MAT 3007 – Optimization

Convexity and Algorithms for Unconstrained Optimization Problems

Lecture 15 *July 13th*

Andre Milzarek SDS / CUHK-SZ

Repetition

Unconstrained	Constrained
First-Order Cond.: x^* local minimum (+ LICQ)	
▸ $\nabla f(x^*) = 0$.	▸ KKT-conditions.
Second-Order Cond.: x^* local minimum (+ LICQ)	
▸ $\nabla f(x^*) = 0$ ▸ $\nabla^2 f(x^*)$ is positive semi-definite (on $\mathbb{R}^n$).	▸ KKT-conditions ▸ $\nabla^2_{xx} L(x^*, \lambda, \mu)$ is positive semidefinite on $\mathcal{C}(x^*)$.
Second-Order Sufficient Cond.	
▸ $\nabla f(x^*) = 0$ and ▸ $\nabla^2 f(x^*)$ is positive definite (on $\mathbb{R}^n$).	▸ x^* is KKT-point and ▸ $\nabla^2_{xx} L(x^*, \lambda, \mu)$ is positive definite on $\mathcal{C}(x^*)$.
$\Longrightarrow$ x^* is strict local minimum	

General Strategy:

- Derive KKT-conditions; [Check LICQ (if required)].
- Discuss different easy cases via the complementarity conditions (set multiplier or constraints to 0) to find all KKT-points.
- Calculate $\mathcal{C}(x^*)$ and $\nabla^2_{xx} L(x^*, \lambda, \mu)$ at KKT-points.
- Check second-order conditions.

Additional Information:

- Check if f is coercive or if Ω is bounded $\rightsquigarrow$ the problem has global solutions (which must be KKT-points)!
- If the LICQ holds, then λ and μ are always unique!
- Finding maximizer: apply all steps to $-f$.

Logistics & Agenda

Logistics:

- The fifth sheet is online since Saturday. It is due on Monday, July 20th, 11:00 am.
- The midterm project is due on Saturday, July 18th, 11:00 pm.
- The tentative final examination period for summer courses is from August 24th to September 5th.
- CTE will be conducted from July 20th to July 24th. (Online system).

Agenda:

- Convexity.
- First Algorithms for Unconstrained Problems.

Convex Functions and Convex Problems

Motivation: So far we have been discussing local minimizers:

- When is a local minimizer also a global minimizer?
- We discuss a class of optimization problems that guarantees this property $\rightsquigarrow$ convex optimization.

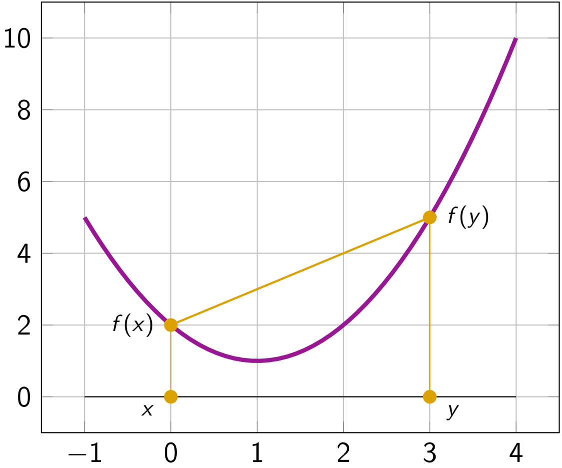
Definition: Convex Function

A function f on a convex set Ω is said to be convex if for every $x_1, x_2 \in \Omega$ and any $0 \le \lambda \le 1$,

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

- We call f a concave function if and only if $-f$ is convex.

Illustration of Convex Functions



Convexity and Concavity via Derivatives: Updated

Theorem: Convexity via Hessian

Let Ω be a convex set and let f be twice cont. differentiable on an open set containing Ω. Then f is convex on Ω if and only if its Hessian matrix is positive semidefinite, i.e.,

$$d^\top \nabla^2 f(x) d \geq 0 \quad \forall\, d \in \mathbb{R}^n, \quad \forall\, x \in \Omega.$$

Theorem: Concavity via Hessian

Let Ω be a convex set and let f be twice cont. differentiable on an open set containing Ω. Then f is concave on Ω if and only if its Hessian matrix is negative semidefinite, i.e.,

$$d^\top \nabla^2 f(x) d \leq 0 \quad \forall\, d \in \mathbb{R}^n, \quad \forall\, x \in \Omega.$$

Lemma: Sum Rule

If $a_1, \ldots, a_m \geq 0$, and $f_1, \ldots, f_m$ are convex (concave) functions, then $a_1 f_1 + \cdots + a_m f_m$ is a convex (concave) function.

