Unconstrained optimization

Source: Ahmadi, ORF 363 slides.

Optimization problems – basic notations

Unconstrained optimization

First order condition for optimality

Second order conditions for optimality

General form of optimization

$$\min \; f(x)$$

subject to $x \in \Omega$

General form of optimization

$$\min f(x)$$

subject to $x \in \Omega$

where

$$x \in \mathbb{R}^n, \quad f : \mathbb{R}^n \to \mathbb{R} \quad \Omega \subseteq \mathbb{R}^n$$

General form of optimization

$$\min \; f(x)$$

subject to $x \in \Omega$

Optimal solution:

$$x^* = \operatorname{argmin} f(x) \quad \text{s.t.} \quad x \in \Omega$$

General form of optimization

x^* minimizes f over Ω:

$$f(x) \geq f(x^*), \quad \forall x \in \Omega$$

General form of optimization

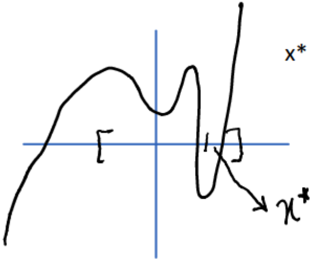
x^* minimizes f over Ω:

$$f(x) \geq f(x^*), \quad \forall x \in \Omega$$

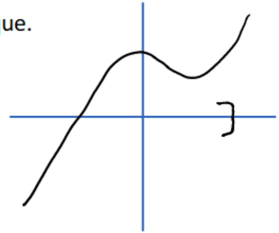
Optimal value:

$$f^* = f(x^*) \qquad \text{(if } x^* \text{ exists)}$$

Examples



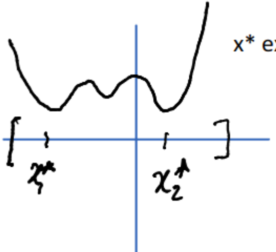
x* exists and is unique.



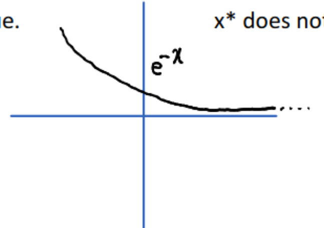
x* does not exist.

$f^* = -\infty$

Problem is "unbounded."



x* exists, but not unique.



x* does not exist.

$f^* = 0$

Maximization

What if we want to maximize an objective function instead?

- Just multiply f by a minus sign:

$$\max f(x) = -\min -f(x)$$

Optimization problems – basic notations

Unconstrained optimization

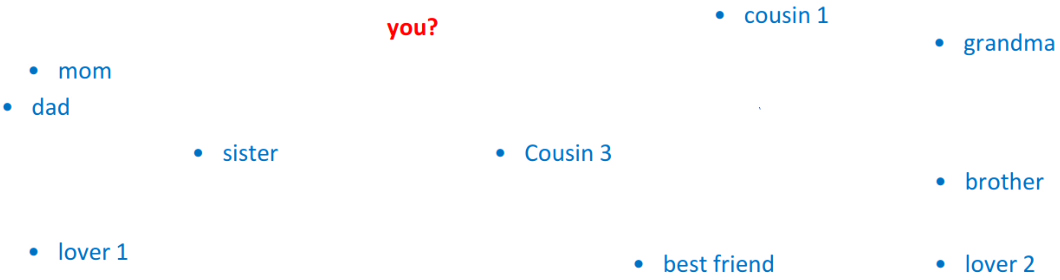
First order condition for optimality

Second order conditions for optimality

Unconstrained optimization: $\Omega = \mathbb{R}^n$

Example 1: The Fermat-Weber problem

You have a list of loved ones who live in given locations in Thailand. You would like to decide where to live so you are as close to them all as possible; say, you want to minimize the sum of distances to each person.



The Fermat-Weber problem

Location of person i:

Your location:

Problem:

Weighted sum of distances

Variant: also given weights w_i for each person
(your mom says you should care more about her than lover 1)

The Fermat-Weber problem

- As we'll see later, this optimization problem is "easy" to solve, because it has a nice structure (called convexity).

