Moving towards nonlinear optimization

CE 377K

April 9, 2015

Review

Simplex tableau

Big M method

Geometric interpretation of sensitivity analysis

OUTLINE

- Formulas for sensitivity analysis (to changes in $\mathbf{c}$ or $\mathbf{b}$)
- Moving towards nonlinear optimization
- Convex functions and sets

Change to the requirement vector

Assume that the j-th component of $\mathbf{b}$ changes from b_j to $b'_j = b_j + \delta$.

If the previous optimal basis is feasible, then it is still optimal.

Why? The reduced costs are $\overline{c}^k = c^k - \mathbf{c^B B^{-1} A^k}$.

These do not depend on $\mathbf{b}$, so if the reduced costs were nonnegative with $\mathbf{b}$, they will also be nonnegative under $\mathbf{b}'$.

However, if δ is large enough, the old basis is no longer feasible (and thus not optimal).

We have a formula for the new basic variables: $\mathbf{x^B} = \mathbf{B}^{-1}\mathbf{b}$ so

$$(\mathbf{x^B})' = \mathbf{B}^{-1}\mathbf{b}' = \mathbf{x^B} + \delta \mathbf{B}^{-1}\mathbf{e_j}$$

where $\mathbf{e_j}$ is a column vector of all zeroes (except for a value of 1 in the j-th row)

Since $(\mathbf{x^B})' = \mathbf{B}^{-1}\mathbf{b}' = \mathbf{x^B} + \delta\mathbf{B}^{-1}\mathbf{e_j}$, we know that the old optimal solution $\mathbf{x^B}$ is perturbed by $\delta\mathbf{B}^{-1}\mathbf{e_j}$.

Remember the bottom right portion of the tableau represents $\mathbf{B}^{-1}\mathbf{A}$.

So, if one of the columns of the constraint matrix is e_j, we just need to look at that column of the final tableau.

If we were using the big-M method, *there will always be such a column.*

So, if $(\mathbf{x^B})' = \mathbf{x^B} + \delta \mathbf{B^{-1}} \mathbf{e_j} \geq \mathbf{0}$, then the old basis is still feasible.

This vector equation breaks down into m scalar equations of the form

$$x_i + \delta \mathcal{T}_{ik} \geq 0$$

This means that if $\mathcal{T}_{ik} > 0$, we need $\delta \geq -x_i/\mathcal{T}_{ik}$ for the current basis to remain feasible.

If $\mathcal{T}_{ik} < 0$, we need $\delta \leq -x_i/\mathcal{T}_{ik}$.

Since *each* of the m constraints imposes such a restriction on δ, we have that the current basis remains optimal iff

$$\max_{i:\mathcal{T}_{ik}>0} -\frac{x_i}{\mathcal{T}_{ik}} \leq \delta \leq \min_{i:\mathcal{T}_{ik}<0} -\frac{x_i}{\mathcal{T}_{ik}}$$

Example

The problem

$$\begin{aligned} \min_{x_1,x_2,x_3,x_4} \quad & -5x_1 - x_2 + 12x_3 \\ \text{s.t.} \quad & 3x_1 + 2x_2 + x_3 && = 10 \\ & 5x_1 + 3x_2 + x_4 && = 16 \\ & x_1, x_2, x_3, x_4 && \geq 0 \end{aligned}$$

has optimal tableau

12	0	0	2	7
2	1	0	−3	2
2	0	1	5	−3

What is the optimal solution? How much can the right-hand side of the first constraint change without affecting the optimal basis?

12	0	0	2	7
2	1	0	-3	2
2	0	1	5	-3

Since the third column of **A** was $\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, the third column of the tableau gives us the information we need.

If $\delta \in [-2/5, 2/3]$ the current basis remains feasible.

Change to the cost vector **c**

Now, assume that the requirement vector **b** is the same, but one of the cost coefficients is changed from c_i to $c_i' = c_i + \delta$

Here, there are no concerns about feasibility since none of the constraints have changed.

However, the old solution may no longer be optimal.

The reduced costs are the key to checking optimality.

As before, assume that we have the optimal solution (and optimal tableau) for the original optimization problem.

The old basis remains optimal as long as all of the reduced costs remain nonnegative.

