Numerical Optimization

Graduate Course

Constrained smooth optimization

Part I: Optimality conditions

Wen Huang

School of Mathematical Sciences
Xiamen University

Compiled on March 15, 2022

Preliminaries and Basic Theories

- Nonlinear programming

$$
\begin{aligned}
&\min_x f(x) \\
c_i(x) &= 0, \quad i \in \mathcal{E} \\
c_i(x) &\geq 0, \quad i \in \mathcal{I}
\end{aligned}
$$

- f and c_i are smooth
- Optimality conditions

Sometimes, nonsmooth objective function or nonsmooth constraints can be reformulated into smooth constrained optimization problem

Sometimes, nonsmooth objective function or nonsmooth constraints can be reformulated into smooth constrained optimization problem

- Nonsmooth constraints:
  - Problem: $\min_{x \in \mathbb{R}^2} f(x)$ s.t. $|x_1| + |x_2| \leq 1$.
  - Constraints:

$$
\begin{aligned}
&|x_1| + |x_2| \leq 1 \Longrightarrow \\
&x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1
\end{aligned}
$$

Sometimes, nonsmooth objective function or nonsmooth constraints can be reformulated into smooth constrained optimization problem

- Nonsmooth constraints:
  - Problem: $\min_{x \in \mathbb{R}^2} f(x)$ s.t. $|x_1| + |x_2| \leq 1$.
  - Constraints:

$$|x_1| + |x_2| \leq 1 \Longrightarrow$$
$$x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1$$

- Objectives:
  - Problem: $\min_{x \in \mathbb{R}} f(x) = \max(x^2, x)$
  - Reformulate as:

$$\min_{(t,x) \in \mathbb{R}^2} t, \qquad \text{s.t. } t \geq x, t \geq x^2.$$

Example with an equality constraint

$$\min_x f(x) = x_1 + x_2 \qquad \text{such that } h(x) = 2 - x_1^2 - x_2^2 = 0$$

with global minimizer: $x^* = \begin{bmatrix} -1 & -1 \end{bmatrix}^T$

First order analysis:

- $\nabla f(x) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$, and $\nabla h(x) = \begin{bmatrix} -2x_1 & -2x_2 \end{bmatrix}^T$
- Feasibility: $0 = h(x+s) \approx h(x) + \nabla h^T(x)s$
- Descend: $0 > f(x+s) - f(x) \approx \nabla f^T(x)s$
- Find such an s?

Intuitions of First Order Optimality Conditions

Example 1

In case that $\nabla f(x)$ and $\nabla h(x)$ are not parallel

- Define $\tilde{s} = -\left(I - \frac{\nabla h(x) \nabla^T h(x)}{\|\nabla h(x)\|^2}\right) \nabla f(x) \neq 0$
- $\nabla h^T(x) \tilde{s} = 0$ and $\nabla f^T(x) \tilde{s} < 0$

In case that $\nabla f(x)$ and $\nabla h(x)$ are parallel

- No direction to move to the first order
- Parallel $\Longrightarrow$

$$\exists \lambda \in \mathbb{R}, \quad \nabla f(x) - \lambda \nabla h(x) = 0$$

- For this example ($\lambda^* = 1/2$):

$$\nabla f(x^*) - \lambda^* \nabla h(x^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- The sign of λ^* can be changed (along with the constraint)

- Lagrangian function: $\mathcal{L}(x, \lambda) = f(x) - \lambda h(x)$
- At x^*, optimality condition:

$$\exists \lambda^*, \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \lambda^* \nabla h(x^*) = 0$$

- Necessary, not sufficient:

$$\tilde{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T, \tilde{\lambda} = -\frac{1}{2}$$
$$\nabla_x \mathcal{L}(\tilde{x}, \tilde{\lambda}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- $\tilde{x}$ is a maximizer

Example with an equality constraint

$$\min_x f(x) = x_1 + x_2 \qquad \text{such that } g(x) = 2 - x_1^2 - x_2^2 \geq 0$$

with global minimizer: $x^* = \begin{bmatrix} -1 & -1 \end{bmatrix}^T$

First order analysis:

- $\nabla f(x) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$, and $\nabla g(x) = \begin{bmatrix} -2x_1 & -2x_2 \end{bmatrix}^T$
- Feasibility: $0 \leq g(x+s) \approx g(x) + \nabla g^T(x) s$
- Descend: $0 > f(x+s) - f(x) \approx \nabla f^T(x) s$
- Find such an s?

