Any typos should be emailed to a_a_a@princeton.edu.

In the previous couple of lectures, we've been focusing on the theory of convex sets. In this lecture, we shift our focus to the other important player in convex optimization, namely, convex functions. Here are some of the topics that we will touch upon:

- Convex, concave, strictly convex, and strongly convex functions

- First and second order characterizations of convex functions

- Optimality conditions for convex problems

1 Theory of convex functions

1.1 Definition

Let's first recall the definition of a convex function.

Definition 1. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain is a convex set and for all x, y in its domain, and all $\lambda \in [0, 1]$, we have*

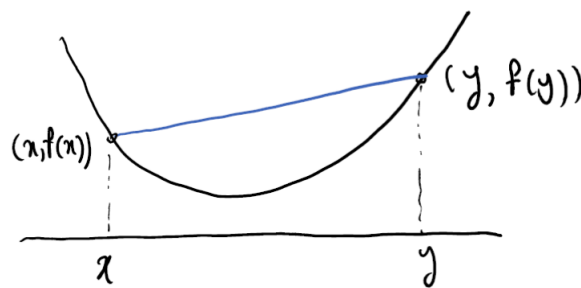$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Figure 1: An illustration of the definition of a convex function

- In words, this means that if we take any two points x, y, then f evaluated at any convex combination of these two points should be no larger than the same convex combination of $f(x)$ and $f(y)$.

- Geometrically, the line segment connecting $(x, f(x))$ to $(y, f(y))$ must sit above the graph of f.

- If f is continuous, then to ensure convexity it is enough to check the definition with $\lambda = \frac{1}{2}$ (or any other fixed $\lambda \in (0, 1)$). This is similar to the notion of midpoint convex sets that we saw earlier.

- We say that f is *concave* if $-f$ is convex.

1.2   Examples of univariate convex functions

This is a selection from [1]; see this reference for more examples.

- e^{ax}

- $-\log(x)$

- x^a (defined on $\mathbb{R}_{++}$), $a \geq 1$ or $a \leq 0$

- $-x^a$ (defined on $\mathbb{R}_{++}$), $0 \leq a \leq 1$

- $|x|^a$, $a \geq 1$

- $x \log(x)$ (defined on $\mathbb{R}_{++}$)

Can you formally verify that these functions are convex?

1.3   Strict and strong convexity

Definition 2. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is*

- Strictly convex *if $\forall x, y, x \neq y, \forall \lambda \in (0, 1)$*

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y).$$

- Strongly convex, *if $\exists \alpha > 0$ such that $f(x) - \alpha ||x||^2$ is convex.*

Lemma 1. *Strong convexity $\Rightarrow$ Strict convexity $\Rightarrow$ Convexity.*
(But the converse of neither implication is true.)

<u>Proof:</u> The fact that strict convexity implies convexity is obvious.

To see that strong convexity implies strict convexity, note that strong convexity of f implies

$$f(\lambda x + (1-\lambda)y) - \alpha||\lambda x + (1-\lambda)y||^2 \leq \lambda f(x) + (1-\lambda)f(y) - \lambda\alpha||x||^2 - (1-\lambda)\alpha||y||^2.$$

But

$$\lambda\alpha||x||^2 + (1-\lambda)\alpha||y||^2 - \alpha||\lambda x + (1-\lambda)y||^2 > 0, \ \ \forall x, y, x \neq y, \ \ \forall \lambda \in (0,1),$$

because $||x||^2$ is strictly convex (why?). The claim follows.

To see that the converse statements are not true, observe that $f(x) = x$ is convex but not strictly convex and $f(x) = x^4$ is strictly convex but not strongly convex (why?).

