Lecture 9: Optimization over a Convex Set

Shi Pu

School of Data Science (SDS)
The Chinese University of Hong Kong, Shenzhen

Outline

Outline

Throughout this lecture we will consider the constrained optimization problem (P) given by

$$(P) \qquad \begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in C \end{aligned}$$

- C - closed convex subset of $\mathbb{R}^n$.
- f - continuously differentiable[1] over C. Not necessarily convex.

Definition (Stationarity)

Let f be a continuously differentiable function over a closed and convex set C. Then $\mathbf{x}^*$ is called a stationary point of (P) if

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \text{ for all } \mathbf{x} \in C$$

[1]We use the convention that a function is differentiable over a given set D if it is differentiable over an open set containing D

Theorem

Let f be a continuously differentiable function over a nonempty closed convex set C, and let $\mathbf{x}^*$ be a local minimum of (P). Then $\mathbf{x}^*$ is a stationary point of (P).

Theorem

Let f be a continuously differentiable function over a nonempty closed convex set C, and let $\mathbf{x}^*$ be a local minimum of (P). Then $\mathbf{x}^*$ is a stationary point of (P).

Proof.

- Let $\mathbf{x}^*$ be a local minimum of (P), and assume in contradiction that $\mathbf{x}^*$ is not a stationary point of (P) $\Rightarrow$ there exists $\mathbf{x} \in C$ such that $\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) < 0$
- Thus, $f'(\mathbf{x}^*; \mathbf{d}) < 0$ where $\mathbf{d} = \mathbf{x} - \mathbf{x}^*$.
- Therefore $\exists \epsilon \in (0,1)$ s.t. $f(\mathbf{x}^* + t\mathbf{d}) < f(\mathbf{x}^*), \forall t \in (0, \epsilon)$.
- Since $\mathbf{x}^* + t\mathbf{d} = (1-t)\mathbf{x}^* + t\mathbf{x} \in C, \forall t \in (0, \epsilon)$, we conclude that $\mathbf{x}^*$ is *not* a local optimum point of (P). Contradiction.

$\square$

$C = \mathbb{R}^n$

- $\mathbf{x}^*$ is a stationary point of (P) iff

$$(\star) \quad \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$$

- We will show that the above condition is equivalent to $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Indeed, if $\nabla f(\mathbf{x}^*) = \mathbf{0}$. then obviously $(\star)$ is satisfied.
- Suppose that $(\star)$ holds.
- Plugging $\mathbf{x} = \mathbf{x}^* - \nabla f(\mathbf{x}^*)$ in the above implies $-\|\nabla f(\mathbf{x}^*)\|^2 \geq 0$.
- Thus, $\nabla f(\mathbf{x}^*) = \mathbf{0}$

$C = \mathbb{R}^n_+$

- $\mathbf{x}^* \in \mathbb{R}^n_+$ is a stationary point iff $\nabla f(\mathbf{x}^*)^\top(\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
- $\Leftrightarrow \nabla f(\mathbf{x}^*)^\top \mathbf{x} - \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
- $\Leftrightarrow \nabla f(\mathbf{x}^*) \geq \mathbf{0}$ and $\nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \leq 0$
- $\Leftrightarrow \nabla f(\mathbf{x}^*) \geq \mathbf{0}$ and $x_i^* \frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0, \quad i = 1, 2, \cdots, n$.
- $\Leftrightarrow$

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \begin{cases} = 0 & x_i^* > 0 \\ \geq 0 & x_i^* = 0 \end{cases}$$

feasible set	explicit stationarity condition
$\mathbb{R}^n$	$\nabla f(\mathbf{x}^*) = \mathbf{0}$
$\mathbb{R}^n_+$	$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \begin{cases} = 0 & x_i^* > 0 \\ \geq 0 & x_i^* = 0 \end{cases}$
$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{e}^\top \mathbf{x} = 1\}$	$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{x}^*)$
$B[\mathbf{0}, 1]$	$\nabla f(\mathbf{x}^*) = \mathbf{0}$ or $\\|\mathbf{x}^*\\| = 1$ and $\exists \lambda \leq 0 : \nabla f(\mathbf{x}^*) = \lambda \mathbf{x}^*$

For convex problems, stationarity is a necessary and sufficient condition

Theorem

Let f be a continuously differentiable convex function over a nonempty closed and convex set $C \subseteq \mathbb{R}^n$. Then $\mathbf{x}^*$ is a stationary point of

$$(P) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & \mathbf{x} \in C \end{array}$$

iff $\mathbf{x}^*$ is an optimal solution of (P).

Proof.