- Examples: $x_1^2 + x_2^2$, $e^x + |x|$.

Lemma: Composition with Linear Functions

If $f : \mathbb{R}^m \to \mathbb{R}$ is convex (concave) and $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ are given, then $g : \mathbb{R}^n \to \mathbb{R}$, $g(x) := f(Ax + b)$, is convex (concave).

- Examples: e^{2x+3}, $(x_1 - x_2)^2 + (x_2 + x_3)^2$, $\|Ax - b\|$, $\log(-2x_1 + 3x_2 + 5)$ (concave).

Further Properties

Lemma: Taking Maximum

If $f_1, \ldots, f_m$ are convex functions, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is a convex function (this can be extended to uncountably many).

- Examples: $|x| = \max\{-x, x\}$, $\max\{a_i^\top x + b_i\}$.

Lemma: Taking Minimum

If $f_1, \ldots, f_m$ are concave function, then $f(x) = \min\{f_1(x), \ldots, f_m(x)\}$ is a concave function (this can be extended to uncountably many).

- Examples: $-|x| = \min\{-x, x\}$, $\min\{a_i^\top x + b_i\}$.

Consider the linear program

$$\begin{aligned} \text{minimize}_x \quad & c^\top x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Given A and b fixed, the optimal value function is a function of c. We denote the function by $V(c)$.

- In sensitivity analysis, we studied how $V(c)$ changes with c.

Theorem: Properties of V

V is a concave function of c.

- V is the minimum of a set of linear functions

$$V(c) = \min_{\{x : Ax = b, x \geq 0\}} \{c^\top x\}.$$

Theorem: Convexity and Global Solutions

Let $f : \Omega \to \mathbb{R}$ be a convex function and $\Omega \subset \mathbb{R}^n$ be a convex set. Then any local minimizer of the problem:

$$\begin{aligned} \text{minimize}_x \quad & f(x) \\ \text{s.t.} \quad & x \in \Omega \end{aligned}$$

is a global minimizer.

Proof: By contradiction. Assume x^* is a local minimizer, however, there exists $\bar{x} \in \Omega$ such that $f(\bar{x}) < f(x^*)$. Then, using convexity, we have

$$f(\lambda\bar{x} + (1-\lambda)x^*) \leq \lambda f(\bar{x}) + (1-\lambda)f(x^*) < f(x^*)$$

for any $0 < \lambda < 1$. This is a contradiction to: x^* is a local min. $\square$

Theorem: Stationarity & Global Optimality

Let f be convex and suppose that $\Omega := \{x : g(x) \leq 0, h(x) = 0\}$ is a convex set. Then, the KKT conditions for the problem

$$\text{minimize}_x \quad f(x) \qquad \text{s.t.} \quad x \in \Omega$$

are sufficient for global optimality.

Remarks:

- In a Nutshell: If f and Ω are convex, then stationary points and KKT-points are already local and global minimizer!
- If f is concave and Ω is convex, then stationary points and KKT-points of the problem $\min_{x \in \Omega} -f(x)$ are local and global maximizer of f.

Convexity/concavity plays a very important role in optimization problems!

We call the optimization problems of the form:

- Minimize a convex function over a convex feasible region
- Maximize a concave function over a convex feasible region

convex optimization problems.

Otherwise, the problem is called a non-convex optimization problem.

In optimization, convexity and non-convexity typically determine whether a problem is easy or hard.

Convex Constraints

Constraint Types

What constraints would make the feasible region convex?

Lemma: Convex Level Sets

Let f be a convex (concave) function. Then, for any c, the level set $L_{\leq c} = \{x : f(x) \leq c\}$ ($L_{\geq c} = \{x : f(x) \geq c\}$) is a convex set.

Observation:

- If we have constraints of the form $g(x) \leq 0$ and g is convex, then this is a convex constraint!
- If we have constraints of the form $g(x) \geq 0$ and g is concave, then this is a convex constraint!
- Linear constraints are always convex constraints.
- Sometimes, even if a constraint does noot appear to be in the above form, it still could be a convex constraint.

Being able to identify convex problems is an important skill.

Example I

Is this a convex optimization problem?

$$\begin{array}{ll} \text{minimize} & 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subject to} & x_1^2 + x_2^2 \leq 5 \\ & 3x_1 + x_2 \geq 3 \end{array}$$

- Answer: Yes

What if we change the constraint $x_1^2 + x_2^2 \leq 5$ to $x_1^2 + x_2^2 \geq 5$?

