & x_1 - 3/2 \leq 0 \\ & -x_1 - 2 \leq 0 \end{array}$$

Remark 26.1.3 We may assume $f(x)$ is a *linear* function, i.e., $f(x) = c^T x$.

Remark 26.1.4 We can rewrite (P) as

$$\begin{array}{ll} \min & \lambda \\ s.t. & \lambda \geq f(x) \\ & g_i(x) \leq 0 \quad (i = 1, \ldots, k) \end{array} \qquad (Q)$$

The optimal solution to (Q) will have $\lambda = f(x)$.

Remark 26.1.5 Recall $x_j \in \{0, 1\} \iff x_j(1 - x_j) = 0$. Thus NLPs generalize binary IPs:

$$\begin{array}{ll} \max & c^T x \\ s.t. & Ax \leq b \\ & x_j \in \{0,1\} \quad (j = 1, \ldots, n) \end{array} \iff \begin{array}{cl} \min & -c^T x \\ s.t. & Ax \leq b \\ x_j(1 - x_j) \leq 0 & (j = 1, \ldots, n) \\ -x_j(1 - x_j) \leq 0 & (j = 1, \ldots, n) \end{array}$$

Remark 26.1.6 Recall $x_j \in \mathbb{Z} \iff \sin(\pi x) = 0$. Thus NLPs generalize pure IPs:

$$\begin{array}{ll} \max & c^T x \\ s.t. & Ax \leq b \\ & x_j \in \mathbb{Z} \quad (j = 1, \ldots, n) \end{array} \iff \begin{array}{ll} \min & -c^T x \\ s.t. & Ax \leq b \\ & \sin(\pi x) = 0 \quad (j = 1, \ldots, n) \end{array}$$

26.2 Convex Sets

Definition 26.2.1 Consider (P) := $\min\{f(x) : x \in S\}$. We call $x \in S$ a **local optimum** if there exists $\delta > 0$ such that

$$\forall x' \in S : \|x' - x\| \leq \delta \implies f(x) \leq f(x').$$

Proposition 26.2.2 Consider (P) := $\min\{f(x) : x \in S\}$. If S is *convex* and x is a *local optimum*, then x is optimal.

Proof. Let x be the local optimum. Suppose to the contrary that $\exists x' \in S$ with $c^T x' < c^T x$. Let $y = \lambda x' + (1-\lambda)x$ for $\lambda > 0$ small. Since S is convex, $y \in S$. For λ small, $\|y - x\| \leq \delta$, then

$$\begin{aligned} c^T y &= c^T(\lambda x' + (1-\lambda)x) \\ &= \lambda c^T x' + (1-\lambda)c^T x \\ &< \lambda c^T x + (1-\lambda)c^T x \\ &= c^T x. \end{aligned}$$

This contradicts the optimality of x. □

26.3 Convex Functions

Definition 26.3.1 A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if for all $a, b \in \mathbb{R}^n$, $f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda)f(b)$ for all $0 \leq \lambda \leq 1$.

Proposition 26.3.2 Let $g : \mathbb{R}^n \to \mathbb{R}$ be a *convex function* and $\beta \in \mathbb{R}$. It follows that $S = \{x \in \mathbb{R}^n : g(x) \leq \beta\}$ is a convex set.

Proof. Let $a, b \in S$ and $\lambda \in [0,1]$. Let $x = \lambda a + (1-\lambda)b$. We want to show that $x \in S$, i.e., $g(x) \leq \beta$:

$$\begin{aligned} g(x) &= g(\lambda a + (1-\lambda)b) \\ &\leq \lambda g(a) + (1-\lambda)g(b) && \text{convexity of } g \\ &\leq \lambda\beta + (1-\lambda)\beta && (a, b \in S) \\ &= \beta && \square \end{aligned}$$

Proposition 26.3.3 If all constraint functions g_i are convex for (P), then the feasible region of (P) is convex.

Proof. Let $S_i = \{x : g_i(x) \leq 0\}$. By Proposition 26.3.2, S_i is convex. The feasible region of (P) is $S_1 \cap S_2 \cap \cdots \cap S_k$. Since the intersection of convex sets is convex, the result follows. □

26.4 Convex Functions vs. Convex Sets

Definition 26.4.1 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. The **epigraph** of f is given by

$$epi(f) = \left\{ \begin{pmatrix} y \\ x \end{pmatrix} : y \geq f(x), x \in \mathbb{R}^n \right\} \subseteq \mathbb{R}^{n+1}.$$

