MAT 3007 – Optimization

Optimality for Constrained Problems and Convexity

Lecture 14

July 9th

Andre Milzarek

SDS / CUHK-SZ

Repetition

Recap: Lagrangian

Setup – General Nonlinear Optimization Problem:

$$\begin{aligned} \text{minimize}_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \quad \forall\ i = 1, \dots, m, \\ & h_j(x) = 0, \quad \forall\ j = 1, \dots, p. \end{aligned}$$

- The feasible set is $\Omega = \{x \in \mathbb{R}^n : g(x) \le 0,\ h(x) = 0\}$.
- For $x \in \Omega$, the set $\mathcal{A}(x) := \{i : g_i(x) = 0\}$ denotes the set of active constraints.
- The set of inactive constraints is $\mathcal{I}(x) := \{i : g_i(x) < 0\}$.

Lagrangian:

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Theorem: KKT Conditions

If x is a local minimizer and if a regularity condition $(\star)$ holds, then there exist λ and μ such that:

1. Main Condition

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{p} \mu_j \nabla h_j(x) = 0.$$

2. Dual Feasibility

$$\lambda_i \geq 0 \quad i = 1, ..., m.$$

3. Complementarity

$$\lambda_i \cdot g_i(x) = 0 \quad \forall\ i = 1, ..., m.$$

We often add primal feasibility as part of the KKT conditions:

4. Primal Feasibility

$$g_i(x) \leq 0, \quad h_j(x) = 0 \quad \forall\ i, \quad \forall\ j.$$

Recap: CQ and KKT Points

Linear Independence Constraint Qualification (LICQ): We require the collection of gradients

$$\{\nabla g_i(x) : i \in \mathcal{A}(x)\} \cup \{\nabla h_j(x) : j = 1, ..., p\} \qquad (\star)$$

to be linearly independent or to have full rank.

- A feasible point x satisfying the LICQ is called regular.

$\rightsquigarrow$ In this course, we usually assume that this CQ is satisfied.

- A (feasible) point satisfying the KKT conditions is called a KKT point.

Logistics:

- The fourth sheet is due on Sunday, July 12th, 11:00 am.
- The fifth exercise sheet will be available on Thursday or Friday.

Agenda:

- More examples and applications.
- Second-order optimality conditions for constrained problems.
- Visualization, connections, summary.
- Convexity.

Examples: Formulating and Using KKT Conditions

Example IV

Task: We want to build a box with a given volume of at least 64 cubic inches. We want to minimize the total amount of material used.

We can formulate the optimization problem as:

$$\begin{array}{ll} \text{minimize} & 2xy + 2yz + 2xz \\ \text{s.t.} & xyz \geq 64 \end{array}$$

Set $g(x, y, z) = 64 - xyz$ and let $\lambda \geq 0$ be the dual multiplier for this problem.

The KKT conditions say:

$$2\begin{pmatrix} y+z \\ x+z \\ x+y \end{pmatrix} = \lambda \begin{pmatrix} yz \\ xz \\ xy \end{pmatrix}, \quad \lambda \geq 0, \quad \lambda \cdot (xyz - 64) = 0$$

Case 1: $\lambda = 0$. Then we must have $x = y = z = 0$, however, this point does not satisfy the constraint.

Case 2: $\lambda > 0$ and $xyz = 64$. By the first equality, we have

$$\lambda = 2\left(\frac{1}{x} + \frac{1}{y}\right) = 2\left(\frac{1}{y} + \frac{1}{z}\right) = 2\left(\frac{1}{x} + \frac{1}{z}\right)$$

Thus, $x = y = z = 4$ is the only solution of the KKT conditions.

Since this problem must have a finite optimal solution, it must be the optimal solution.

Second-Order Optimality Conditions for Constrained Problems

- The KKT-conditions are necessary first-order optimality conditions.

$\rightsquigarrow$ KKT-points are potential candidates for local minimizer.

Question:

- Is it possible to use second-order conditions as in the unconstrained case? Answer: $\rightsquigarrow$ Yes!
- We assume that f, g_i, and h_j are twice cont. differentiable.

The Hessian of the Lagrangian is given by:

$$\nabla^2_{xx} L(x, \lambda, \mu) = \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x) + \sum_{j=1}^{p} \mu_j \nabla^2 h_j(x).$$

Second-Order Necessary Conditions

We define the so-called critical cone:

$$\mathcal{C}(x) := \{d \in \mathbb{R}^n : \nabla f(x)^\top d = 0,\ \nabla g_i(x)^\top d \le 0,\ \forall\ i \in \mathcal{A}(x),$$
$$\nabla h_j(x)^\top d = 0,\ \forall\ j\}.$$

The second-order necessary conditions take the following form:

Theorem: SONC for Constrained Problems

Let x^* be a regular point and local min. Then, the KKT-conditions hold and there are unique multiplier $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that:

$$\nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu = 0,$$
$$g(x^*) \le 0,\ h(x^*) = 0, \quad \lambda \ge 0, \quad \lambda_i \cdot g_i(x^*) = 0 \quad \forall\ i$$

and we have:

$$d^\top \nabla^2_{xx} L(x^*, \lambda, \mu) d \ge 0 \quad \forall\ d \in \mathcal{C}(x^*).$$

Remark:

- The uniqueness of λ and μ in the SONC follows from the LICQ. (This can be helpful in calculations).