1.4   Examples of multivariate convex functions

- **Affine functions:** $f(x) = a^T x + b$ (for any $a \in \mathbb{R}^n, b \in \mathbb{R}$). They are convex, but not strictly convex; they are also concave:

$$\forall \lambda \in [0,1], \ f(\lambda x + (1-\lambda)y) = a^T(\lambda x + (1-\lambda)y) + b$$
$$= \lambda a^T x + (1-\lambda)a^T y + \lambda b + (1-\lambda)b$$
$$= \lambda f(x) + (1-\lambda)f(y).$$

  In fact, affine functions are the only functions that are both convex and concave.

- **Some quadratic functions:** $f(x) = x^T Q x + c^T x + d$.

  - Convex if and only if $Q \succeq 0$.
  - Strictly convex if and only if $Q \succ 0$.
  - Concave if and only if $Q \preceq 0$; strictly concave if and only if $Q \prec 0$.
  - The proofs are easy if we use the second order characterization of convexity (coming up).

- **Any norm**: Recall that a norm is any function f that satisfies:

  a. $f(\alpha x) = |\alpha|f(x), \forall \alpha \in \mathbb{R}$
  b. $f(x+y) \leq f(x) + f(y)$
  c. $f(x) \geq 0, \forall x, f(x) = 0 \Rightarrow x = 0$

<u>Proof:</u> $\forall \lambda \in [0, 1]$,

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

where the inequality follows from triangle inequality and the equality follows from the homogeneity property. (We did not even use the positivity property.) $\square$
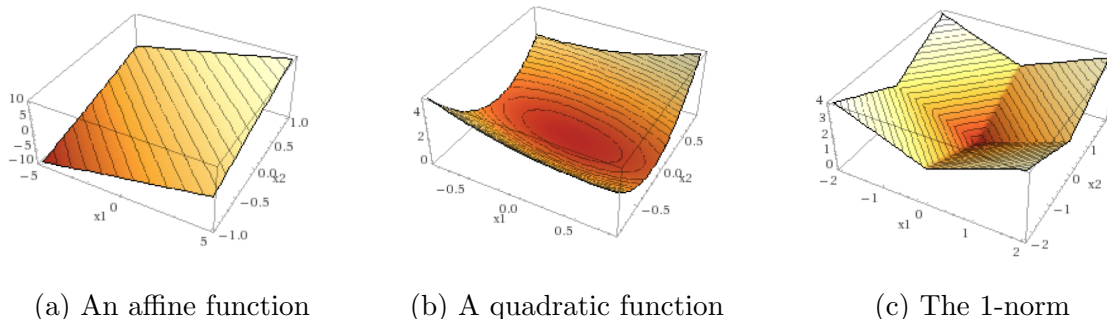

(a) An affine function   (b) A quadratic function   (c) The 1-norm

Figure 2: Examples of multivariate convex functions

1.5 Convexity = convexity along all lines

Theorem 1. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(t) = f(x + ty)$ is convex (as a univariate function) for all x in domain of f and all $y \in \mathbb{R}^n$. (The domain of g here is all t for which $x + ty$ is in the domain of f.)*

<u>Proof:</u> This is straightforward from the definition.

- The theorem simplifies many basic proofs in convex analysis but it does not usually make verification of convexity that much easier as the condition needs to hold for *all* lines (and we have infinitely many).

- Many algorithms for convex optimization iteratively minimize the function over lines. The statement above ensures that each subproblem is also a convex optimization problem.
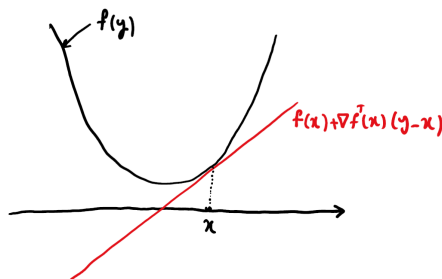
2 First and second order characterizations of convex functions

Theorem 2. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable over an open domain. Then, the following are equivalent:*
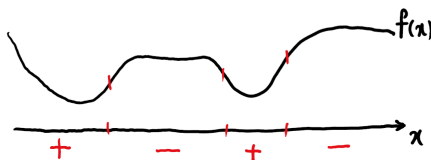
(i) f is convex.