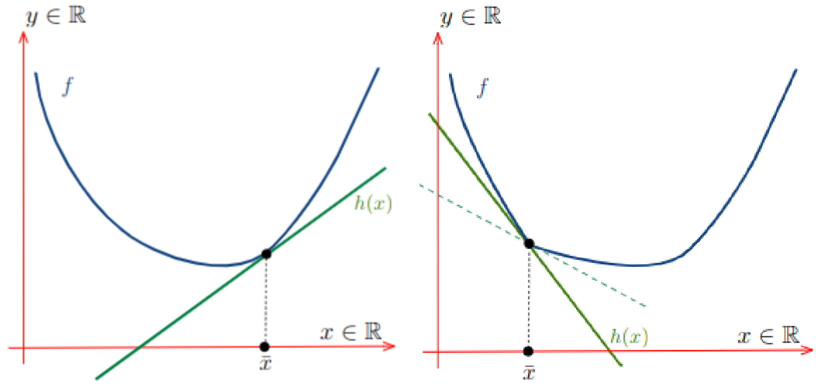
Proposition 26.4.2 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. It follows that f is convex if and only if $epi(f)$ is a convex set.

Proof. Suppose f is convex. Pick $\begin{pmatrix} \alpha \\ a \end{pmatrix}, \begin{pmatrix} \beta \\ b \end{pmatrix} \in epi(f)$ and $\lambda \in [0,1]$. Observe $f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda)f(b) \leq \lambda\alpha + (1-\lambda)\beta$ as $f(a) \leq \alpha$ and $f(b) \leq \beta$. Thus $\lambda \begin{pmatrix} \alpha \\ a \end{pmatrix} + (1-\lambda) \begin{pmatrix} \beta \\ b \end{pmatrix} = \begin{pmatrix} \lambda\alpha + (1-\lambda)\beta \\ \lambda a + (1-\lambda)b \end{pmatrix} \in epi(f)$.

The other direction is left for exercise. □

Lecture 27. KKT Theorem

27.1 Subgradient



Definition 27.1.1 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $\bar{x} \in \mathbb{R}^n$. Then, $s \in \mathbb{R}^n$ is a **subgradient** of f at $\bar{x}$ if $h(x) := f(\bar{x}) + s^T(x - \bar{x}) \le f(x)$ for all $x \in \mathbb{R}^n$.

Remark 27.1.2 There are three key points in this definition:

- $h(x)$ is affine: Observe $h(x) = s^T x + b$ where $b = f(\bar{x}) - s^T \bar{x}$ is a constant vector.
- $h(\bar{x}) = f(\bar{x})$: The affine function $h(x)$ touches the convex function $f(x)$ at $\bar{x}$.
- h is a lower bound for f: We restrict that for all $x \in \mathbb{R}^n$, $h(x) \le f(x)$.

Remark 27.1.3 Remarks on subgradients and subdifferentials.

- The subgradient generalize the derivative to convex functions which are not necessarily differentiable. They arise in convex analysis, the study of convex functions, often in connection to convex optimization.
- Let $f : I \to \mathbb{R}$ be a real-valued function defined on an open interval of the real line. Such a function need not be differentiable at all points. For example, the absolute value function $f(x) = |x|$ is non-differentiable when $x = 0$. However, for any x_0 in the domain of the function one can draw a line which goes through the point $(x_0, f(x_0))$ and which is everywhere either touching or below the graph of f. The slope of such a line is called a subgradient or subderivative (because the line is under the graph of f).
- Rigorously, a subderivative of a function $f : I \to \mathbb{R}$ at a point $x_0 \in I$ open is $c \in \mathbb{R}$ such that $f(x) - f(x_0) \ge c(x - x_0)$ for all $x \in I$. We can show that the set of subderivatives at x_0 for a convex function is a non-empty closed interval $[a, b]$, where a and b are one-sided limits as $x \to x_0$, which are guaranteed to exist and satisfy $a \le b$. The set $[a, b]$ of all subderivatives is called the *subdifferential* of the function f at x_0. Since f is convex, if its subdifferential at x_0 contains exactly one subderivative, then f is differentiable at x_0. Putting together, a convex function $f : I \to \mathbb{R}$ is differentiable at x_0 if and only if the subdifferential is made up of only one point, which is the derivative at x_0.
- A point x_0 is a global minimum of a convex function f if and only if zero is contained in the subdifferential, that is, one may draw a horizontal "subtangent line" to the graph of f at

$(x_0, f(x_0))$. This is a generalization of the fact that the derivative of a function differentiable at a local minimum is zero.

- The concepts of subderivative and subdifferential can be generalized to functions of several variables. If $f : U \to \mathbb{R}$ is a real-valued convex function defined on a convex open set in the Euclidean space $\mathbb{R}^n$, a vector v in the space is called a subgradient at $x_0 \in U$ if for any $x \in U$ one has $f(x) - f(x_0) \geq v \cdot (x - x_0)$, where the dot denotes the dot product. The set of all subgradients at x_0 is called the subdifferential at x_0 and is denoted $\partial f(x_0)$. The subdifferential is always a non-empty convex compact set.

