Lecture 17: Optimality Conditions

Zizhuo Wang

Institute of Data and Decision Analytics (iDDA)
Chinese University of Hong Kong, Shenzhen

Nov 7, 2018

Announcements

- Homework 6 posted, due next Wednesday (11/14)

Recap: Optimality Conditions for Unconstrained Problems

Theorem (First-Order Necessary Condition)

If $\mathbf{x}^*$ is a local minimizer of $f(\cdot)$ for an unconstrained problem, then we must have $\nabla f(\mathbf{x}^*) = 0$.

Theorem (Second-Order Necessary Condition)

If $\mathbf{x}^*$ is a local minimizer of $f(\cdot)$ for an unconstrained problem, then we must have

1. $\nabla f(\mathbf{x}^*) = 0$;
2. $\nabla^2 f(\mathbf{x}^*)$ *is positive semi-definite.*

- There are several ways to check positive semi-definiteness of a matrix
- These necessary conditions can be used to find candidates for local minimizers
- However, these conditions are not sufficient

Recap: Second-Order Sufficient Condition

Theorem

Let f be second-order continuously differentiable. If $\mathbf{x}^*$ satisfies:

1. $\nabla f(\mathbf{x}^*) = 0$
2. $\nabla^2 f(\mathbf{x}^*)$ *is positive definite*

Then $\mathbf{x}^*$ is a local minimizer of f for the unconstrained problem.

- Positive definite is equivalent as the eigenvalues are all strictly positive
- Can be used to verify candidate solutions
- May not be necessary ($f(x) = x^4$).

For Maximization Problems

Our conditions are derived for minimization problems.

- For maximization problems, it is just the opposite direction

Theorem (FONC for Maximization)

If $\mathbf{x}^*$ *is a local maximizer of* $f(\cdot)$ *for the unconstrained problem, then we must have* $\nabla f(\mathbf{x}^*) = 0$.

Theorem (SONC for Maximization)

If $\mathbf{x}^*$ *is a local maximizer of* $f(\cdot)$ *for an unconstrained problem, then we must have 1)* $\nabla f(\mathbf{x}^*) = 0$*; 2)* $\nabla^2 f(\mathbf{x}^*)$ *is negative semi-definite*

Theorem (SOSC for Maximization)

Let f *be second-order differentiable. If* $\mathbf{x}^*$ *satisfies 1)* $\nabla f(\mathbf{x}^*) = 0$*; 2)* $\nabla^2 f(\mathbf{x}^*)$ *is negative definite, then* $\mathbf{x}^*$ *is a local maximizer.*

Example

Find the minimum of

$$f(x, y, z) = 2x^2 + 2y^2 + z^2 - 2xy - 2xz - 6y + 7$$

We first look at the FONC. We have

$$\nabla f(x, y, z) = (4x - 2y - 2z, 4y - 2x - 6, 2z - 2x)$$

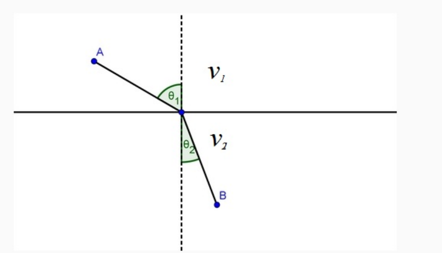
Solve $\nabla f(x, y, z) = 0$, get $x = y = z = 3$.

Now we consider the second-order sufficient condition.

$$\nabla^2 f = \begin{bmatrix} 4 & -2 & -2 \\ -2 & 4 & 0 \\ -2 & 0 & 2 \end{bmatrix}$$

The eigenvalues of this matrix are all positive. Therefore, $\nabla^2 f$ is positive definite and thus (3,3,3) is a local minimum (in fact it is also the global minimum)

Example: Reflection Principle of Lights



Light travels across two medium which separate at $y = 0$. It travels from $A = (0, a)$ to $B = (b, c)$. The speed in medium 1 is v_1 and the speed in medium 2 is v_2. The basic principle is that light will travel in shortest time.

- What path would the light take?

