10.007: Modeling the Systems World

Week 03 - Cohort 1

Unconstrained Optimization with Convex Functions

Term 3, 2017

Singapore University of
Technology & Design

Agenda

- Necessary and sufficient conditions for local optimality.
- Relationship between convexity and optimality conditions.
- Eigenvalues of the Hessian.

Unconstrained optimization

Suppose we have the optimization problem:

$$(\max \text{ or}) \min\{f(x) : x \in \mathbb{R}^n\} \qquad \text{(P)}.$$

This is an unconstrained optimization problem: the feasible region is the entire space $\mathbb{R}^n$.

The feasible region is an unbounded set: it is not clear if a minimum/maximum exists.

For example, $f(x) = e^x$ does not have a maximum in $\mathbb{R}$. It does not even have a minimum, although it is bounded below by 0.

Necessary conditions for local minima

Let x^* be an unconstrained local minimum of $f : \mathbb{R}^n \to \mathbb{R}$.

- First-order condition:

  If f is continuously differentiable, then:

$$\nabla f(x^*) = \vec{0}.$$

- Second-order condition:

  If f is twice continuously differentiable, then:

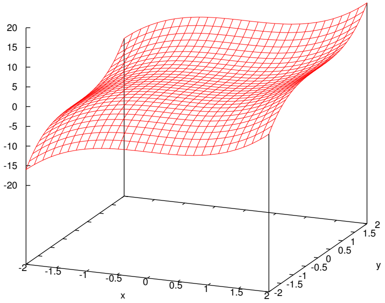
$$H_f(x^*) \text{ is psd.}$$

Necessary $\neq$ sufficient

Example from the lecture: $f(x,y) = x^3 + y^3$.

$$\nabla f(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and $H_f(0,0)$ is psd, so the point $(0,0)$ satisfies both necessary conditions. But the function has no minimum over $\mathbb{R}^2$.



A sufficient condition

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. Suppose $x^* \in \mathbb{R}^n$ satisfies:

$$\nabla f(x^*) = \vec{0}, \qquad H_f(x^*) \text{ is pd.}$$

Then x^* is a strict unconstrained local minimum of f.

Single-variable function $f(x)$:

- Find points with $f'(x) = 0$.
- For all such points: check $f''(x) > 0$.

Two-variable function $f(x, y)$:

- Find points with $\nabla f(x, y) = \vec{0}$.
- For all such points: check $H_f(x, y)$ is pd.

Necessary and Sufficient Conditions for Minimizing Convex Functions

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, and assume that f is continuously differentiable at point x^*. Then x^* is a global minimum of f if and only if $\nabla f(x^*) = \vec{0}$.

This means that we have a very simple procedure to find the global minimum of a convex and differentiable function.

Optimality conditions for convex functions

Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a twice continuously differentiable convex function. We want to solve: $\min_{x \in \mathbb{R}^n} f(x)$.

Because the function is convex, the approach becomes simpler:

1. Find all x^* such that $\nabla f(x^*) = \vec{0}$. These are global minima.
2. (Optional unless requested) Check if $H_f(x^*)$ is pd. If yes, x^* is a strict global minimum.

If we are not interested in checking if global minima are strict, solving Step 1 is sufficient. Step 1 is often very hard even for the convex case.

Activity 1 (15 mins)

Find all the global optima of the following optimization problems, or show that no global optimum exists.

a. $\min_{(x_1,x_2)\in\mathbb{R}^2} x_1 x_2$.
b. $\min_{(x_1,x_2,x_3)\in\mathbb{R}^3} e^{x_1+x_2+x_3} + e^{-x_1-x_2-x_3}$.

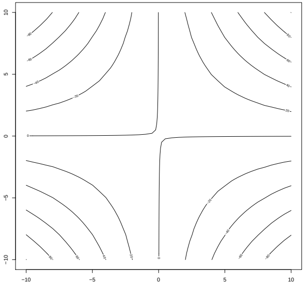
Hint: use first and second-order optimality conditions, and/or convexity.