The second-order sufficient conditions take the following form:

Theorem: SOSC for Constrained Problems

Let x^* be a KKT-point with multiplier λ and μ, i.e., we have

$$\begin{aligned}&\nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu = 0,\\ &g(x^*) \leq 0,\ h(x^*) = 0, \quad \lambda \geq 0, \quad \lambda_i \cdot g_i(x^*) = 0 \quad \forall\ i\end{aligned}$$

and suppose that the condition

$$d^\top \nabla^2_{xx} L(x^*, \lambda, \mu) d > 0 \quad \forall\ d \in \mathcal{C}(x^*) \backslash \{0\}.$$

is satisfied. Then, x^* is a strict local minimizer.

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	

Examples: Applying Second-Order Conditions

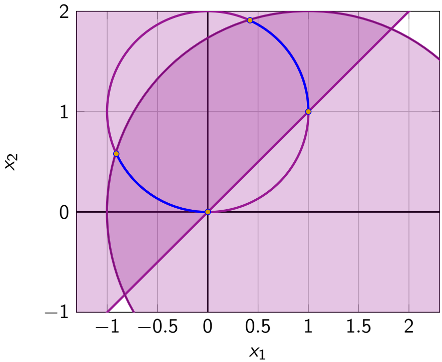
Example I

Consider the problem

$$\min_x \; x_1^3 x_2 - 2x_1^2 + 3x_2 \quad \text{s.t.} \quad g(x) \le 0, \quad h(x) = 0,$$

where $g_1(x) = (x_1 - 1)^2 + x_2^2 - 4$

$$g_2(x) = x_1 - x_2, \quad h(x) = x_1^2 + (x_2 - 1)^2 - 1.$$



Example I: Continued

Consider the point: $\bar{x} = (0, 0)^\top$.

Typical Tasks and Questions:

- Show that the LICQ holds at $\bar{x}$.
- Is $\bar{x}$ a KKT-point? If yes, calculate the associated Lagrangian multiplier λ and μ!
- Compute the critical cone $\mathcal{C}(\bar{x})$ and $\nabla^2_{xx} L(\bar{x}, \lambda, \mu)$.
- Is $\bar{x}$ a local solution of the problem?

Example II

Consider the nonlinear program

$$\min_{x} \; (2x_1 - 1)^2 + x_2^2 \quad \text{s.t.} \quad h(x) = -2x_1 + x_2^2 = 0.$$

Task: Solve this problem and find all global and local solutions!

General Strategy:

- Check LICQ (if required).
- Derive KKT-conditions.
- 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:

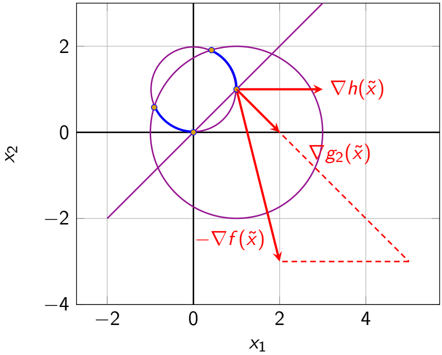
- Check if f is coercive or if Ω is bounded $\rightsquigarrow$ the problem has global solutions (which must be KKT-points)!
- If the LICQ holds, then λ and μ are always unique!
- Finding maximizer: apply all steps to $-f$.

Visualization and Interpretation of the KKT Conditions

Reconsider problem I at $\tilde{x} = (1,1)^\top$. We have $\mathcal{A}(\tilde{x}) = \{2\}$ and

$$\nabla f(\tilde{x}) = (-1,4)^\top,\ \nabla g_2(\tilde{x}) = (1,-1)^\top,\ \nabla h(\tilde{x}) = (2,0)^\top$$

The KKT-conditions mean: $-\nabla f(\tilde{x}) = \nabla g_2(\tilde{x})\lambda_2 + \nabla h(\tilde{x})\mu$ for some $\lambda_2 \geq 0$:



Following our derivation, the KKT-conditions imply that the LP

$$\begin{array}{ll} \max_{\lambda \geq 0, \mu} & 0 \\ \text{s.t.} & \nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) = 0 \end{array}$$

is feasible. The dual of the problem is:

$$\begin{array}{ll} \min_d & -\nabla f(x^*)^\top d \\ \text{s.t.} & \nabla g_i(x^*)^\top d \geq 0, \ \forall\ i \in \mathcal{A}(x^*), \ \nabla h(x^*)^\top d = 0. \end{array}$$