In case that x is an interior point, i.e., $g(x) > 0$

- No descent direction $\Longrightarrow$ any small p, $p^T \nabla f(x) \geq 0 \Longrightarrow \nabla f(x) = 0$
- Otherwise, $s = -\alpha \nabla f(x)$ for small α preserves feasibility and descend

---

In case that x is on the boundary, i.e., $g(x) = 0$

- s exists if $\nabla f^T(x) s < 0$ and $\nabla g^T(x) s \geq 0$
- Two half spaces no interection only if $\nabla f(x)$ and $\nabla g(x)$ in the same direction
- $\exists\ \lambda \geq 0$

$$\nabla f(x) - \lambda \nabla g(x) = 0$$

So if no first order feasible direction exists at x^* then for the two cases of this problem we have

$$\text{defining:} \quad \mathcal{L}(x, \lambda) = f(x) - \lambda g(x)$$
$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \lambda^* \nabla g(x^*) = 0, \quad \lambda^* g(x^*) = 0 \text{ and } \lambda^* \geq 0$$

- $\lambda^* g(x^*) = 0$ is called complementarity or complentary slackness
- $\lambda^* > 0 \to g(x^*) = 0$, i.e., $\lambda^* > 0$ only when $g(x^*)$ is active.
- Case 1: $g(x^*) > 0 \to \lambda^* = 0 \to \nabla f(x^*) = 0$
- Case 2: $g(x^*) = 0 \to \lambda^* \geq 0 \to \nabla f(x^*) - \lambda^* \nabla g(x^*) = 0$

- Geometry: easy to understand, difficult to use
- Algebra: easy to use, difficulty to understand

[Summary and Prediction]

Define the Lagrangian $\mathcal{L}(x,\lambda) = f(x) - \sum_{i\in\mathcal{E}} c_i(x)\lambda_i - \sum_{i\in\mathcal{I}} c_i(x)\lambda_i$, where λ_i are the Lagrange multiplier.
Conjecture: If x^* is a local minimizer, then

$$\nabla_x \mathcal{L}(x^*,\lambda^*) = 0$$

$$c_i(x^*) = 0 \quad i \in \mathcal{E}, \qquad c_i(x^*) \geq 0 \quad i \in \mathcal{I}$$

$$\lambda_i^* \geq 0 \quad i \in \mathcal{I}, \qquad \lambda_i^* c_i(x^*) = 0 \quad i \in \mathcal{I}.$$

No descent direction: Geometry

No descent direction: Geometry

x is a local minimizer if there does not exist a sequence $z_i \in \Omega$ such that $\lim_{i\to\infty} z_i = x$ and $f(z_i) < x$, where Ω denotes the feasible region.

Definition 1

The vector d is a tangent vector to the set $\mathcal{K}$ at x if there is a feasible sequence $\{z_k\}$ converging to x and sequence of positive scalars $\{\tau_k\}$, with $\tau_k \to 0$, such that

$$d = \lim_{k \to \infty} \frac{z_k - x}{\tau_k}$$

The collection of all tangent vectors at x denoted $T_{\mathcal{K}}(x)$.

Definition 2

- A set $S \subseteq \mathbb{R}^n$ is a cone if $x \in S \rightarrow \forall \alpha > 0, \;\; \alpha x \in S$.
- $T_{\mathcal{K}}(x)$ is called the tangent cone of $\mathcal{K}$ at x.