- Then it no longer is a convex optimization problem.

Example II

How about

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}} \quad & x^\top Q x - c^\top x \\ \text{s.t.} \quad & Ax = b \\ & Cx \geq d \\ & x \geq 0 \end{aligned}$$

- The constraints are linear.
- It is a convex optimization problem if and only if Q is PSD.

Example III

Consider the optimization problem:

$$\begin{aligned} \text{maximize}_{x,y,z} \quad & xyz \\ \text{s.t.} \quad & x + 2y + 3z \le 3 \\ & x, y, z \ge 0 \end{aligned}$$

In order for a maximization problem to be a convex optimization problem, we need the objective function to be concave.

- However, xyz is not a concave function in x, y, z.

But we can transform this into maximizing $\log(xyz)$. The problem becomes:

$$\begin{aligned} \text{maximize} \quad & \log x + \log y + \log z \\ \text{s.t.} \quad & x + 2y + 3z \le 3 \\ & x, y, z \ge 0 \end{aligned}$$

which is a convex optimization problem.

Strategies:

- Often, we can apply monotone transformations or variable substitutions.
- Sometimes one has to look at the defined region explicitly.

Examples:

- $\{x : x^3 - 1 \le 0\}$.
  $g(x) = x^3$ is not a convex function. However, this constraint defines a convex feasible region ($\equiv \{x : x \le 1\}$).
- $\{z^2 - xy \le 0,\ x, y, z \ge 0\}$.
  $g(x, y, z) = z^2 - xy$ is not a convex function. The Hessian is

$$\nabla^2 g(x, y, z) = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

  with eigenvalues -1, 1 and 2. But this gives a convex region.

Software to Solve Convex Optimization Problems

MATLAB has its own functions: `fmincon` and `fminunc`.

- However, in general they are not very scalable nor fast. (One has to provide the right information + adjust parameter).

We suggest to use CVX. CVX can solve a large range of nonlinear optimization problems.

- CVX can only solve convex optimization problems (that is what it is named for).
- It can only recognize certain classes of convex functions.
- Sometimes, one has to manually convert a problem into a recognizable form before inputting into CVX.

Examples

Example 1:

$$\begin{array}{ll} \text{minimize} & (x_1 - 1)^2 + (x_2 - 1)^2 \\ \text{s.t.} & x_1 + x_2 = 1 \end{array}$$

Example 2:

$$\begin{array}{ll} \text{minimize} & e^{x_1 + x_2} + (x_1 - 0.5x_2)^2 + 2.75x_2^2 \\ \text{s.t.} & x_1 + 2x_2 = 1 \end{array}$$

Let $y^1, y^2, ..., y^k \in \mathbb{R}^2$ be k different points.

We want to find a circle in $\mathbb{R}^2$ with minimum radius that contains all of these points:

$$\begin{array}{ll} \min_{y \in \mathbb{R}^2, r \in \mathbb{R}} & r \\ \text{subject to} & \|y - y^1\| \le r, \;\; \|y - y^2\| \le r, \;\; \dots \;\; , \|y - y^k\| \le r, \\ & r \ge 0. \end{array}$$

- This is a convex optimization problem.
- The equivalent (differerentiable) formulation $\|y - y^i\|^2 \le r^2$ will be rejected by CVX!

Algorithms for Unconstrained Problems

Upcoming Agenda

We now discuss how to solve nonlinear optimization problems.

- In many cases, the KKT conditions can be used to solve the optimization problem.
- However, those are ad hoc situations. In most situation, it is too complicated to directly find the optimal solution from the KKT conditions.
- We want to have a robust procedure (an algorithm) that allows to solve the optimization problem.

We start with the unconstrained problem:

$$\text{minimize}_{x \in \mathbb{R}^n} \quad f(x)$$

We are going to study the following methods:

- Bisection search.
- Golden section search.
- Gradient descent method.
- Newton's method

Typically, optimization algorithms are iterative procedures:

- Starting from some point x^0, we generate a sequence of iterates $\{x^k\}$.
- The sequence terminates when either no progress can be made or when we know that the current step is already satisfactory.
- Typically, we want to have $f(x^{k+1}) < f(x^k)$, i.e., each step we can improve the objective value.
- And hopefully, the sequence $\{x^k\}$ converges to a local minimizer x^* (or global minimizer).