(ii) $f(y) \geq f(x) + \nabla f(x)^T (y - x)$, for all $x, y \in dom(f)$.

(iii) $\nabla^2 f(x) \succeq 0$, for all $x \in dom(f)$.

Intepretation:
Condition (ii): The first order Taylor expansion at any point is a global under estimator of the function.

Condition (iii): The function f has nonnegative curvature everywhere. (In one dimension $f''(x) \geq 0, \forall x \in dom(f)$.)

Sign of the curvature
(sign of f'')

Proof ([2],[1]):

We prove $(i) \Leftrightarrow (ii)$ then $(ii) \Leftrightarrow (iii)$.

$(i) \Rightarrow (ii)$ If f is convex, by definition

$$f(\lambda y + (1 - \lambda)x) \leq \lambda f(y) + (1 - \lambda)f(x), \forall \lambda \in [0, 1], x, y \in dom(f).$$

After rewriting, we have

$$f(x + \lambda(y - x)) \leq f(x) + \lambda(f(y) - f(x))$$
$$\Rightarrow f(y) - f(x) \geq \frac{f(x + \lambda(y - x)) - f(x)}{\lambda}, \forall \lambda \in (0, 1]$$

As $\lambda \downarrow 0$, we get

$$f(y) - f(x) \geq \nabla f^T(x)(y - x).$$

$$(1)$$

$(ii) \Rightarrow (i)$ Suppose (1) holds $\forall x, y \in dom(f)$. Take any $x, y \in dom(f)$ and let

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

We have

$$f(x) \geq f(z) + \nabla f^T(z)(x - z) \tag{2}$$
$$f(y) \geq f(z) + \nabla f^T(z)(y - z) \tag{3}$$

Multiplying (2) by λ, (3) by $(1 - \lambda)$ and adding, we get

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(z) + \nabla f^T(z)(\lambda x + (1 - \lambda)y - z)$$
$$= f(z)$$
$$= f(\lambda x + (1 - \lambda)y). \ \square$$

$(ii) \Leftrightarrow (iii)$ We prove both of these claims first in dimension 1 and then generalize.

$(ii) \Rightarrow (iii)$ $(n = 1)$ Let $x, y \in dom(f)$, $y > x$. We have

$$f(y) \geq f(x) + f'(x)(y - x) \tag{4}$$
$$\text{and } f(x) \geq f(y) + f'(y)(x - y) \tag{5}$$

$$\Rightarrow f'(x)(y-x) \leq f(y) - f(x) \leq f'(y)(y-x)$$

using (4) then (5). Dividing LHS and RHS by $(y-x)^2$ gives

$$\frac{f'(y) - f'(x)}{y - x} \geq 0, \ \forall x, y, \ x \neq y.$$

As we let $y \to x$, we get

$$f''(x) \geq 0, \ \forall x \in dom(f).$$

$(iii) \Rightarrow (ii) \ (n = 1)$ Suppose $f''(x) \geq 0, \ \forall x \in dom(f)$. By the mean value version of Taylor's theorem we have

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2}f''(z)(y-x)^2, \text{ for some } z \in [x, y].$$
$$\Rightarrow f(y) \geq f(x) + f'(x)(y-x).$$

Now to establish $(ii) \Leftrightarrow (iii)$ in general dimension, we recall that convexity is equivalent to convexity along all lines; i.e., $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $g(\alpha) = f(x_0 + \alpha v)$ is convex $\forall x_0 \in dom(f)$ and $\forall v \in \mathbb{R}^n$. We just proved this happens iff

$$g''(\alpha) = v^T \nabla^2 f(x_0 + \alpha v)v \geq 0,$$

$\forall x_0 \in dom(f), \forall v \in \mathbb{R}^n$ and $\forall \alpha$ s.t. $x_0 + \alpha v \in dom(f)$. Hence, f is convex iff $\nabla^2 f(x) \succeq 0$ for all $x \in dom(f)$. $\square$

Corollary 1. *Consider an unconstrained optimization problem*

$$\min f(x)$$
$$s.t. \ x \in \mathbb{R}^n,$$

where f is convex and differentiable. Then, any point $\bar{x}$ that satisfies $\nabla f(\bar{x}) = 0$ is a global minimum.