So, we need

$$\overline{\mathbf{c}}' = \mathbf{c}' - \mathbf{c^{B}}'\mathbf{B^{-1}A} \geq 0$$

which is equivalent to n scalar equations of the form $c_k' - \mathbf{c^{B}}'\mathbf{B^{-1}A^{k}} \geq 0$

The easy case: assume that x_i is nonbasic. Then $\mathbf{c^{B'}} = \mathbf{c^B}$, and all of the equations stay the same except for the one involving x_i.

This one is simply $\mathbf{c^B B^{-1} A^i} \leq c^i + \delta$ or $\delta \geq -\overline{c}_i = -\mathcal{T}_{0i}$ which is easily obtained from the final tableau.

Example

The problem

$$\begin{aligned}
\min_{x_1,x_2,x_3,x_4} \quad & -5x_1 - x_2 + 12x_3 & \\
\text{s.t.} \quad & 3x_1 + 2x_2 + x_3 & = 10 \\
& 5x_1 + 3x_2 + x_4 & = 16 \\
& x_1, x_2, x_3, x_4 & \geq 0
\end{aligned}$$

has optimal tableau

$$\begin{array}{|c|cccc|}
\hline
12 & 0 & 0 & 2 & 7 \\
\hline
2 & 1 & 0 & -3 & 2 \\
2 & 0 & 1 & 5 & -3 \\
\hline
\end{array}$$

How much can x_3 change for the current solution to remain optimal?

12	0	0	2	7
2	1	0	-3	2
2	0	1	5	-3

Since x_3 is nonbasic, we simply need $\delta \geq -\overline{c}_3 = -2$. For any such δ the current solution remains optimal.

The slightly harder case is if x_i is basic (say it is the ℓ-th basic variable). In this case $\mathbf{c^B}$ changes as well as $\mathbf{c}$.

So, the new reduced cost equation is $\mathbf{c^B B^{-1} A} + \delta \mathbf{e}_\ell \mathbf{B^{-1} A} \leq \mathbf{c} + \mathbf{e_i}$ which simplifies to

$$\delta[\mathbf{B^{-1} A^j}]_\ell \leq c_j - \mathbf{c^B B^{-1} A} = \overline{c}_j$$

or, from the tableau,

$$\mathcal{T}_{\ell i}\delta \leq \overline{c}_j$$

for all j.

As we did for changes in the requirement vector, this last formula simplifies to

$$\max_{j \neq i : \mathcal{T}_{\ell j} < 0} \left\{ \frac{\overline{c}_j}{\mathcal{T}_{\ell j}} \right\} \leq \delta \leq \min_{j \neq i : \mathcal{T}_{\ell j} > 0} \left\{ \frac{\overline{c}_j}{\mathcal{T}_{\ell j}} \right\}$$

Example

The problem

$$
\begin{aligned}
\min_{x_1,x_2,x_3,x_4} \quad & -5x_1 - x_2 + 12x_3 \\
\text{s.t.} \quad & 3x_1 + 2x_2 + x_3 && = 10 \\
& 5x_1 + 3x_2 + x_4 && = 16 \\
& x_1, x_2, x_3, x_4 && \geq 0
\end{aligned}
$$

has optimal tableau

12	0	0	2	7
2	1	0	-3	2
2	0	1	5	-3

How much can x_1 change for the current solution to remain optimal?

12	0	0	2	7
2	1	0	-3	2
2	0	1	5	-3

Since x_1 is basic, we need to compare the ratios of reduced costs for each nonbasic variable to the tableau entry: $\delta \in [-2/3, 7/2]$ for the current basis to remain optimal.

NONLINEAR OPTIMIZATION

Many optimization problems cannot be put into the form of a linear program.

$$
\begin{aligned}
\min_{\mathbf{n}} \quad & D(\mathbf{n}) = \sum_{r \in R} \frac{d_r T_r}{2 n_r} & \\
\text{s.t.} \quad & n_r \geq L_r & \forall r \in R \\
& n_r \leq U_r & \forall r \in R
\end{aligned}
$$

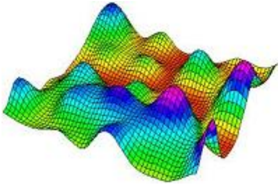
The next few weeks discuss techniques that can be used when the objective and/or constraints are nonlinear.