Recall the algorithms we have studied so far: the simplex method and the interior point method.

They both follow the above paradigm.

Definition: Convergence

Let $\{x^k\}$ be a sequence of real vectors. Then $\{x^k\}$ converges to x^* if and only if for every $\epsilon > 0$, there exists a positive integer K such that $\|x^k - x^*\| < \epsilon$ for all $k \geq K$.

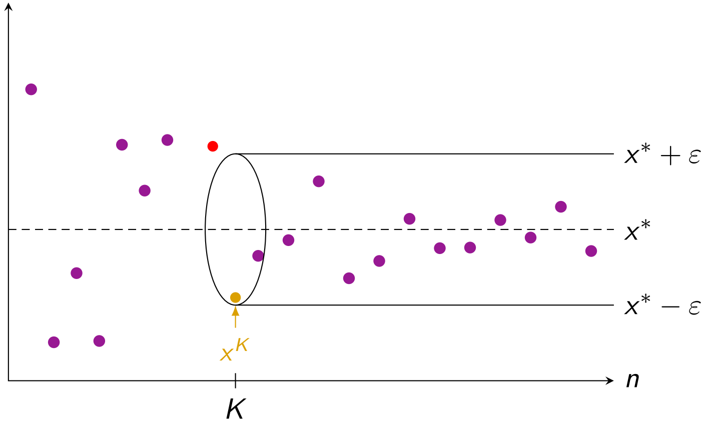
In all our discussions, we assume that $\|\cdot\|$ is the Euclidean norm, which means:

$$\|x\| = \sqrt{x^\top x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

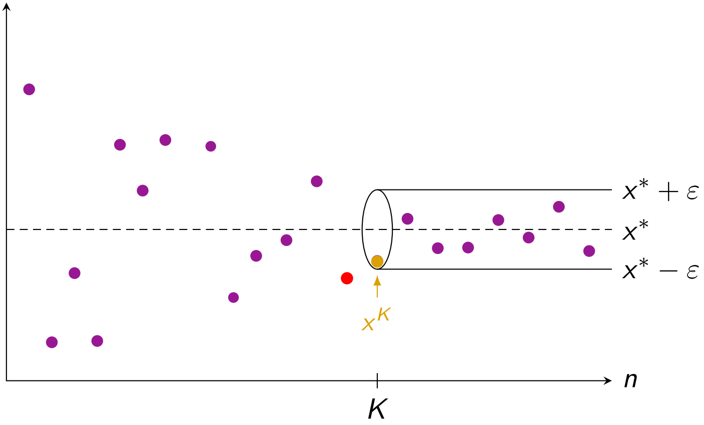
Examples of convergent sequences:

- $x^k := 1/k$ for all k; then $x^k \to 0$.
- $x^k := (1/2)^k$ for all k; then $x^k \to 0$.

Convergence: Illustration



Convergence: Illustration



Problems in $\mathbb{R}$

Assume $f : \mathbb{R} \to \mathbb{R}$ is a single variable function.

Our Objective: find a local minimizer of f.

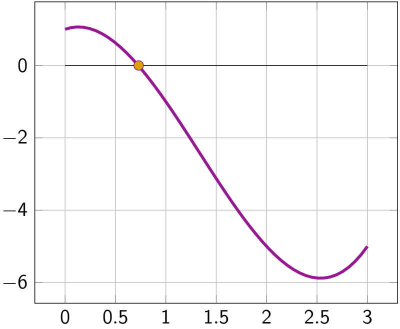
We introduce two methods:

- Bisection method.
- Golden section method.

Bisection method uses the idea that the local minimizer must satisfy the first-order necessary conditions: $f'(x) = 0$.

Therefore, the problem becomes a root-finding problem for

$$g(x) = f'(x) = 0.$$



Root Finding Algorithm: Bisection Method

Assume we can find x_ℓ and x_r such that $g(x_\ell) < 0$ and $g(x_r) > 0$.

By the intermediate value theorem, if g is continuous, there must exist a root of g in $[x_\ell, x_r]$.

Bisection Method

1. Define $x_m = \frac{x_\ell + x_r}{2}$.
2. If $g(x_m) = 0$, then output x_m.
3. Otherwise:
    - If $g(x_m) > 0$, then let $x_r = x_m$.
    - If $g(x_m) < 0$, then let $x_\ell = x_m$.
4. If $|x_r - x_\ell| < \epsilon$: stop and output $\frac{x_\ell + x_r}{2}$, otherwise go back to step 1.

