# MAT 3007 – Optimization

## The Interior Point Method and Nonlinear Programming

*Lecture 11*

*July 2nd*

Andre Milzarek

iDDA / CUHK-SZ

# Interior Point Method

# Start: A New Perspective on Solving LPs

- We have studied duality theory for linear optimization.
- We now consider the simplex method from a new perspective.

We start with the optimality conditions for LPs (in standard form):

1. Primal Feasibility: $Ax = b$, $x \geq 0$.
2. Dual Feasibility: $A^\top y \leq c$.
3. Complementarity: $x_i \cdot s_i = x_i \cdot (c_i - A_i^\top y) = 0$ for each $i$.

These conditions are necessary and sufficient for $x$ and $y$ being the optimal solutions of the primal and dual problems.

In the simplex method, we search among basic feasible solutions:

$\rightsquigarrow$ We always maintain primal feasibility!

For any basis $B$, define $y = (A_B^{-1})^\top c_B$. Note that the reduced costs are by $c^\top - y^\top A$. For basic variables, the reduced costs are zero. Consequently:

$$x_i \cdot (c_i - A_i^\top y) = 0 \quad \forall\ i.$$

$\rightsquigarrow$ The complementarity conditions are always satisfied!

During the simplex method, the reduced costs may be negative:

- Hence, the dual solution $y = (A_B^{-1})^\top c_B$ may not always satisfy the constraint $A^\top y \leq c$.
- The method stops whenever the reduced costs are nonnegative.
- We seek $y$ satisfying $A^\top y \leq c$, i.e., dual feasibility.

**Proposition: Optimality Properties of the Simplex Method**

During each iteration, the simplex method maintains primal feasibility and the complementarity conditions. It seeks a solution that is dual feasible.

What if we choose to maintain the other two conditions?

- This will result in the dual-simplex method and the interior point method.

One can view the dual simplex method to as a simplex method applied to the dual problem of an LP.

- It maintains dual feasibility.
- It maintains the complementarity conditions.
- However, primal feasibility does not need to be satisfied during the iterations. It seeks for primal feasible BFS.

The tableau can be seen as a rotated variant of the primal one.

There are cases where using the dual simplex method can be more convenient:

- If a dual BFS is available (but we don't have a primal BFS).
- We have mentioned a scenario in the discussion of the sensitivity analysis (when $b$ is changed by a large amount or a constraint is added)

Optimality Conditions for LPs:

1. Primal Feasibility: $Ax = b$, $x \geq 0$.
2. Dual Feasibility: $A^\top y \leq c$.
3. Complementarity: $x_i \cdot s_i = x_i \cdot (c_i - A_i^\top y) = 0$ for each $i$.

The interior point method maintains both primal feasibility and dual feasibility during its iterations and seeks for a pair of solutions that satisfy the complementarity conditions.

# High-Level Idea

We want to find $x$, $y$ and $s$ such that:

$$\begin{aligned} Ax = b, &\quad x \geq 0 \\ A^\top y + s = c, &\quad s \geq 0 \\ x_i \cdot s_i = 0, &\quad \forall\ i. \end{aligned}$$

This is a set of nonlinear equations. It is not obvious to find a solutions of this system.

# High-Level Idea: Continued

We consider a relaxed version of the problem.

$$\begin{aligned} Ax = b, &\quad x \geq 0 \\ A^\top y + s = c, &\quad s \geq 0 \\ x_i \cdot s_i \leq \mu, &\quad \forall\ i. \end{aligned}$$

We call $\mu > 0$ the complementarity gap.

Idea: If we have found a solution for a certain $\mu$, then it might be possible to find a solution for a smaller $\mu$. Then we keep decreasing $\mu$ until we can find a solution of the LP.

- The essential step in the interior point method is to show that this is indeed doable – the approach is similar to Newton's method, which we discuss in the second half of the semester.