The standard form for a nonlinear optimization problem is

$$
\begin{array}{lrl}
\min\limits_{\mathbf{x}} & f(\mathbf{x}) & \\
\text{s.t.} & g_1(\mathbf{x}) & \leq 0 \\
& \vdots & \\
& g_l(\mathbf{x}) & \leq 0 \\
& h_1(\mathbf{x}) & = 0 \\
& \vdots & \\
& h_m(\mathbf{x}) & = 0
\end{array}
$$

The objective function is to be minimized; all other constraints are of the form $\leq$ or $=$.

The *general* nonlinear optimization problem (where f, g, and h can be any functions whatever) is extremely difficult and probably impossible.



However, if the objective and constraints are “nice” functions, there are efficient algorithms for finding the global minimum.

At the start of this class, we saw some of these conditions (continuity, differentiability, unimodality, coercivity, boundedness, etc.)

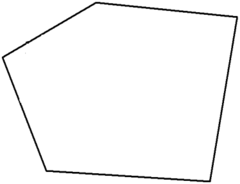
For nonlinear optimization problems the most important condition in practice is *convexity*.

There are actually *two* definitions of convexity, one applies to sets and the other applies to functions.

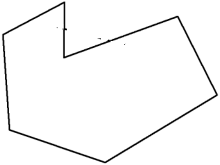
We will see that finding the global minimum of a convex function over a convex feasible set is achievable.

CONVEX SET

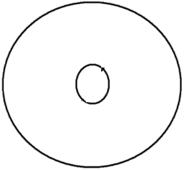
Intuitively, a convex set does not have any “holes” or “bites” in it.



Convex

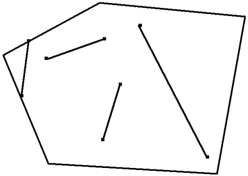


Not convex

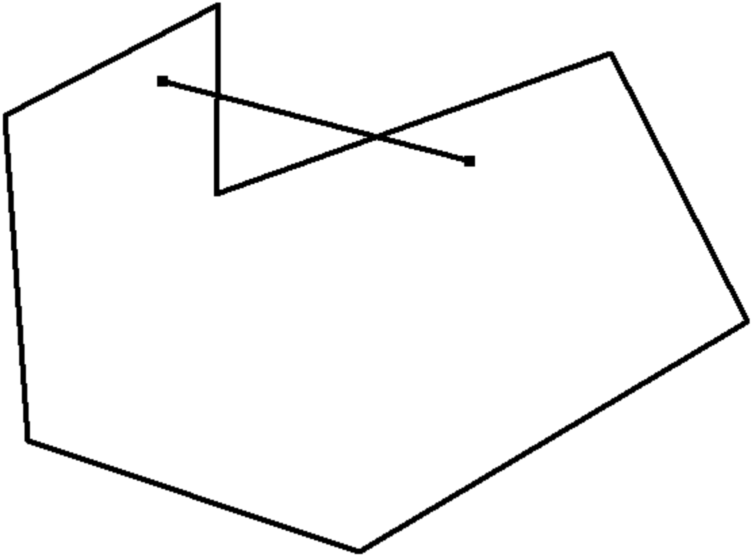


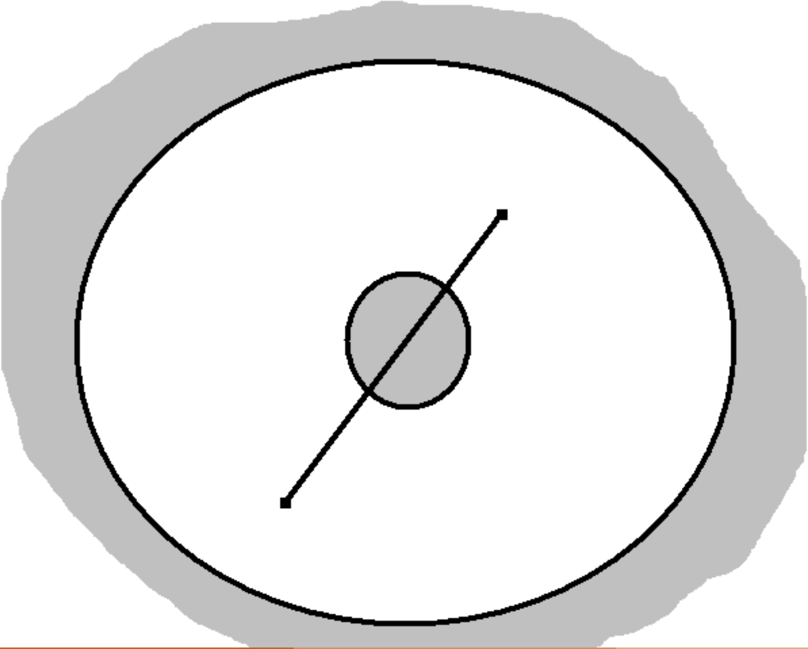
Not convex

The more precise definition is that **for any two points in the set, the straight line connecting those two points also lies in the set.**