Proof: From the first order characterization of convexity, we have

$$f(y) \geq f(x) + \nabla f^T(x)(y-x), \ \forall x, y$$

In particular,

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

Since $\nabla f(\bar{x}) = 0$, we get

$$f(y) \geq f(\bar{x}), \ \forall y. \ \square$$

Remarks:

- Recall that $\nabla f(x) = 0$ is always a necessary condition for local optimality in an unconstrained problem. The theorem states that for convex problems, $\nabla f(x) = 0$ is not only necessary, but also sufficient for local *and global* optimality.

- In absence of convexity, $\nabla f(x) = 0$ is not sufficient even for local optimality (e.g., think of $f(x) = x^3$ and $\bar{x} = 0$).

- Another necessary condition for (unconstrained) local optimality of a point x was $\nabla^2 f(x) \succeq 0$. Note that a convex function automatically passes this test.

3   Strict convexity

3.1   Characterization of Strict Convexity

Recall that a fuction $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if $\forall x, y, x \neq y, \forall \lambda \in (0, 1)$,

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y).$$

Like we mentioned before, if f is strictly convex, then f is convex (this is obvious from the definition) but the converse is not true (e.g., $f(x) = x, \ x \in \mathbb{R}$).

Second order sufficient condition:

$$\nabla^2 f(x) \succ 0, \ \forall x \in \Omega \Rightarrow \ f \text{ strictly convex on } \Omega.$$

The converse is not true though (why?).

First order characterization: A function f is strictly convex on $\Omega \subseteq \mathbb{R}^n$ if and only if

$$f(y) > f(x) + \nabla f^T(x)(y - x), \forall x, y \in \Omega, x \neq y.$$

- There are similar characterizations for strongly convex functions. For example, f is strongly convex if and only if there exists $m > 0$ such that

$$f(y) \geq f(x) + \nabla^T f(x)(y - x) + m||y - x||^2, \ \forall x, y \in dom(f),$$

  or if and only if there exists $m > 0$ such that

$$\nabla^2 f(x) \succeq mI, \ \forall x \in dom(f).$$

- One of the main uses of strict convexity is to ensure uniqueness of the optimal solution. We see this next.

3.2   Strict Convexity and Uniqueness of Optimal Solutions

Theorem 3. *Consider an optimization problem*

$$\min f(x)$$
$$s.t. \ x \in \Omega,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex on Ω and Ω is a convex set. Then the optimal solution (assuming it exists) must be unique.

<u>Proof:</u> Suppose there were two optimal solutions $x, y \in \mathbb{R}^n$. This means that $x, y \in \Omega$ and

$$f(x) = f(y) \leq f(z), \forall z \in \Omega. \tag{6}$$

But consider $z = \frac{x+y}{2}$. By convexity of Ω, we have $z \in \Omega$. By strict convexity, we have

$$
\begin{aligned}
f(z) = f\left(\frac{x+y}{2}\right) \\
< \frac{1}{2}f(x) + \frac{1}{2}f(y) \\
= \frac{1}{2}f(x) + \frac{1}{2}f(x) = f(x).
\end{aligned}
$$

But this contradicts (6). $\square$

<u>Exercise:</u> For each function below, determine whether it is convex, strictly convex, strongly convex or none of the above.

- $f(x) = (x_1 - 3x_2)^2$

- $f(x) = (x_1 - 3x_2)^2 + (x_1 - 2x_2)^2$

- $f(x) = (x_1 - 3x_2)^2 + (x_1 - 2x_2)^2 + x_1^3$

- $f(x) = |x| \ (x \in \mathbb{R})$

- $f(x) = ||x|| \ (x \in \mathbb{R}^n)$

3.3   Quadratic functions revisited

Let $f(x) = x^T A x + bx + c$ where A is symmetric. Then f is convex if and only if $A \succeq 0$, independently of b, c. (why?)