The Fermat-Weber problem

- As we'll see later, this optimization problem is ”easy” to solve, because it has a nice structure (called convexity).

- If you also wanted to be ”far” from some of your friends and family, this would have been a hard problem to solve!

The Fermat-Weber problem

- As we'll see later, this optimization problem is "easy" to solve, because it has a nice structure (called convexity).
- If you also wanted to be "far" from some of your friends and family, this would have been a hard problem to solve!
- A tiny variation in the problem makes the problem much harder.
- By the end of the course, you will learn techniques that will help you make such distinctions.

Unconstrained optimization: $\Omega = \mathbb{R}^n$

Example 2: Least squares

Given: $m \times n$ matrix

$m \times 1$ vector

Solve: $\min_{x \in \mathbb{R}^n} \|Ax - b\|^2$

In expanded notation, we are solving:

Some applications

Data fitting

Some applications

Overdetermined system of linear equations
A linear predictor for a stock price of a company:
$s(t)$: Stock price at day t

$$s(t) = a_1 s(t-1) + a_2 s(t-2) + a_3 s(t-3) + a_4 s(t-4)$$

We have three months of daily stock price data to train our model (lots of 5-day windows). How to find the best a_1, a_2, a_3, a_4 for future prediction?

Unconstrained local and global minima

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$
A point x^* is said to be a:

- Local minimum: $f(x^*) \leq f(x) \; \forall x$ near x^* ($\|x - x^*\|$ small)

Unconstrained local and global minima

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$
A point x^* is said to be a:

- Local minimum: $f(x^*) \leq f(x)$ $\forall x$ near x^* ($\|x - x^*\|$ small)

- Strict local minimum: $f(x^*) < f(x)$ $\forall x$ near x^* ($\|x - x^*\|$ small)

Unconstrained local and global minima

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$
A point x^* is said to be a:

- Local minimum: $f(x^*) \leq f(x) \; \forall x$ near x^* ($\|x - x^*\|$ small)

- Strict local minimum: $f(x^*) < f(x) \; \forall x$ near x^* ($\|x - x^*\|$ small)

- Global minimum: $f(x^*) \leq f(x) \; \forall x \in \mathbb{R}^n$

Unconstrained local and global minima

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$
A point x^* is said to be a:

- Local minimum: $f(x^*) \leq f(x) \; \forall x$ near x^* ($\|x - x^*\|$ small)

- Strict local minimum: $f(x^*) < f(x) \; \forall x$ near x^* ($\|x - x^*\|$ small)

- Global minimum: $f(x^*) \leq f(x) \; \forall x \in \mathbb{R}^n$

- Strict global minimum: $f(x^*) < f(x) \; \forall x \in \mathbb{R}^n$

Local/global maxima are defined analogously.

Example

- In general, finding local minima is easier than finding global minima.

- In general, finding local minima is easier than finding global minima.
- There are important problems where we can find global minima efficiently.
- On the other hand, there are problems where finding even a local minimum is intractable.

Optimization problems – basic notations

Unconstrained optimization

First order condition for optimality

Second order conditions for optimality

Notation reminder

$$f(x) = f(x_1, x_2, \ldots, x_n) \qquad (f : \mathbb{R}^n \to \mathbb{R})$$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} \quad \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n^2}(x) \end{pmatrix}$$

Gradient vector

Hessian matrix

First order condition

Theorem. **(First Order Necessary Condition for (Local) Optimality)**
If x^* is an unconstrained local minimizer of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, then we must have:

$$\nabla f(x^*) = 0$$

Fermat
(1607-1665)

Remarks

- This condition is necessary but not sufficient for local optimality.

Remarks

- This condition is necessary but not sufficient for local optimality.

- Nevertheless, it is useful because any local minimum must satisfy this condition. So, we can look for local (or global) minima only among points that make the gradient of the objective function vanish.

Remarks

- This condition is necessary but not sufficient for local optimality.

- Nevertheless, it is useful because any local minimum must satisfy this condition. So, we can look for local (or global) minima only among points that make the gradient of the objective function vanish.