Specifically, the set X is convex if, for any $x_1 \in X$, $x_2 \in X$, and $\lambda \in [0, 1]$, the point $\lambda x_1 + (1 - \lambda)x_2 \in X$. (Such a point is a **convex combination** of x_1 and x_2.





Example

The one-dimensional set $X = \{x : x \geq 0\}$ is convex.

Pick any $x_1 \geq 0$, $x_2 \geq 0$, and $\lambda \in [0, 1]$.

Because all three of these are nonnegative, so is $\lambda x_1 + (1 - \lambda)x_2$.

Therefore the set is convex.

Example

The plane $X = \{(x, y, z) : 3x + 4y - 3z = 1\}$ is convex.

Pick any (x_1, y_1, z_1) and (x_2, y_2, z_2) in X, and any $\lambda \in [0, 1]$.

Then the convex combination is $(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2, \lambda z_1 + (1 - \lambda)z_2)$. Does this satisfy the conditions to be part of X?

We know $3x_1 + 4y_1 - 3z_1 = 1$ and $3x_2 + 4y_2 - 3z_2 = 1$.

Therefore $\lambda(3x_1 + 4y_1 - 3z_1) = \lambda$ and $(1-\lambda)(3x_2 + 4y_2 - 3z_2) = 1 - \lambda$.

Adding these shows that the convex combination $(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2, \lambda z_1 + (1-\lambda)z_2)$ also satisfies the equation of the plane, so it is convex.

Example

Is the region $X = \{(x, y) : x^2 + y^2 \geq 1\}$ convex?

The points $(1, 0)$ and $(-1, 0)$ are in X. Pick $\lambda = 1/2$.

The resulting point $(0, 0)$ is *not* in X, so X is not convex.

To show that a set is convex, you have to show that *every* convex combination of *every* two points in the set lie within the set. To show that a set is not convex, you only need one case where that is false.

CONVEX FUNCTIONS

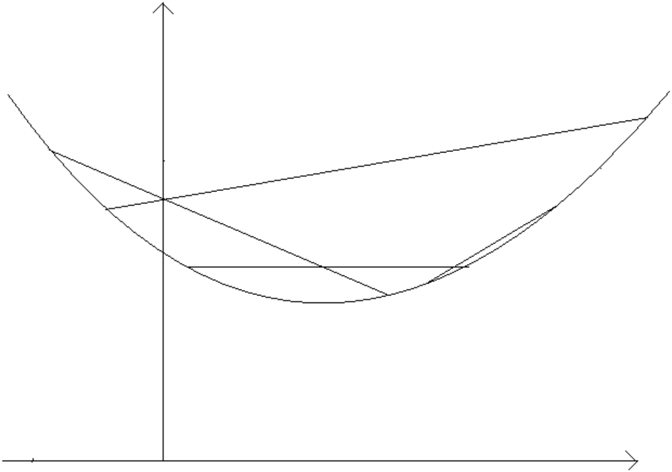
Function convexity is a bit different than set convexity.

We have already seen one definition of convexity early in the class (a one-dimensional, twice-differentiable function is convex if $f''(x) \geq 0$ everywhere.)

We will now generalize this definition to higher-dimension functions and to functions which are not twice differentiable.

Throughout this discussion, assume that the function's domain is a convex set.

Intuitively, a convex set lies below its secant lines.