Example Continued

Suppose it crosses the two medium at $(x, 0)$. Then x must minimize

$$\frac{\sqrt{a^2 + x^2}}{v_1} + \frac{\sqrt{(x-b)^2 + c^2}}{v_2}$$

The first-order condition is:

$$\frac{1}{v_1}\frac{x}{\sqrt{x^2 + a^2}} + \frac{1}{v_2}\frac{x-b}{\sqrt{(x-b)^2 + c^2}} = 0$$

Therefore

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{x}{\sqrt{x^2 + a^2}} \Big/ \frac{b-x}{\sqrt{(x-b)^2 + c^2}} = \frac{v_1}{v_2}$$

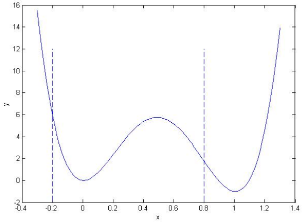
- Heron's reflection principle of light

Constrained Problems

We have derived necessary and sufficient conditions for the local minimum for unconstrained problems.

- What is the difference between constrained and unconstrained problems?

Consider the example $f(x) = 100x^2(1-x)^2 - x$ with constraint $-0.2 \le x \le 0.8$.



In addition to the original local minimizer ($x_1 = 0.013$), there is one more local minimizer on the boundary ($x = 0.8$).

Constrained Problems

At the boundary ($x^* = 0.8$), the FONC is not satisfied

$$f'(0.8) < 0$$

However, at this point, in order to stay feasible, we can only go leftward. That is, in the Taylor expansion

$$f(x^* + d) = f(x^*) + df'(x^*) + o(d)$$

we can only take d to be negative (otherwise it won't be feasible).

Thus $f(x^* + d) > f(x^*)$ in a small neighborhood of x^* in the feasible region. Thus x^* is a local minimizer.

Feasible Directions

Now we formalize the above arguments.

Definition (Feasible Direction)

Given $\mathbf{x} \in F$, we call $\mathbf{d}$ to be a *feasible direction* at $\mathbf{x}$ if there exists $\bar{\alpha} > 0$ such that $\mathbf{x} + \alpha\mathbf{d} \in F$ for all $0 \le \alpha \le \bar{\alpha}$.

For example,

- If $F = \{\mathbf{x} | A\mathbf{x} = \mathbf{b}\}$, then the feasible directions at $\mathbf{x}$ is $\{\mathbf{d} | A\mathbf{d} = 0\}$
- If $F = \{\mathbf{x} | A\mathbf{x} \ge \mathbf{b}\}$, then the feasible directions at $\mathbf{x}$ is $\{\mathbf{d} | \mathbf{a}_i^T\mathbf{d} \ge 0 \text{ if } \mathbf{a}_i^T\mathbf{x} = b_i\}$

FONC for Constrained Problems

Theorem (FONC for Constrained Problems)

If $\mathbf{x}^*$ is a local minimum, then for any feasible direction $\mathbf{d}$ at $\mathbf{x}^*$, we must have $\nabla f(\mathbf{x}^*)^T \mathbf{d} \geq 0$

Remark

In unconstrained problems, all directions are feasible, thus we must have $\nabla f(\mathbf{x}^*) = 0$.

An Alternative View

Definition (Descent Direction)

Let f be continuously differentiable. Then $\mathbf{d}$ is called a *descent direction* at $\mathbf{x}$ if and only if $\nabla f(\mathbf{x})^T\mathbf{d} < 0$.

Remark

If $\mathbf{d}$ is a descent direction at $\mathbf{x}$, then there exists $\bar{\gamma} > 0$ such that $f(\mathbf{x} + \gamma\mathbf{d}) < f(\mathbf{x})$ for all $0 < \gamma \leq \bar{\gamma}$.

If we denote the set of feasible directions at $\mathbf{x}$ by $S_F(\mathbf{x})$ and the set of descent directions at $\mathbf{x}$ by $S_D(\mathbf{x})$. Then the first order necessary condition can be written as:

$$S_F(\mathbf{x}^*) \cap S_D(\mathbf{x}^*) = \emptyset$$

Or in other words, there cannot be any feasible descent directions.

Nonlinear Optimization with Equality Constraints

Consider

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \end{aligned}$$

- The feasible direction set is $\{\mathbf{d} | A\mathbf{d} = 0\}$.
- The descent direction set is $\{\mathbf{d} | \nabla f(\mathbf{x})^T \mathbf{d} < 0\}$.