- If $\mathbf{x}^*$ is an optimal solution of (P), then we already showed that it is a stationary point of (P).
- Assume that $\mathbf{x}^*$ is a stationary point of (P).
- Let $\mathbf{x} \in C$. Then

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq f(\mathbf{x}^*).$$

- Establishing the optimality of $\mathbf{x}^*$.

$\square$

Outline

The Orthogonal Projection Operator

Definition

Given a nonempty closed convex set C, the orthogonal projection operator $P_C : \mathbb{R}^n \to C$ is defined by

$$P_C(\mathbf{x}) = \mathrm{argmin}\{\|\mathbf{y} - \mathbf{x}\|^2 : \mathbf{y} \in C\}.$$

The Second Projection Theorem

Theorem

Let C be a nonempty closed convex set and let $\mathbf{x} \in \mathbb{R}^n$. Then $\mathbf{z} = P_C(\mathbf{x})$ if and only if

$$(\mathbf{x} - \mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \leq 0 \text{ for any } \mathbf{y} \in C. \tag{1}$$

The Second Projection Theorem

Theorem

Let C be a nonempty closed convex set and let $\mathbf{x} \in \mathbb{R}^n$. Then $\mathbf{z} = P_C(\mathbf{x})$ if and only if

$$(\mathbf{x} - \mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \le 0 \text{ for any } \mathbf{y} \in C. \tag{1}$$

Proof.

- $\mathbf{z} = P_C(\mathbf{x})$ iff it is the optimal solution of the problem

$$\begin{aligned} \min \quad & g(\mathbf{y}) \equiv \|\mathbf{y} - \mathbf{x}\|^2 \\ \text{s.t.} \quad & \mathbf{y} \in C \end{aligned}$$

- By the previous theorem, $\mathbf{z} = P_C(\mathbf{x})$ if and only if

$$\nabla g(\mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \ge 0 \text{ for any } \mathbf{y} \in C$$

which is the same as (1).

□

Theorem

Let C be a nonempty closed and convex set.

1. *For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$:*

$$(P_C(\mathbf{v}) - P_C(\mathbf{w}))^\top (\mathbf{v} - \mathbf{w}) \geq \|P_C(\mathbf{v}) - P_C(\mathbf{w})\|^2. \tag{2}$$

2. **(non-expansiveness)** *For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$:*

$$\|P_C(\mathbf{v}) - P_C(\mathbf{w})\| \leq \|\mathbf{v} - \mathbf{w}\|. \tag{3}$$

Proof.

- For any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in C$: $(\mathbf{x} - P_C(\mathbf{x}))^\top(\mathbf{y} - P_C(\mathbf{x})) \leq 0$, $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in C$. Substituting $\mathbf{x} = \mathbf{v}$, $\mathbf{y} = P_C(\mathbf{w})$, we have

$$(\mathbf{v} - P_C(\mathbf{v}))^\top(P_C(\mathbf{w}) - P_C(\mathbf{v})) \leq 0. \tag{4}$$

- Now, by substituting $\mathbf{x} = \mathbf{w}, \mathbf{y} = P_C(\mathbf{v})$, we obtain

$$(\mathbf{w} - P_C(\mathbf{w}))^\top(P_C(\mathbf{v}) - P_C(\mathbf{w})) \leq 0. \tag{5}$$

Adding the two inequalities (4) and (5),

$$(P_C(\mathbf{w}) - P_C(\mathbf{v}))^\top(\mathbf{v} - \mathbf{w} + P_C(\mathbf{w}) - P_C(\mathbf{v})) \leq 0,$$

and hence, $(P_C(\mathbf{v}) - P_C(\mathbf{w}))^\top(\mathbf{v} - \mathbf{w}) \geq \|P_C(\mathbf{v}) - P_C(\mathbf{w})\|^2$.

Proof Contd.

- To prove (3), note that if $P_C(\mathbf{v}) = P_C(\mathbf{w})$, the inequality is trivial. Assume then that $P_C(\mathbf{v}) \neq P_C(\mathbf{w})$. By the Cauchy-Schwarz inequality we have

$$(P_C(\mathbf{v}) - P_C(\mathbf{w}))^\top(\mathbf{v} - \mathbf{w}) \leq \|P_C(\mathbf{v}) - P_C(\mathbf{w})\| \cdot \|\mathbf{v} - \mathbf{w}\|,$$

which combined with (2) yields the inequality

$$\|P_C(\mathbf{v}) - P_C(\mathbf{w})\| \cdot \|\mathbf{v} - \mathbf{w}\| \geq \|P_C(\mathbf{v}) - P_C(\mathbf{w})\|^2$$

Dividing by $\|P_C(\mathbf{v}) - P_C(\mathbf{w})\|$, implies (3).