One can also set the stopping criterion based on $|g(x)| < \epsilon$.

Bisection Method

In the bisection method, each iteration will divide the search interval to half.

Therefore, to find an ϵ approximation of x^*, we need at most $\log_2 \frac{x_r - x_\ell}{\epsilon}$ many iterations.

Applying the bisection method to f', we can find an approximate stationary point. If f is convex, this is an (approximate) global minimizer of f.

- Although simple, the bisection method is very useful in practice because it is easy to implement.

Example: Use bisection method to maximize:

$$f(x) = \frac{xe^{-x}}{1 + e^{-x}} \quad \rightsquigarrow \quad f'(x) = \frac{e^{-x}(1 - x + e^{-x})}{(1 + e^{-x})^2}$$

Bisection Method

```
function [x,gx] = bisection(g,xl,xr,options)

% Compute intial function values
gr  = g(xr); gl  = g(xl); sl = sign(gl);

if gl*gr > 0
    fprintf(1,'The input data not suitable!');
    x = []; gx = []; return
end

for i = 1:options.maxit
    xm  = (xl + xr)/2; gm  = g(xm);

    if abs(gm) < options.tol || abs(xl-xr) < options.tol
        x = xm; gx = gm; return
    end

    if gm > 0
        if sl < 0, xr = xm; else, xl = xm; end
    else
        if sl < 0, xl = xm; else, xr = xm; end
    end
end
```

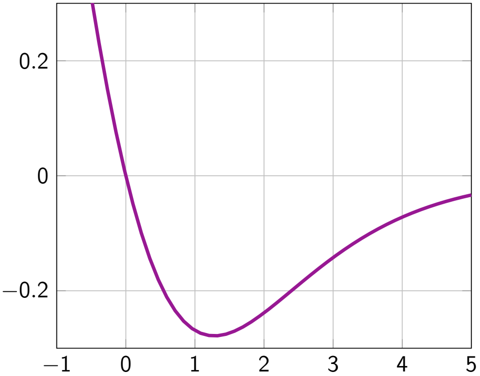
Drawback of the bisection method: When solving (single variable, unconstrained) optimization problems, we require the knowledge (and computation) of f'.

- Sometimes, f' is not available. For example, f sometimes is only a black box, which does not admit an analytical form (thus, the derivative is hard to compute)

However, if we know that f has a unique local minimum x^* in the range $[x_\ell, x_r]$, then we still have a very efficient way to find x^*:

- We call f unimodal if it only has one single stationary point (on $\mathbb{R}$).
- Unimodal functions have the property that the local minimum is already global. (Similarly, if the stationary point is a local maximum).

Consider $f(x) = -\frac{xe^{-x}}{1+e^{-x}}$:



This is a unimodal function, but not a concave function.

Golden Section Method

Assume we start with $[x_\ell, x_r]$. Assume $0 < \phi < 0.5$.

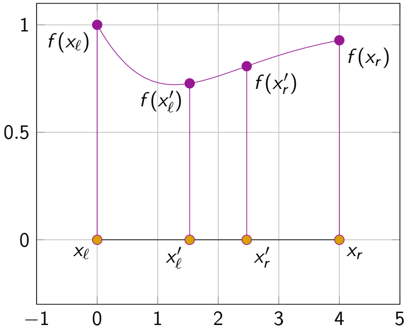
1. Set $x_\ell' = \phi x_r + (1-\phi)x_\ell$ and $x_r' = (1-\phi)x_r + \phi x_\ell$.
2. If $f(x_\ell') < f(x_r')$, then the minimizer must lie in $[x_\ell, x_r']$, so set $x_r = x_r'$.
3. Otherwise, the minimizer must lie in $[x_\ell', x_r]$, so set $x_\ell = x_\ell'$.
4. If $x_r - x_\ell < \epsilon$, output $\frac{x_\ell + x_r}{2}$, otherwise go back to step 1.

- Suppose we update $x_r = x_r'$. We want to choose ϕ such that x_r' of the new iteration coincides with x_ℓ' of the old iteration.

$\rightsquigarrow$ This allows to save one function evaluation!

- This is true when

$$\phi = \frac{3-\sqrt{5}}{2} \quad \text{and} \quad 1-\phi = \frac{\sqrt{5}-1}{2} = 0.618.$$



Both the bisection and golden section method can be easily adapted for maximization problems. (Just adjust the comparison).

Example Revisited: Use the Golden section method to maximize:

$$f(x) = \frac{xe^{-x}}{1 + e^{-x}}$$

Questions?