The FONC says that at local minimum, there cannot be a solution to both systems (feasible and descent direction)

Theorem (Alternative System)

The system $A\mathbf{d} = 0$ *and* $\nabla f(\mathbf{x})^T \mathbf{d} < 0$ *does not have a solution if and only if there exists* $\mathbf{y}$ *such that*

$$A^T \mathbf{y} = \nabla f(\mathbf{x})$$

Nonlinear Optimization with Equality Constraints

Therefore, the first-order necessary condition for

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \end{aligned} \tag{1}$$

is that there exists $\mathbf{y}$ such that

$$A^T\mathbf{y} = \nabla f(\mathbf{x})$$

Theorem

If $\mathbf{x}^*$ is a local minimum for (1), then there must exist $\mathbf{y}$ such that

$$A^T\mathbf{y} = \nabla f(\mathbf{x}^*)$$

Proof

First it is easy to see that if there exists $\mathbf{y}$ such that $A^T\mathbf{y} = \nabla f(\mathbf{x})$. Then we can't have a $\mathbf{d}$ such that $A\mathbf{d} = 0$ and $\nabla f(\mathbf{x})^T\mathbf{d} < 0$ (multiplying $\mathbf{d}^T$ to both sides of the equation will reach a contradiction).

To prove the reverse, consider the LP:

$$\begin{aligned} \text{minimize}_{\mathbf{d}} \quad & \nabla f(\mathbf{x})^T\mathbf{d} \\ \text{s.t.} \quad & A\mathbf{d} = 0 \end{aligned}$$

If there doesn't exist $\mathbf{d}$ satisfying $A\mathbf{d} = 0$ and $\nabla f(\mathbf{x})^T\mathbf{d} < 0$, then the optimal value of this LP must be 0.

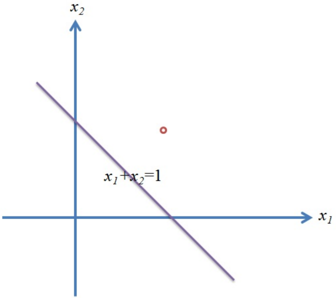
Therefore, by the strong duality theorem, its dual problem must also be feasible (and the optimal value is 0). However, the dual constraint is $A^T\mathbf{y} = \nabla f(\mathbf{x})$. Thus the theorem is proved. $\square$

Example

Consider the problem:

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

- This problem finds the nearest point on the line $x_1 + x_2 = 1$ to the point $(1, 1)$



Example Continued

By the FONC, $\mathbf{x} = (x_1, x_2)$ is a local minimizer if there exists y such that

$$A^T y = \nabla f(\mathbf{x})$$

Here $A = (1, 1)$. And $\nabla f(\mathbf{x}) = (2x_1 - 2; 2x_2 - 2)$.

Thus it means there exists y such that

$$2x_1 - 2 = y \quad 2x_2 - 2 = y$$

Also combined with the constraint $x_1 + x_2 = 1$. We have

$$x_1 = x_2 = 1/2$$

is the only candidate for local minimum. And it is indeed a local minimizer (also a global minimizer)

Another Example

Consider a constrained version of the least squares problem:

$$\begin{array}{rl} \text{minimize}_{\boldsymbol{\beta}} & \|X\boldsymbol{\beta} - \mathbf{y}\|_2^2 \\ \text{s.t.} & W\boldsymbol{\beta} = \xi \end{array}$$

The gradient is $2(X^T X\boldsymbol{\beta} - X^T\mathbf{y})$.

Therefore, the FONC is that there exists $\mathbf{z}$ such that

$$W^T\mathbf{z} = 2(X^T X\boldsymbol{\beta} - X^T\mathbf{y})$$

Therefore, an optimal $\boldsymbol{\beta}$ must satisfy:

$$W\boldsymbol{\beta} = \xi, \quad X^T X\boldsymbol{\beta} = \frac{1}{2}W^T\mathbf{z} + X^T\mathbf{y}$$

$$W\beta = \xi, \quad X^TX\beta = \frac{1}{2}W^T\mathbf{z} + X^T\mathbf{y}$$

We can write this as:

$$\begin{bmatrix} W & 0 \\ X^TX & -\frac{1}{2}W^T \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} \xi \\ X^T\mathbf{y} \end{bmatrix}$$

Here the size of X be $m \times n$, and the size of W be $d \times n$. Then these are $n + d$ linear equations with $n + d$ unknowns.