# Why “Interior Point” Method?

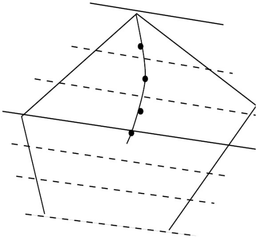


Figure: Central path in an interior point method

- In the simplex method, we only search among the extreme points (at the boundary of the polytope).
- In the interior point method, we start in the interior of the feasible region.
- Until we reach an optimal solution, we keep $x > 0$ and $s > 0$.
- The optimal solution recovered by the interior point method may not be a BFS (if the solution is unique, it must be a BFS).

We also need to resolve the issue of finding an initial basic solution in the interior point method:

- This can be resolved by solving an auxiliary problem (called the homogeneous self-dual problem).

Complexity: The interior point method is a polynomial-time algorithm with overall complexity $\approx O(n^{3.5})$.

- There are several variants of the interior point method. The one we introduced is called the primal-dual type of interior point method.
- The main ideas of other variants are similar, i.e., going through the interior of the feasible region and seeking complementarity.

# Practical Performance

On average, the speed of the simplex method is comparable with the speed of the interior point method despite their difference in theoretical complexity.

- For some problems, simplex method can be very fast (within a few iterations); for other problems, the simplex method requires some extra time (worst-case exponential).
- In contrast, the running time of the interior point method is quite stable, it does not vary much from problem to problem (given a fixed size).

**Theorem: Quality of Solutions**

The interior point method will always find the optimal solution with the maximum possible number of non-zeros.

Consider a simple case:

$$\begin{aligned} \text{minimize}_x \quad & x_1 + x_2 + x_3 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 1 \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

- The simplex method will give a BFS as optimal solution (one of $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$).
- The interior point method will return $(1/3, 1/3, 1/3)$.

# Discussion

In the case of multiple optimal solutions:

- If we want a high-rank solution (with the maximum possible non-zeros), then choose the interior point method.
- If we want a low-rank solution (with small number of non-zeros), then choose the simplex method.

- The optimal solution output of the simplex method may depend on the initial solution (as well as on the pivoting rules).

Either situation could be desirable in practice.

High-rank:

- In the multi-firm alliance problem, a high-rank (fair) allocation.

Low-rank:

- In portfolio problems, we want to minimize the number of stocks chosen (reduce the transaction cost).
- In graph problems, we want to use fewer nodes/edges.
- In other cases, we prefer integer solutions over fractional solutions (given that their objective values are equal). This usually corresponds to low-rank solutions.

# Software Choices

Both the simplex method and the interior point method are used in major commercial software:

- In MATLAB, we can specify which method we want to use in the function called `linprog`.
- Same in CPLEX.
- CVX uses the interior point method.
- Excel uses the simplex method.

## Nonlinear Programming

So far we have discussed linear optimization problems. However, in practice, there are many interesting optimization problems that do not take a linear form.

In general, we can write a nonlinear optimization problem as:

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

We call $\Omega$ the feasible set and $x \in \Omega$ are feasible points.

In the following, we study such nonlinear optimization problems:

- Geometric properties and optimality conditions.
- How can we find the optimal solution?
- We always assume that we are solving a minimization problem.

Let $\Omega \subseteq \mathbb{R}^n$ be a nonempty set and let $f : \Omega \to \mathbb{R}$ be given. We define $B_\varepsilon(y) := \{x \in \mathbb{R}^n : \|x - y\| < \varepsilon\}$ to be the open ball in $\mathbb{R}^n$ with center $y$ and radius $\varepsilon > 0$.