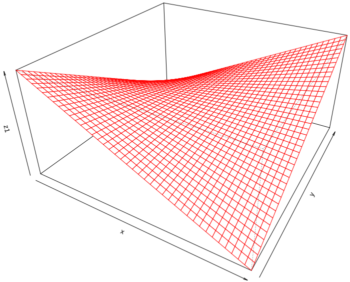
Activity 1 (solutions)

a. min $x_1 x_2$.

$$\nabla f(x_1, x_2) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad H_f(0,0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Hessian is indefinite everywhere on the domain (check eigenvalues). So there is no minimum. Notice that $x_1 x_2$ is not convex.





Activity 1 (solutions)

b. $\min e^{x_1+x_2+x_3} + e^{-x_1-x_2-x_3}$.

We observe that $f(x_1, x_2, x_3)$ is a convex function (from operations that preserve convexity).

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} e^{x_1+x_2+x_3} - e^{-x_1-x_2-x_3} \\ e^{x_1+x_2+x_3} - e^{-x_1-x_2-x_3} \\ e^{x_1+x_2+x_3} - e^{-x_1-x_2-x_3} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

First-order conditions give us $x_1 + x_2 + x_3 = 0$.

By convexity, $x^* = (x_1^*, x_2^*, x_3^*)$ with $x_1^* + x_2^* + x_3^* = 0$ is a global minimum.

Review: Taylor series at a critical point

Suppose we have $f : \mathbb{R}^n \to \mathbb{R}$ and a critical point: $\nabla f(\bar{x}) = \vec{0}$. Then:

$$f(x) \approx \hat{\hat{f}}_{\bar{x}}(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T H_f(\bar{x})(x - \bar{x})$$
$$= f(\bar{x}) + \frac{1}{2}(x - \bar{x})^T H_f(\bar{x})(x - \bar{x}).$$

For simplicity, we assume $\bar{x} = \vec{0}$. We can always translate the origin to satisfy this condition. We obtain:

$$f(x) - f(\vec{0}) \approx \frac{1}{2} x^T H_f(\vec{0}) x.$$

The difference in function value with respect to $\bar{x} = \vec{0}$ depends on the Hessian (up to second order). Remembering this will be useful in the next activity.

Activity 2 (30 minutes)

Let $f(x, y) = (x-2)(y-2)$.

a. Find a critical point of f, call it (x^*, y^*), and calculate the Hessian at (x^*, y^*) with its eigen-decomposition.
b. Discuss what happens to the function value if we move by a small step of length $\epsilon > 0$ starting from (x^*, y^*) along the following vectors:
   0.1 $(1, 0)$ and $(-1, 0)$.
   0.2 $(0, 1)$ and $(0, -1)$.
   0.3 $(1, 1)$ and $(-1, -1)$.
   0.4 $(1, -1)$ and $(-1, 1)$.

   Does the function increase or decrease?
c. Compare your findings with the directions of the eigenvectors. Try to relate to the sufficient conditions for optimality.
d. Repeat this analysis for the functions $f(x, y) = x^2 + y^2 - 2xy$ and $f(x, y) = x^2 + y^3$.

Activity 2 (solutions)

a. $f(x,y) = (x-2)(y-2)$, $\nabla f(x,y) = \begin{pmatrix} y-2 \\ x-2 \end{pmatrix}$, $H_f(x,y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

There is a critical point at $x = 2, y = 2$.

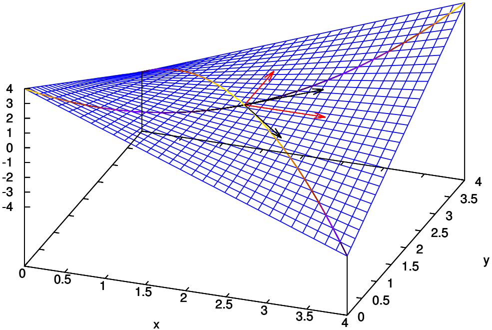
Eigenvalues and corresponding eigenvectors:

$$\lambda_1 = 1, v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \lambda_2 = -1, v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Activity 2 (solutions)

b. The function is constant along the directions $(1, 0), (-1, 0), (0, 1)$, and $(0, -1)$, so these directions do not help us understand whether $(2, 2)$ is a local optimum.