This is a system of linear equations with $n + d$ equations and $n + d$ unknowns. Solving this equation will yield the unique candidate for local minimizer (provided the left hand side matrix is of full rank).

Now we consider an inequality constrained problem:

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & A\mathbf{x} \geq \mathbf{b} \end{aligned} \tag{2}$$

What should be the necessary optimality conditions?

Theorem

If $\mathbf{x}^*$ is a local minimum of (2), then there exists some $\mathbf{y} \geq 0$ satisfying

$$\nabla f(\mathbf{x}^*) = A^T \mathbf{y}$$
$$y_i \cdot (\mathbf{a}_i^T \mathbf{x}^* - b_i) = 0, \quad \forall i$$

where $\mathbf{a}_i^T$ is the ith row of A.

Proof

We consider the descent directions and the feasible directions at $\mathbf{x}^*$. First it is easy to see that the descent directions are:

$$S_D(\mathbf{x}^*) = \{\mathbf{d} : \nabla f(\mathbf{x}^*)^T \mathbf{d} < 0\}$$

For the feasible directions, it is

$$S_F(\mathbf{x}^*) = \{\mathbf{d} : \mathbf{a}_i^T \mathbf{d} \geq 0, \text{ if } \mathbf{a}_i^T \mathbf{x}^* = b_i\}$$

Local optimality requires that $S_D(\mathbf{x}^*) \cap S_F(\mathbf{x}^*) = \emptyset$. We define $A(\mathbf{x}) = \{i : \mathbf{a}_i^T \mathbf{x} = b_i\}$ to be the *active constraints* at $\mathbf{x}$, then the necessary condition should be:

There does not exist $\mathbf{d}$ such that

1. $\nabla f(\mathbf{x}^*)^T \mathbf{d} < 0$
2. $\mathbf{a}_i^T \mathbf{d} \geq 0$ for $i \in A(\mathbf{x}^*)$

Proof Continued

The nonexistence of $\mathbf{d}$ such that

1. $\nabla f(\mathbf{x})^T \mathbf{d} < 0$
2. $\mathbf{a}_i^T \mathbf{d} \geq 0$ for $i \in A(\mathbf{x})$

is equivalent to the existence of $\mathbf{y} \geq 0$, such that

$$\nabla f(\mathbf{x}) = \sum_{i \in A(\mathbf{x})} \mathbf{a}_i y_i$$

This can be further written as the following conditions:

- There exists $\mathbf{y} \geq 0$ such that

$$\nabla f(\mathbf{x}) = A^T \mathbf{y}$$
$$y_i \cdot (\mathbf{a}_i^T \mathbf{x} - b_i) = 0, \quad \forall i$$

More General Cases — KKT Conditions

We have discussed cases with linear equality constraints or linear inequality constraints and derived the (necessary) optimality conditions

- We want to extend them to more general cases — KKT conditions
- We call the first-order necessary conditions for a general optimization problem the KKT conditions
- Solutions that satisfy the KKT conditions are called KKT points.
- KKT points are candidate points for local optimal solutions.
- The KKT conditions were originally named after H. Kuhn and A. Tucker, who first published the conditions in 1951. Later scholars discovered that the conditions had been stated by W. Karush in his master's thesis in 1939.

Find KKT Conditions

We consider the general nonlinear optimization problem:

$$\begin{array}{rll} \text{minimize}_{\mathbf{x}} & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) \geq 0 & i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 & i = 1, \ldots, p \\ & \ell_i(\mathbf{x}) \leq 0 & i = 1, \ldots, r \\ & x_i \geq 0 & i \in M \\ & x_i \leq 0 & i \in N \\ & x_i \text{ free} & i \notin M \cup N \end{array}$$

One can use the feasible/descent directions arguments to find the KKT conditions. But it is not very convenient.

- In the next lecture, we present a direct approach