Example 27.1.4 Consider $f : \mathbb{R}^2 \to \mathbb{R}$ where $f(x) = -x_1 + x_2^2$ and $\bar{x} = (1,1)^T$. We claim that $(-1,2)^T$ is a subgradient of f at $\bar{x}$.

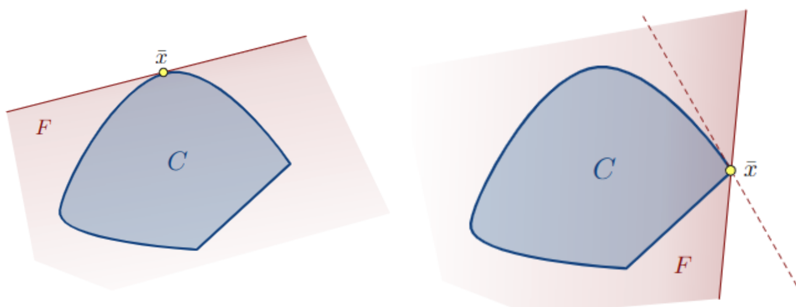
- Evaluate $h(x)$:

$$h(x) = f(\bar{x}) + s^T(x - \bar{x}) = 0 + (-1,2)(x - (1,1)^T) = -x_1 + 2x_2 - 1 \implies h(1,1) = f(1,1).$$

- Check $h(x) \leq f(x)$ for all $x \in \mathbb{R}^n$:

$$(x_2 - 1)^2 = x_2 - 2x_2 + 1 \geq 0 \implies -x_1 + 2x_2 - 1 \leq -x_1 + x_2^2. \quad \square$$

27.2 Supporting Halfspace



Definition 27.2.1 Let $C \in \mathbb{R}^n$ be a convex set and $\bar{x} \in C$. The halfspace $F = \{x : s^T x \leq \beta\}$ is supporting C at $\bar{x}$ if

1. $C \subseteq F$ and
2. $s^T \bar{x} = \beta$. That is, $\bar{x}$ is on the boundary of F.

Remark 27.2.2 Remarks on halfspaces.

- A halfspace is either of the two parts into which a plane divides the 3D Euclidean space. More generally, it is either of the two parts into which a hyperplane divides an affine plane. That is, the points that are not incident to the hyperplane are partitioned into two convex sets (i.e., half-spaces), such that any subspace connecting a point in one set to a point in the other must intersect the hyperplane.
- A 2D halfspace is called halfplane; a 1D halfspace is a ray.

- An open halfspace is specified by $a_1x_1 + \cdots + a_nx_n > b$; a closed halfspace is specified by $a_1x_1 + \cdots a_nx_n \geq b$. Here, one assumes that not all $a_1, \ldots, a_n$ are zero.
- The epigraph of a real-valued function f is a halfspace if and only if f is a real valued-affine function.

Remark 27.2.3. Remarks on support hyperplane.

- A supporting hyperplane of a set S in Euclidean space $\mathbb{R}^n$ is a hyperplane that has both of the following properties:
    1. S is entirely contained in one of the two closed halfspaces bounded by the hyperplane.
    2. S has at least one boundary point on the hyperplane.
- In 1D, C is a segment, F is a ray, and they have one endpoint together.

Proposition 27.2.4

- Let $g : \mathbb{R}^n \to \mathbb{R}$ be convex and let $\bar{x}$ where $g(\bar{x}) = 0$.
- Let s be a subgradient of g at $\bar{x}$.
- Let $C = \{x : g(x) \leq 0\}$.
- Let $F = \{x : h(x) := g(\bar{x}) + s^T(x - \bar{x}) \leq 0\}$.
- Then, F is a supporting halfspace of C at $\bar{x}$.

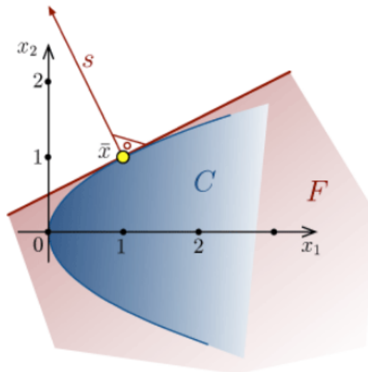
Remark 27.2.5 We make the following observations:

1. By 26.2.2, since g is a convex, C is a convex set.
2. F is a halfspace as $h(x)$ is an affine function.
3. $h(\bar{x}) = g(\bar{x}) = 0$, thus, $\bar{x}$ is on the boundary of F.

Proof.