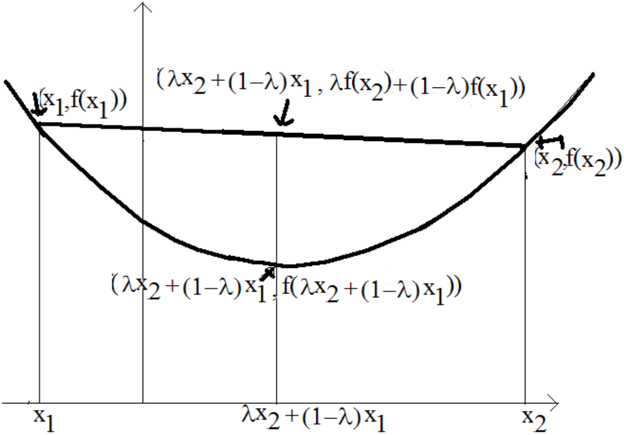
The mathematical way to express this is:

A function $f : X \to \mathbb{R}$ is *convex* if, for every $x_1, x_2 \in X$ and every $\lambda \in (0, 1)$,

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2) \tag{1}$$

Such a function is *strictly convex* if the $\leq$ can be replaced by $<$

Compare this definition with the figure:



Example

Is the function $f(x) = |x|$ convex? Is it strictly convex?

Pick any x_1, x_2, and $\lambda \in (0, 1)$.

$$f((1-\lambda)x_1 + \lambda x_2) = |(1-\lambda)x_1 + \lambda x_2|$$

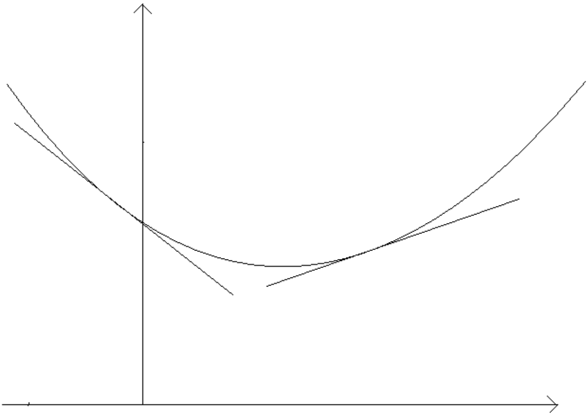
$\leq |(1-\lambda)x_1| + |\lambda x_2|$ by the triangle inequality

$= (1-\lambda)|x_1| + \lambda|x_2|$

$= (1-\lambda)f(x_1) + \lambda f(x_2).$

This definition can be unwieldy to work with, so there are alternative characterizations.

If the function is differentiable, convexity can be characterized in terms of a function's *tangent* lines.



The function f is convex if it lies above all of its tangents.

Mathematically, if f is differentiable on its domain, then f is convex if and only if

$$f(x_2) \geq f(x_1) + f'(x_1)(x_2 - x_1)$$

for all $x_1, x_2 \in X$.

Example

Is x^2 convex?

Pick any x_1, x_2. Since $f'(x_1) = 2x_1$, we need to show that

$$x_2^2 \geq x_1^2 + 2x_1(x_2 - x_1)$$

This is equivalent to or

$$(x_1 - x_2)^2 \geq 0$$

which is always true, so f is convex.

If f is twice differentiable on its domain, then f is convex if and only if $f''(x) \geq 0$ everywhere.

Example: x^2 is convex because $f''(x) = 2 \geq 0$.

This is the definition we used earlier in the class.

When f is a function of multiple variables, the convexity conditions involving first and second derivatives must change.

The analogue of the first derivative is the *gradient* vector
$\nabla f = \begin{bmatrix} \partial f/\partial x_1 & \partial f/\partial x_2 & \cdots & \partial f/\partial x_n \end{bmatrix}^T$

The analogue of the second derivative is the *Hessian* matrix

$$Hf = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 & \cdots & \partial^2 f/\partial x_1 x_n \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 & \cdots & \partial^2 f/\partial x_2 x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2 f/\partial x_n \partial x_1 & \partial^2 f/\partial x_n \partial x_2 & \cdots & \partial^2 f/\partial x_n^2 \end{bmatrix}$$

For twice-differentiable multidimensional functions, f is convex if any of these equivalent conditions are satisfied:

1. For all x_1 and x_2 in X,

$$f(\lambda x_2 + (1-\lambda)x_1) \leq \lambda f(x_2) + (1-\lambda) f(x_1)$$

2. For all x_1 and x_2 in X,

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1)$$

3. For all x in X, $H(x)$ is positive semidefinite (that is, $y^T H(x) y \geq 0$ for all vectors y).

These conditions can be tedious to check. In this class I will not ask you to apply these definitions directly to multidimensional functions.

However, there are some facts which we can use (even in higher dimensions):

- Any linear function is convex.
- A nonnegative multiple of a convex function is convex.
- The sum of convex functions is convex.
- The composition of convex functions is convex.