- Terminology: A point x that satisfies $\nabla f(x) = 0$ is called a **stationary point** or a **critical point** of f. stationary point or a critical point of

Optimization problems – basic notations

Unconstrained optimization

First order condition for optimality

Second order conditions for optimality

Linear algebra review

Symmetric matrix: $A = A^T$

Theorem. Eigenvalues of a real symmetric matrix are real.

A square matrix A is said to be

- Positive semidefinite (psd) if:

- Positive definite (pd) if:

Linear algebra review

Recall that when we talk of psd or pd, we can assume that our matrix is symmetric: If A was not symmetric, we can replace A by $\frac{A^T + A}{2}$ as:

$$x^T A x = x^T \left(\frac{A^T + A}{2} \right) x$$

Linear algebra review

Theorem. A matrix is positive semidefinite iff all its eigenvalues are nonnegative. A matrix is positive definite iff all its eigenvalues are positive.

Examples:

1. $A = \begin{bmatrix} 2 & 4 \\ 4 & 5 \end{bmatrix}$
2. $A = \begin{bmatrix} 2 & 1 \\ 1 & 8 \end{bmatrix}$
3. $A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$

Second order condition

Theorem. (Second Order Necessary Condition for (Local) Optimality)
If x^* is an unconstrained local minimizer of a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, then, in addition to $\nabla f(x^*) = 0$, we must have:

$$\nabla^2 f(x^*) \succeq 0$$

(i.e., the Hessian at x^* is positive semidefinite.)

Second order condition

Theorem. **(Second Order sufficient Condition for (Local) Optimality)**
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, $\nabla f(x^*) = 0$ and

$$\nabla^2 f(x^*) \succ 0$$

(i.e. the Hessian at x^* is positive definite), then x^* is a strict local minimum of f.

Remarks

- $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$ is not sufficient for local optimality.

$$f(x) = x^3$$

- $\nabla^2 f(x) \succ 0$ is not necessary for (even strict global) optimality

$$f(x) = x^4$$

What are the questions in practice?

- How would we use all these optimality conditions to find local solutions and certify their optimality?

What are the questions in practice?

- How would we use all these optimality conditions to find local solutions and certify their optimality?
- Is it easy to find points satisfying these conditions? e.g., is it easy to solve $\nabla f(x) = 0$?

What are the questions in practice?

- How would we use all these optimality conditions to find local solutions and certify their optimality?
- Is it easy to find points satisfying these conditions? e.g., is it easy to solve $\nabla f(x) = 0$?
- Suppose you found that a given point is locally optimal, how would you go about checking if it is also globally optimal?

Exercise: State the analogues of the three theorems for local maxima.

Example 1: Least squares

Given: A $m \times n$ matrix (columns of A are independent)

b $m \times 1$ vector

Solve: $\min_{x \in \mathbb{R}^n} \|Ax - b\|^2$

Example 1: Least squares

Given: A $m \times n$ matrix (columns of A are independent)

b $m \times 1$ vector

Solve: $\min_{x \in \mathbb{R}^n} \|Ax - b\|^2$

Steps:

1. Write $f(x) = \|Ax - b\|^2$ as inner product of vectors
2. Solve $\nabla f(x) = 0$ for x
3. Show that $\nabla^2 f(x) = 2A^T A \succ 0$

Example 2

Find all the local minima and maxima of the following function:

$$f(x) = \frac{1}{2}x_1^2 + x_1 x_2 + 2x_2^2 - 4x_1 - 4x_2 - x_2^3$$

$\nabla f(x) =$

$\nabla f(x) = 0 \Rightarrow$

$$f(x) = \frac{1}{2}x_1^2 + x_1 x_2 + 2x_2^2 - 4x_1 - 4x_2 - x_2^3$$

$\nabla^2 f(x) =$

$x_2 = \qquad \Rightarrow \nabla^2 f(x) =$

$$f(x) = \frac{1}{2}x_1^2 + x_1x_2 + 2x_2^2 - 4x_1 - 4x_2 - x_2^3$$

$\nabla^2 f(x) =$

$x_2 = \qquad \Rightarrow \nabla^2 f(x) =$