1. $C \subseteq F$: Let $x \in C$, i.e., $g(x) \leq 0$. By definition of a subgradient, we know that $h(x) \leq g(x)$. It follows that $h(x) \leq g(x) \leq 0$. Hence $x \in F$.
2. $h(\bar{x}) = 0$: $h(\bar{x}) = g(\bar{x}) = 0$. □

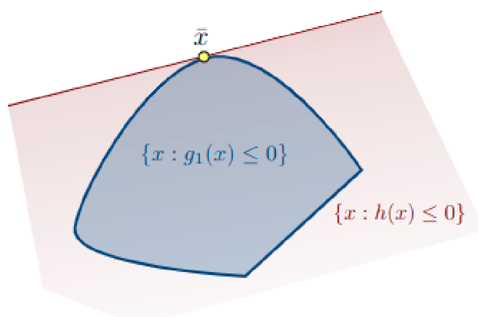
Example 27.2.6 Given $g(x) = x_2^2 - x_1$, $\bar{x} = (1, 1)^T$, $s = (-1, 2)^T$.



The subgradient $h(x)$ at $\bar{x}$ is $0 + (-1, 2)[(x_1, x_2)^T - (1, 1)^T] = -x_1 + 2x_2 - 1$. Let $F = \{x : -x_1 + 2x_2 \leq 1\}$. We see that F is the supporting halfspace of C.

27.3 LP Relaxation of NLP

Remark 27.3.1 We can use this proposition to construct relaxations of NLPs. Suppose we are given NLP $\min\{c^T x : g_i(x) \leq 0, i = 1, \ldots k\}$. Let $\bar{x}$ be a feasible solution, g_1 be convex, $g_1(\bar{x}) = 0$, and s be a subgradient for g_1 at $\bar{x}$. If we replace the nonlinear constraint $g_1(x) \leq 0$ with the linear constraint $h(x) = g_1(\bar{x}) + s^T(x - \bar{x}) \leq 0$, we get a relaxation.



This motivates the following proposition.

Proposition 27.3.2 Given NLP $\min\{c^T x : g_i(x) \leq 0, i = 1, \ldots k\}$, suppose $g_1, \ldots, g_k$ are all convex, $\bar{x}$ is a feasible solution, $g_i(\bar{x}) = 0$ for all $i \in I$ and $s^{(i)}$ is a subgradient for g_i at $\bar{x}$ for all $i \in I$. If $-c \in \text{cone}\{s^{(i)} : i \in I\}$, then $\bar{x}$ is optimal.

Proof. Rewrite the constraints as $P_i = s^{(i)}x \leq s^{(i)}\bar{x} - g_i(\bar{x})$ and we get a relaxation $\max\{-c^T x : P_i, i \in I\}$. From previous theorem, $\bar{x}$ is optimal for the relaxation if $-c \in \text{cone}\{s^{(i)} : i \in I\}$. This means that $\bar{x}$ is also optimal for the NLP. □

Example 27.3.3

$$\begin{aligned} \min \quad & -x_1 - x_2 \\ s.t. \quad & -x_2 + x_1^2 \leq 0 \\ & -x_1 + x_2^2 \leq 0 \\ & -x_1 + \tfrac{1}{2} \leq 0 \end{aligned}$$

Observe $\bar{x} = (1,1)^T$ is feasible, $I = \{1,2\}$, $(2,-1)^T$ is a subgradient for g_1 at $\bar{x}$, $(-1,2)^T$ is a subgradient for g_2 at $\bar{x}$. Since $-(-1,-1) \in \textbf{cone}\{(2,-1),(-1,2)\}$, we get $\bar{x}$ optimal.

27.4 KKT

Proposition 27.4.1 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $\bar{x} \in \mathbb{R}^n$. If the gradient $\nabla f(\bar{x})$ of f exists at $\bar{x}$, then it is a subgradient.

Definition 27.4.2 A feasible solution to $\bar{x}$ is a **Slater point** of $\min\{c^T x : g_i(x) \leq 0, i \in J\}$ if $g_i(\bar{x}) < 0$ for all $i \in J$, i.e., every inequality is satisfied strictly by $\bar{x}$.

Theorem 27.4.3 [KKT] Given NLP $\min\{c^T x : g_i(x) \leq 0, i \in J\}$, if

1. g_i is convex for all $i \in J$,
2. There exists a Slater point,
3. $\bar{x}$ is a feasible solution,
4. I is the set of indices i for which $g_i(\bar{x}) = 0$, and
5. For all $i \in I$ there exists a gradient $\nabla g_i(\bar{x})$ of g_i at $\bar{x}$.

Then $\bar{x}$ is optimal if and only if $-c \in \textbf{cone}\{\nabla g_i(\bar{x}) : i \in I\}$.

Proof. We proved the $\Longleftarrow$ direction in 27.3.5. The other direction is omitted. □