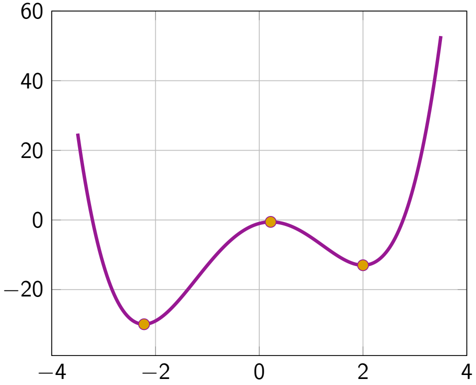
The point $x^* \in \mathbb{R}^n$ is said to be a:

- local minimizer, if $x^* \in \Omega$ and there exists $\varepsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \Omega \cap B_\varepsilon(x^*)$.
- strict local minimizer, if $x^* \in \Omega$ and there is $\varepsilon > 0$ with $f(x) > f(x^*)$ for all $x \in (\Omega \cap B_\varepsilon(x^*)) \backslash \{x^*\}$.
- global minimizer, if $x^* \in \Omega$ and we have $f(x) \geq f(x^*)$ for all $x \in \Omega$.
- strict global minimizer, if $x^* \in \Omega$ and we have $f(x) > f(x^*)$ for all $x \in \Omega \backslash \{x^*\}$.

- Remark: global minimizer $\equiv$ global solution $\equiv$ optimal sol.
- The def. for maximizer is identical, changing: $\geq / > \to \leq / <$.

We consider the unconstrained problem

$$\mathrm{minimize}_{x\in\mathbb{R}}\ f(x) := x^4 - 9x^2 + 4x - 1.$$



- Assume $f(\boldsymbol{x}) = f(x_1, x_2, ..., x_n)$ is continuously differentiable. Then we denote the gradient of $f$ by (an $n \times 1$ vector):

$$\nabla f(x) = \left( \frac{\partial f}{\partial x_1}; \frac{\partial f}{\partial x_2}; ...; \frac{\partial f}{\partial x_n} \right)$$

The first-order Taylor expansion yields:

$$f(x + td) = f(x) + t\nabla f(x)^\top d + o(t), \quad t \to 0.$$

- If $f$ is twice continuously differentiable, then the Hessian of $f$ (an $n \times n$ matrix) is given by:

$$\nabla^2 f(x) = \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j}$$

By a second-order Taylor expansion, we obtain:

$$f(x + td) = f(x) + t\nabla f(x)^\top d + \frac{1}{2}t^2 d^\top \nabla^2 f(x) d + o(t^2), \quad t \to 0.$$

Suppose:

$$f(x_1, x_2, x_3) := x_1^2 + x_1 x_2 + x_1 e^{x_3} + x_2 \log x_3$$

Then:

$$\nabla f(x) = \left(2x_1 + x_2 + e^{x_3}, x_1 + \log x_3, x_1 e^{x_3} + \frac{x_2}{x_3}\right)^\top$$

and

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 1 & e^{x_3} \\ 1 & 0 & \frac{1}{x_3} \\ e^{x_3} & \frac{1}{x_3} & x_1 e^{x_3} - \frac{x_2}{x_3^2} \end{pmatrix}.$$

# Optimality Conditions

In the following, we first study what conditions an optimal solution has to satisfy:

$\rightsquigarrow$ First- and second-order optimality conditions.

- We will first start with local optimal solutions.

Let us fix $\Omega = \mathbb{R}^n$ (unconstrained problems).

What are the optimality conditions for local minimizers for unconstrained problems?

Claim: We must have:

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

Reason: If $\nabla f(x) \neq 0$, then we can find a vector $d$ such that $\nabla f(x)^\top d < 0$. Therefore, by Taylor expansion:

$$f(x + td) = f(x) + t\nabla f(x)^\top d + o(t), t \to 0.$$

By choosing $t$ small enough, we can find a point $\bar{x} = x + td$ in the neighborhood of $x$ such that $f(\bar{x}) < f(x)$.

# First-Order Optimality Conditions

## First-Order Necessary Conditions

If $x^*$ is a local minimizer of the unconstr. problem $\min_{x \in \mathbb{R}^n} f(x)$, then we must have $\nabla f(x^*) = 0$.

Remark:

- Points $x$ with $\nabla f(x)$ are all candidates for local minimizers.

Example: $f(x) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2$.

The FONC is:

$$2x_1 - x_2 = 0, \quad -x_1 + 2x_2 = 3.$$

There is a unique solution ($x_1 = 1$, $x_2 = 2$), which turns out to be the global minimizer of $f$.