Lemma 3

Let $\{z_k\}$ be a feasible sequence converging to x and $\{\tau_k\}$ be the associated sequence of positive scalars , with $\tau_k \rightarrow 0$, used to define a tangent vector d to $\mathcal{K}$ at x. For any z_k, τ_k the following holds

$$z_k = x + \tau_k d + o(\tau_k).$$

Proof: This follows directly from the definition of the tangent vectors.

Lemma 4

If x^* is a local minimizer then

$$\forall d \in T_{\mathcal{K}}(x) \quad \nabla f^T(x^*)d \geq 0$$

First Order Optimality Conditions: Geometry

Alternative necessary condition

Definition 5

The normal cone at feasible point x of $f(x)$ is the set

$$N_{\mathcal{K}}(x) = \{v \mid \forall w \in T_{\mathcal{K}}(x) \;\; v^T w \leq 0\}.$$

That is, all vectors, v, such that its angle θ with any tangent vector satisfies $-\pi \leq \theta \leq -\pi/2$ or $\pi \geq \theta \geq \pi/2$.

Lemma 6

If x^* is a local minimizer then

$$-\nabla f(x^*) \in N_{\mathcal{K}}(x)$$

The optimality conditions from geometry are not easy to use

The optimality conditions from geometry are not easy to use

How to characterize the tangent cone/normal cone?

Definition 7

The active set $\mathcal{A}(x)$ at any feasible x consists of the equality constraint indices from $\mathcal{E}$ together with the indices of the inequality constraints i for which $c_i(x) = 0$; that is,

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} : c_i(x) = 0\}.$$

At a feasible point x, the inequality constraint $i \in \mathcal{I}$ is said to be active if $c_i(x) = 0$ and inactive if the strict inequality $c_i(x) > 0$ is satisfied.

Definition 8

Give a feasible point x, the set of linearized feasible directions $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d : \begin{array}{ll} d^T \nabla c_i(x) = 0, & \text{for all } i \in \mathcal{E} \\ d^T \nabla c_i(x) \geq 0, & \text{for all } i \in \mathcal{A} \cap \mathcal{I} \end{array} \right\}.$$

It is easy to verify that $\mathcal{F}(x)$ is a cone.

Definition 8

Give a feasible point x, the set of linearized feasible directions $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d : \begin{array}{ll} d^T \nabla c_i(x) = 0, & \text{for all } i \in \mathcal{E} \\ d^T \nabla c_i(x) \geq 0, & \text{for all } i \in \mathcal{A} \cap \mathcal{I} \end{array} \right\}.$$

It is easy to verify that $\mathcal{F}(x)$ is a cone.

Question: Is $\mathcal{F}(x)$ the same as $\mathrm{T}_\Omega(x)$?

Example 1:

$$c(x) = x_1^2 + x_2^2 - 1 = 0$$

- $x^* = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$
- $$\left. \begin{array}{c} \mathrm{T}_\Omega(x) = \{d \in \mathbb{R}^2 : d_1 = 0, d_2 \in \mathbb{R}\} \\ \mathcal{F}(x) = \{d \in \mathbb{R}^2 : d_1 = 0, d_2 \in \mathbb{R}\} \end{array} \right\} \implies \mathrm{T}_\Omega(x) = \mathcal{F}(x)$$

Example 2:

$$\tilde{c}(x) = (x_1^2 + x_2^2 - 1)^2 = 0$$

- $x^* = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$
- $$\left. \begin{array}{c} \mathrm{T}_\Omega(x) = \{d \in \mathbb{R}^2 : d_1 = 0, d_2 \in \mathbb{R}\} \\ \mathcal{F}(x) = \mathbb{R}^2 \end{array} \right\} \implies \mathrm{T}_\Omega(x) \neq \mathcal{F}(x)$$

Definition 9

Given a point x and active set $\mathcal{A}(x)$, x is said to be a regular point or equivalently the linear independence constraint qualification (LICQ) holds at x if the gradients of the active constraints are linearly independent.

Lemma 10

If x^* is a feasible point then

1. $\mathrm{T}_{\mathcal{K}}(x^*) \subseteq \mathcal{F}(x^*)$;
2. *and if x^* is a regular point (LICQ holds)* $\mathrm{T}_{\mathcal{K}}(x^*) = \mathcal{F}(x^*)$.