The other two directions are the same as the eigenvectors. f increases along the eigenvector with positive eigenvalue, decreases along the eigenvector with negative eigenvalue.



Activity 2 (solutions)

c. We have seen before that when the function is translated to have $\bar{x} = 0$:

$$\hat{\hat{f}}_{\bar{x}=\vec{0}}(x) - f(\vec{0}) = \frac{1}{2} x^T H_f(\vec{0}) x.$$

When x is an eigenvector of $H_f(\vec{0})$ with corresponding eigenvalue λ:

$$\hat{\hat{f}}_{\bar{x}=\vec{0}}(x) - f(\vec{0}) = \frac{1}{2} \lambda \|x\|^2.$$

So the eigenvectors corresponding to non-zero eigenvalues give information about the directions towards which the function increases or decreases. If all eigenvalues are positive, the function bends upwards everywhere and we have a local minimum – the sufficient condition for optimality.

When the eigenvalue is 0, the conclusion is not clear.

Activity 2 (solutions)

d1. $f(x, y) = x^2 + y^2 - 2xy$, $\nabla f = \begin{pmatrix} 2x - 2y \\ 2y - 2x \end{pmatrix}$, $H_f = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$.

There is a critical point at $x = 0, y = 0$.

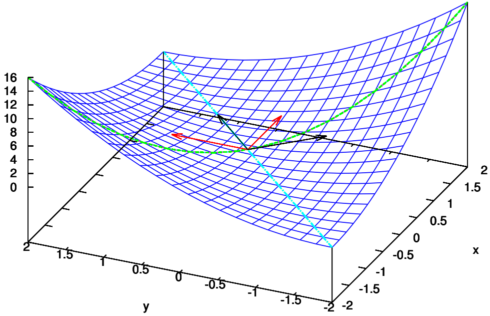
Eigenvalues and corresponding eigenvectors:

$$\lambda_1 = 0, v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \lambda_2 = 4, v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Activity 2 (solutions)

d1. Along the canonical basis $(1, 0), (0, 1)$ the function increases, but this does not show that we have a minimum.

One eigenvalue is positive and one is zero: f increases along the first eigenvector, but it is unclear what happens along the second. In this case, it stays constant: we have a minimum.



Activity 2 (solutions)

d2. $f(x,y) = x^2 + y^3, \qquad \nabla f = \begin{pmatrix} 2x \\ 3y^2 \end{pmatrix}, \qquad H_f = \begin{pmatrix} 2 & 0 \\ 0 & 6y \end{pmatrix}.$

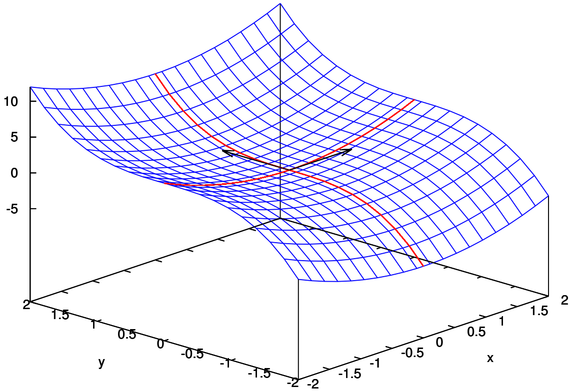
There is a critical point at $x = 0, y = 0$.

Eigenvalues and corresponding eigenvectors:

$$\lambda_1 = 2, v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \lambda_2 = 0, v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Activity 2 (solutions)

d2. One eigenvalue is positive and one is zero: f increases along the first eigenvector, but it is unclear what happens along the second. In this case, f may decrease along the direction of second eigenvector and we have a saddle point.