Assume a variable $y$ is affected by $n$ factors $x_1, ..., x_n \in \mathbb{R}^m$. We know that they approximately have a linear relationship:

$$y \approx \beta_1 x_1 + \cdots + \beta_n x_n = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}.$$

Now, we want to specify this relationship (find parameters $\beta$).

- We have $m$ observations ($m > n$):

$$\{(x_{i1}, ..., x_{in}), y_i\}, \quad i = 1, ..., m.$$

- Ideally, we want to find $\beta = (\beta_1, ..., \beta_n)^\top$ such that $y = X\beta$.

$\rightsquigarrow$ However, this may not be possible (the equation $y = X\beta$ is an overdetermined linear system).

- Usually the observations do not follow $y = X\beta$ exactly $\rightsquigarrow$ noisy observations.

# Example: Least Squares – Continued

Instead, we try to minimize the sum of the squared errors:

$$\text{minimize}_\beta \quad \sum_{i=1}^{m} \left( y_i - \sum_{j=1}^{n} \beta_j x_{ij} \right)^2$$

The matrix form of this problem is:

$$\text{minimize}_\beta \quad ||X\beta - y||^2 = \beta^\top X^\top X\beta - 2\beta^\top X^\top y + y^\top y$$

where $||w||^2 = w^\top w = w_1^2 + \cdots + w_n^2$.

Facts:

- If $f(x) = x^\top Mx$ ($M$ is symmetric), then: $\nabla f(x) = 2Mx$.
- If $f(x) = c^\top x$, then $\nabla f(x) = c$.

Therefore, the FONC for the least squares problem is:

$$X^\top X\beta - X^\top y = 0.$$

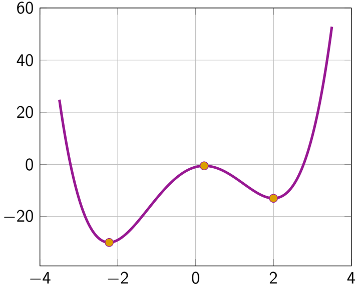
Solving this equation gives candidates for local minimizer.

# FONC is Not Sufficient

Example: $f(x) := x^4 - 9x^2 + 4x - 1$. The FONC is

$$f'(x) = 4x^3 - 18x + 4 = 0$$

with solutions $x_1 = -1 - \sqrt{6}/2$, $x_2 = -1 + \sqrt{6}/2$, and $x_3 = 2$.




We see that FONC is not sufficient!

- In fact, each local maximum also satisfies the FONC!
- Or it could be neither a local minimum nor maximum ($x^3$).

## Second-Order Necessary Conditions

Consider the Taylor expansion again but to the 2nd order (assuming $f$ is twice continuously differentiable):

$$f(x+td) = f(x) + t\nabla f(x)^\top d + \frac{1}{2}t^2 d^\top \nabla^2 f(x) d + o(t^2).$$

When the first-order necessary condition holds, we have:

$$f(x+td) = f(x) + \frac{1}{2}t^2 d^\top \nabla^2 f(x) d + o(t^2).$$

In order for $x$ to be a local minimizer, we also need $d^\top \nabla^2 f(x) d$ to be nonnegative for every $d \in \mathbb{R}^n$.

### Theorem: Second-Order Necessary Conditions

If $x^*$ is a local minimizer of $f$, then it holds that:

1. $\nabla f(x^*) = 0$;
2. For all $d \in \mathbb{R}^n$: $d^\top \nabla^2 f(x^*) d \geq 0$.

### Definition: Semidefiniteness

We call a (symmetric) matrix $A$ positive (negative) semidefinite (PSD/NSD) if and only if for all $x$ we have $x^\top A x \geq 0$ ($\leq 0$).

Remark:

- Therefore, the second-order necessary condition requires the Hessian matrix at $x^*$ to be PSD. In the one-dim. case, this is equivalent to $f''(x^*) \geq 0$.