Hence, by strict duality and setting

$$\mathcal{T}_\ell(x^*) := \{d : \nabla g_i(x^*)^\top d \geq 0, \ \forall\ i \in \mathcal{A}(x^*), \ \nabla h(x^*)^\top d = 0\},$$

the KKT-conditions are equivalent to:

$$\nabla f(x^*)^\top d \geq 0 \quad \forall\ d \in \mathcal{T}_\ell(x^*).$$

Final Comments:

- In the KKT-conditions, we substitute the set of feasible directions $S_\Omega(x^*)$ with the simpler set $\mathcal{T}_\ell(x^*)$.

  $\rightsquigarrow$ We require a CQ that allows us to do this.

- The set $\mathcal{T}_\ell(x^*)$ is called linearized tangent set.

- The objective increases along directions $d \in \mathcal{T}_\ell(x^*)$ such that $\nabla f(x^*)^\top d > 0$.

  $\rightsquigarrow$ In the SOC, we only need to consider directions $d \in \mathcal{C}(x^*)$.

Convexity

So far we have been discussing local minimizers:

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

Definition: Convex Set

A set $\Omega \subseteq \mathbb{R}^n$ is convex if for any x, $y \in \Omega$, and any $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda)y \in \Omega$.

Convex Combination

For any $x_1, ..., x_n$ and $\lambda_1, ..., \lambda_n \geq 0$ satisfying $\lambda_1 + \cdots + \lambda_n = 1$, we call $\sum_{i=1}^{n} \lambda_i x_i$ a convex combination of $x_1, ..., x_n$.

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 $0 \leq \lambda \leq 1$,

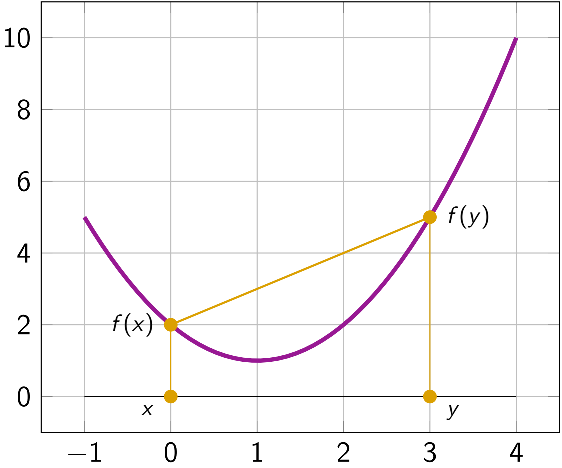
$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$

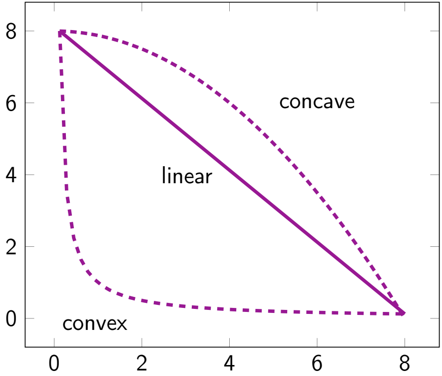
Definition: Concave Function

We call f a concave function if and only if $-f$ is convex, i.e., for any x_1, x_2 and $0 \leq \lambda \leq 1$,

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

Illustration of Convex Functions





Some examples of convex functions:

- $f(x) = x$, $f(x) = x^2$, $f(x) = e^x$, $f(x) = |x|$.

Some examples of concave functions:

- $f(x) = x$, $f(x) = \sqrt{x}$, $f(x) = \log x$

Theorem: Convexity via Hessian

Let f be twice cont. differentiable. Then f is convex (on $\mathbb{R}^n$) if and only if its Hessian matrix is positive semidefinite, i.e.,

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

- In $\mathbb{R}$, this means that the second-order derivative is non-negative.
- Taking second-order derivatives is usually the easiest way to test convexity
- Examples: check whether $x \log x$, $\|x\|^2$ are convex?

Otherwise, convexity is typically tested by definition (or using some rules).

Theorem: Concavity via Hessian

Let f be twice cont. differentiable. Then f is convex (on $\mathbb{R}^n$) if and only if its Hessian matrix is positive semidefinite, i.e.,

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

- In $\mathbb{R}$, this means that the second-order derivative is non-positive.
- Examples: $x^{1/2}$, $\log x$.

Lemma: Sum Rule

If $a_1, ..., a_m \geq 0$, and $f_1, ..., f_m$ 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).

Lemma: Taking Maximum

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

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

Another Example: Linear Programming

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

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

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

Stationarity and Global Optimality

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

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

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.

Questions?