Consider now the unconstrained optimization problem

$$\min_x f(x) \tag{7}$$

We can establish the following claims:

- $A \not\succeq 0$ (f not convex) $\Rightarrow$ problem (7) is unbounded below.
 <u>Proof:</u> Let $\bar{x}$ be an eigenvector with a negative eigenvalue λ. Then

$$A\bar{x} = \lambda\bar{x} \Rightarrow \bar{x}^T A\bar{x} = \lambda\bar{x}^T\bar{x} < 0$$
$$f(\alpha\bar{x}) = \alpha^2\bar{x}^T A\bar{x} + \alpha b\bar{x} + c$$

 So $f(\alpha\bar{x}) \to -\infty$ when $\alpha \to \infty$.

- $A \succ 0 \Rightarrow f$ is strictly convex. There is a unique solution to (7):

$$x^* = -\frac{1}{2}A^{-1}b \ \text{(why?)}$$

- $A \succeq 0 \Rightarrow f$ is convex. If A is positive semidefinite but not positive definite, the problem can be unbounded (e.g., take $A = 0$, $b \neq 0$), or it can be bounded below and have infinitely many solutions (e.g., $f(x) = (x_1 - x_2)^2$). One can distinguish between the two cases by testing whether b lies in the range of A. Furthermore, it is impossible for the problem to have a unique solution if A has a zero eigenvalue (why?).
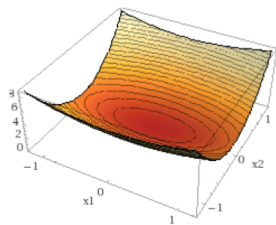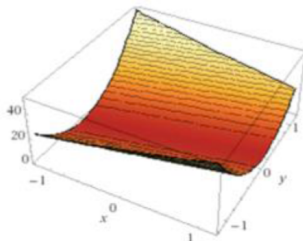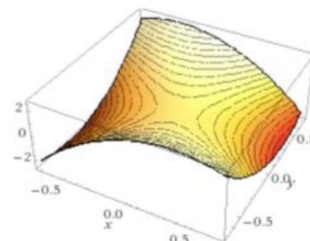
$$A \succ 0 \qquad\qquad A \succeq 0 \qquad\qquad A \nsucceq 0$$
$$(\text{but } A \nsucc 0)$$

Figure 3: An illustration of the different possibilities for unconstrained quadratic minimization

3.3.1 Least squares revisited

Recall the least squares problem

$$\min_{x} ||Ax - b||^2.$$

Under the assumption that the columns of A are linearly independent,

$$x = (A^T A)^{-1} A^T b$$

is the unique global solution because the objective function is strictly convex (why?).

4 Optimality conditions for convex optimization

Theorem 4. *Consider an optimization problem*

$$min. \ f(x)$$
$$s.t. \ x \in \Omega,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable and Ω is convex. Then a point x is optimal if and only if $x \in \Omega$ and

$$\nabla f(x)^T (y - x) \geq 0, \forall y \in \Omega.$$

Remarks:

11

- What does this condition mean?

  - If you move from x towards any feasible y, you will increase f locally.

  - The vector $-\nabla f(x)$ (assuming it is nonzero) serves as a hyperplane that "supports" the feasible set Ω at x. (See figure below.)
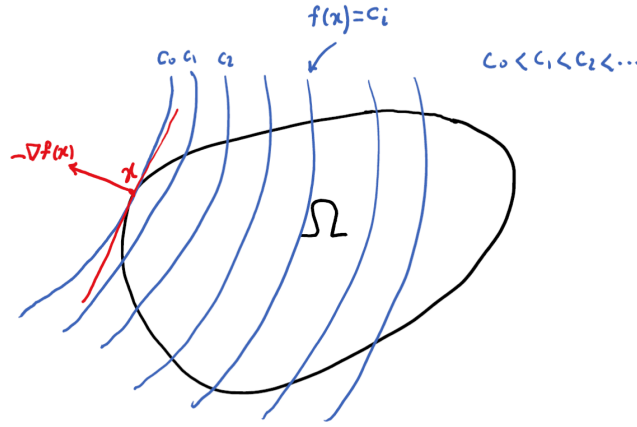

Figure 4: An illustration of the optimality condition for convex optimization

- The necessity of the condition holds independent of convexity of f. Convexity is used in establishing sufficiency.