Here are some useful facts about PSD matrices:

- We usually only talk about PSD properties for symmetric matrices.
- If a matrix $A$ is not symmetric, we use $\frac{1}{2}(A + A^\top)$ to define the PSD properties (because $x^\top Ax = \frac{1}{2}x^\top(A + A^\top)x$).
- A symmetric matrix is PSD if and only if all the eigenvalues are nonnegative.
- A symmetric matrix is PSD if and only if all the principal submatrices have nonnegative determinants.
- For any matrix $A$, $A^\top A$ is a (symmetric) PSD matrix.

# Example Continued

For $f(x) := x^4 - 9x^2 + 4x - 1$, the second-order condition is:

$$f''(x) = 12x^2 - 18 \geq 0$$

Only $x_1 = -1 - \sqrt{6}/2$ and $x_3 = 2$ satisfy the condition. But for the point $x_2 = -1 + \sqrt{6}/2$, we obtain $f''(x_2) = 12(1 - \sqrt{6}) < 0$ (thus, $x_2$ is not a local minimizer).

In the example of least squares problem, we use the following fact:

- If $f(x) = x^\top Mx$ ($M$ is symmetric), then $\nabla^2 f(x) = 2M$.

Therefore, the Hessian matrix in that problem is $2X^\top X$, which is always a PSD matrix. Therefore, the SONC always holds!

# SONC is Not Sufficient

However, even if both the first- and second-order necessary conditions hold, we still can not guarantee that the candidate is a local minimum!

Example: Consider $f(x) = x^3$ at 0.

- $f'(0) = f''(0) = 0$, thus FONC and SONC hold.
- But 0 is not a local minimum

- A point $x$ satisfying $\nabla f(x) = 0$ is called critical point or stationary point.
- The SONC can used to verify that a stationary point is **not** a local minimizer.

$\rightsquigarrow$ By modifying the SONC, we can get a sufficient condition.

# Second-Order Sufficient Conditions

# Second-Order Sufficient Condition (SOSC)

**Theorem: Second-Order Sufficient Conditions**

Let $f$ be twice continuously differentiable. If $x^*$ satisfies:

1. $\nabla f(x^*) = 0$;
2. For all $d \in \mathbb{R}^n \backslash \{0\}$: $d^\top \nabla^2 f(x^*) d > 0$;

then $x^*$ is a strict local minimum of $f$.

**Definition: Definite Matrices**

We call a (symmetric) matrix $A$ positive (negative) definite (PD/ ND) if and only if for all $x \neq 0$: $x^\top A x > 0$ $(< 0)$.

- A PD matrix must be PSD (thus PD is a stronger notion).
- A symmetric matrix is PD $\iff$ all its eigenvalues are positive.
- A symmetric matrix is PD $\iff$ the determinants of all leading principal submatrices are positive.

Our conditions are derived for minimization problems. For maximization problems, we just change the inequalities. Let $f \in C^2$.

**Theorem: FONC for Maximization**

If $x^*$ is a local (unconstrained) maximizer of $f$, then we must have $\nabla f(x^*) = 0$.

**Theorem: SONC for Maximization**

If $x^*$ is a local maximizer of $f$, then we must have 1.) $\nabla f(x^*) = 0$; 2.) $\nabla^2 f(x^*)$ is negative semidefinite.

**Theorem: SOSC for Maximization**

If $x^*$ satisfies 1.) $\nabla f(x^*) = 0$; 2.) $\nabla^2 f(x^*)$ is negative definite, then $x^*$ is a strict local maximizer.

## Optimality Conditions for Unconstrained Problems:

- First-order necessary condition.
- Second-order necessary condition.
- Second-order sufficient condition.

In many cases, we can utilize these conditions to identify local and global optimal solutions.

## General Strategy:

- Use FONC and SONC to identify all possible candidates. Then, use the sufficient conditions to verify.
- If a problem only has one stationary point and one can reason that the problem must have a finite optimal solution, then this point must be the (global) optimum.

Questions?