See Nocedal and Wright pp. 323 - 325.

Lemma 11 (Farkas's Lemma)

Define the cone K to be

$$K = \{x \in \mathbb{R}^n \mid x = By + Cw,\ B \in \mathbb{R}^{n\times m},\ \ C \in \mathbb{R}^{n\times p},\ \ y \geq 0\}$$

For any $g \in \mathbb{R}^n$ either

- $g \in K$;
- *or* $\exists d \in \mathbb{R}^n$ *such that*

$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0,$$

i.e., d defines a hyperplane that separates g and K,

but not both.

Theorem 12 (Karush-Kuhn-Tucker Conditions)

Suppose that x^* is a local minimizer, that the function f and c_i are continuously differentiable, and that the LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^*, with components $\lambda_i^*, i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied

$$
\begin{gathered}
\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \\
c_i(x^*) = 0 \quad i \in \mathcal{E}, \qquad\qquad c_i(x^*) \geq 0 \quad i \in \mathcal{I} \\
\lambda_i^* \geq 0 \quad i \in \mathcal{I}, \qquad\qquad \lambda_i^* c_i(x^*) = 0 \quad i \in \mathcal{I}.
\end{gathered}
$$

Lemma 13

Suppose at some $x^* \in \Omega$, all active constraints $c_i, i \in \mathcal{A}(x^*)$, are linear functions, then $\mathcal{F}(x^*) = \mathrm{T}_\Omega(x^*)$.

Other constraint qualifications (not discuss in details here)

- Mangasarian-Fromovitz constraint qualification (MFCQ)
- Abadie's constraint qualification (ACQ)
- Guinard constraint qualifications (GCQ)
- LICQ $\Longrightarrow$ MFCQ $\Longrightarrow$ ACQ $\Longrightarrow$ GCQ

Definition 14

Strict complementarity holds at x^* if either $\lambda_i^* = 0$ or $c_i(x^*) = 0$ but not both for all inequality constraints. In other words $\lambda_i^* > 0$ for any active inequality constraint.
An active inequality with $\lambda_i^* > 0$ is called nondegenerate; and one with $\lambda_i^* = 0$ is called degenerate.

The strict complementarity property usually makes it easier for algorithms to determine the active set $\mathcal{A}(x^*)$.

Lemma 15

If the KKT conditions and LICQ hold at a point x^*, then the vector λ^* of Lagrange multipliers is unique.

LICQ implies the linear independence of B and C. Therefore, the coefficients, Lagrange multipliers, are unique.

Theorem 16

Suppose that x^* is a local minimizer, that the function f and c_i are twice continuously differentiable, and that the LICQ holds at x^*. Let λ^* be the Lagrange multiplier vector for which the KKT conditions are satisfied. Then

$$w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w \geq 0, \forall w \in \mathcal{C}(x^*, \lambda^*),$$

where $\mathcal{C}(x^*, \lambda^*)$ is the critical cone
$\mathcal{C}(x^*, \lambda^*) = \{w \in \mathcal{F}(x^*) : \nabla c_i^T(x^*) w = 0, \text{ all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0\}.$

See detailed proofs in [NW06, P.332].

$w \in \mathcal{C}(x^*, \lambda^*) \implies w^T \nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* w^T \nabla c_i(x^*) = 0 \implies$ directions in $\mathcal{C}(x^*, \lambda^*)$ are unknown to be descent or ascent for f from only first order information

Theorem 17

Suppose that for some feasible point $x^* \in \mathbb{R}^n$ there is a Langrange multiplier vector λ^* such that the KKT conditions are satisfied. Suppose also that

$$w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w > 0, \forall w \in \mathcal{C}(x^*, \lambda^*), w \neq 0.$$

Then x^* is a strict local solution.

See detailed proofs in [NW06, P.333].

Note that LICQ is not needed here.

References I

J. Nocedal and S. J. Wright.
Numerical Optimization.
Springer, second edition, 2006.