- If $\Omega = \mathbb{R}^n$, the condition above reduces to our first order unconstrained optimality condition $\nabla f(x) = 0$ (why?).

- Similarly, if x is in the interior of Ω and is optimal, we must have $\nabla f(x) = 0$. (Take $y = x - \alpha \nabla f(x)$ for α small enough.)

Proof:
(Sufficiency) Suppose $x \in \Omega$ satisfies

$$\nabla f(x)^T(y - x) \geq 0, \forall y \in \Omega. \tag{8}$$

By the first order characterization of convexity, we have:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x), \forall y \in \Omega. \tag{9}$$

12

Then

$$(8) + (9) \Rightarrow f(y) \geq f(x), \ \forall y \in \Omega$$
$$\Rightarrow x \ \text{is optimal.}$$

(Necessity) Suppose x is optimal but for some $y \in \Omega$ we had

$$\nabla f^T(x)(y - x) < 0.$$

Consider $g(\alpha) := f(x + (\alpha(y - x)))$. Because Ω is convex, $\forall \alpha \in [0, 1]$, $x + \alpha(y - x) \in \Omega$. Observe that

$$g'(\alpha) = (y - x)^T \nabla f(x + \alpha(y - x)).$$
$$\Rightarrow g'(0) = (y - x)^T \nabla f(x) < 0.$$

This implies that

$$\exists \delta > 0, \ \text{s.t.} \ g(\alpha) < g(0), \ \forall \alpha \in (0, \delta)$$
$$\Rightarrow f(x + \alpha(y - x)) < f(x), \ \forall \alpha \in (0, \delta).$$

But this contradicts the optimality of x. $\square$

Here's a special case of this theorem that comes up often.

Theorem 5. *Consider the optimization problem*

$$\min f(x) \tag{10}$$
$$s.t. \ Ax = b,$$

where f is a convex function and $A \in \mathbb{R}^{m \times n}$. A point $x \in \mathbb{R}^n$ is optimal to (10) *if and only if it is feasible and $\exists \mu \in \mathbb{R}^m$ s.t.*

$$\nabla f(x) = A^T \mu.$$

<u>Proof:</u> Since this is a convex problem, our optimality condition tells us that a feasible x is optimal iff

$$\nabla f^T(x)(y - x) \geq 0, \ \forall y \ \text{s.t.} \ Ay = b.$$

13

Any y with $Ay = b$ can be written as $y = x + v$, where v is a point in the nullspace of A; i.e., $Av = 0$. Therefore, a feasible x is optimal if and only if $\nabla f^T(x)v \geq 0, \forall v$ s.t. $Av = 0$. For any v, since $Av = 0$ implies $A(-v) = 0$, we see that $\nabla f^T(x)v \leq 0$. Hence the optimality condition reads $\nabla f^T(x)v = 0$, $\forall v$ s.t. $Av = 0$.

This means that $\nabla f(x)$ is in the orthogonal complement of the nullspace of A which we know from linear algebra equals the row space of A (or equivalently the column space of A^T). Hence $\exists \mu \in \mathbb{R}^m$ s.t. $\nabla f(x) = A^T \mu$. $\square$

Notes

Further reading for this lecture can include Chapter 3 of [1].

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, http://stanford.edu/ boyd/cvxbook/, 2004.

[2] A.L. Tits. *Lecture notes on optimal control*. University of Maryland, 2013.