Activity 3 (20 minutes)

Consider the function:

$$f(x, y) = ax^2 + by^2 - cx$$

with these values of the coefficients:

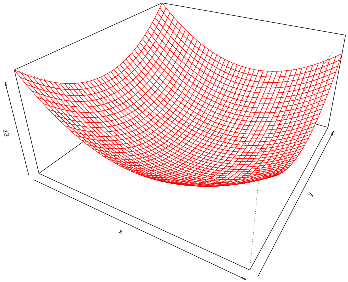
a. $a = b = 1$ and $c = 2$.
b. $a = 1, b = -1$ and $c = 2$.
c. $a = 1, b = 0$ and $c = 2$.

Which of these functions are convex? For each function:

- Compute the gradient as a function of x, y.
- Find the local minima and identify the global minima if any.

Activity 3 (solutions)

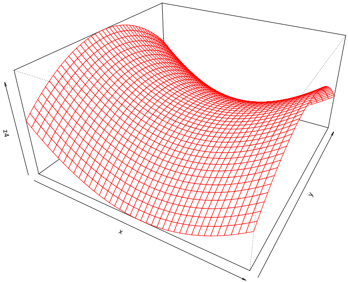
a. $f(x, y) = x^2 + y^2 - 2x$.



This is a convex function (sum of convex functions) with gradient: $\nabla f(x, y) = (2x - 2, 2y)$. The minimum is $(x_0, y_0) = (1, 0)$.

Activity 3 (solutions)

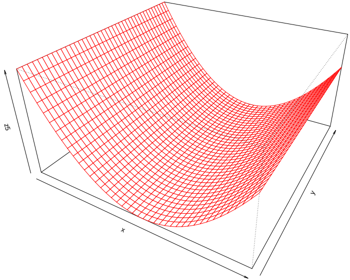
b. $f(x, y) = x^2 - y^2 - 2x$.



This is not a convex function, with gradient: $\nabla f(x, y) = (2x - 2, -2y)$. There is no minimum in $\mathbb{R}^2$: the gradient is zero at $(1, 0)$, but it is a saddle point (function increases in some directions, decreases in others).

Activity 3 (solutions)

c. $f(x, y) = x^2 - 2x$.



This is a convex function (sum of convex functions) with gradient: $\nabla f(x, y) = (2x - 2, 0)$. The minimum is $(1, y)$ for any $y \in \mathbb{R}$.

Wrap-up

When we have a critical point, the eigenvalues of the Hessian can help us understand the local behavior of the function.

If all the eigenvalues are positive, we have a minimum because the function increases in all directions. Along the eigenvectors, the function bends upwards.

If some eigenvalues are zero, the second-order approximation does not give enough information: the error term of the Taylor series becomes important.

If some eigenvalues are positive and some negative, certainly we have a saddle point.

Wrap-up

We now have both necessary and sufficient conditions for local optima of an unconstrained problem. The sufficient conditions are more restrictive.

We can use the necessary conditions to find candidate local optima. We can use the sufficient conditions to find guaranteed local optima.

The conditions studied today can be used to develop general solution methods for these problems. They can also be used to solve small problems by hand.

When the problem size is large, we can find points that satisfy these conditions by specialized algorithms. We will learn a popular method, the steepest descent algorithm, in the next cohort.

Go global!

Solving an optimization problem typically requires finding a global optimum.

However, it is very difficult to write algorithms that find global optima. In most cases, we can only guarantee local optimality.

There are problems for which global optimality follows from local optimality: these problems are considered "easy". This is the "ideal" situation.

Summary

- Necessary and sufficient conditions for local optima.
- Necessary conditions are also sufficient for convex functions.
- Local optima are also global optima for convex functions.
- The Hessian is related to the curvature of the function and important in characterizing critical points.