□

Representation of Stationarity via the Orthogonal Projection Operator

Theorem

Let f be a continuously differentiable function over the nonempty closed convex set C, and let $s > 0$. Then $\mathbf{x}^*$ is a stationary point of

$$(P) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & \mathbf{x} \in C \end{array}$$

if and only if

$$\mathbf{x}^* = P_C(\mathbf{x}^* - s\nabla f(\mathbf{x}^*)).$$

Proof.

- By the second projection theorem, $\mathbf{x}^* = P_C(\mathbf{x}^* - s\nabla f(\mathbf{x}^*))$ iff

$$(\mathbf{x}^* - s\nabla f(\mathbf{x}^*) - \mathbf{x}^*)^\top(\mathbf{x} - \mathbf{x}^*) \leq 0 \text{ for any } \mathbf{x} \in C.$$

- Equivalent to

$$\nabla f(\mathbf{x}^*)^\top(\mathbf{x} - \mathbf{x}^*) \geq 0 \text{ for any } \mathbf{x} \in C,$$

namely to stationarity. □

Outline

The Gradient Mapping

- It is convenient to define the gradient mapping as

$$G_L(\mathbf{x}) = L\left[\mathbf{x} - P_C(\mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x}))\right]$$

  where $L > 0$.
- In the unconstrained case $G_L(\mathbf{x}) = \nabla f(\mathbf{x})$.
- $G_L(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x}$ is a stationary point of (P). This means that we can consider $\|G_L(\mathbf{x})\|^2$ to be optimality measure.

Algorithm 1 The Gradient Projection Method

1: **Input:** $\epsilon > 0$ - tolerance parameter.
2: **Initialization:** pick $\mathbf{x}_0 \in C$ arbitrarily.
3: **General step:**
4: **for** $k = 0, 1, 2, \cdots$ execute the following steps: **do**
5: pick a stepsize t_k by a line search procedure.
6: set $\mathbf{x}_{k+1} = P_C(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k))$.
7: if $\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \leq \epsilon$, then STOP and $\mathbf{x}_{k+1}$ is the output.
8: **end for**

- There are several strategies for choosing the stepsizes t_k.
- When $f \in C_L^{1,1}$, we can choose t_k to be constant and equal to $\frac{1}{L}$.

The Gradient Projection Method with Constant Stepsize

Algorithm 2 The Gradient Projection Method with Constant Stepsize

Input: $\epsilon > 0$ - tolerance parameter. $L > 0$ - an upper bound on the Lipschitz constant of ∇f.
Initialization: pick $\mathbf{x}_0 \in C$ arbitrarily. $\bar{t} > 0$ - constant stepsize.
General step:
for $k = 0, 1, 2, \cdots$ execute the following steps: **do**
    set $\mathbf{x}_{k+1} = P_C(\mathbf{x}_k - \bar{t}\nabla f(\mathbf{x}_k))$
    if $\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \leq \epsilon$, then STOP and $\mathbf{x}_{k+1}$ is the output.
end for

Algorithm 3 Gradient Projection Method with Backtracking

1: **Initialization:** Take $\mathbf{x}_0 \in C$ and $s > 0$, $\alpha \in (0,1)$, $\beta \in (0,1)$.
2: **General step:**
3: **for** $k \geq 1$ **do**
4: Pick $t_k = s$. Then, while

$$f(\mathbf{x}_k) - f(P_C(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k))) < \alpha t_k \|G_{\frac{1}{t_k}}(\mathbf{x}_k)\|^2$$

set $t_k := \beta t_k$.
5: Set $\mathbf{x}_{k+1} = P_C(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k))$
6: **end for**
7: **Stopping Criteria:** $\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \leq \epsilon$

Lemma (sufficient decrease lemma for constrained problems)

Suppose that $f \in C_L^{1,1}(C)$ *for some* $L > 0$, *where* C *is a closed convex set. Then for any* $\mathbf{x} \in C$ *and* $t \in (0, \frac{2}{L})$ *the following inequality holds:*

$$f(\mathbf{x}) - f(P_C(\mathbf{x} - t\nabla f(\mathbf{x}))) \geq t\left(1 - \frac{Lt}{2}\right)\left\|\frac{1}{t}(\mathbf{x} - P_C(\mathbf{x} - t\nabla f(\mathbf{x})))\right\|^2.$$

Proof. In class

Convergence of the Gradient Projection Method

Theorem

Let $\{\mathbf{x}_k\}$ be the sequence generated by the gradient projection method for solving problem (P) with either a constant stepsize $\bar{t} \in (0, \frac{2}{L})$, where L is a Lipschitz constant of ∇f or a backtracking stepsize strategy. Assume that f is bounded below. Then

1. *The sequence $\{f(\mathbf{x}_k)\}$ is nonincreasing.*
2. *$G_d(\mathbf{x}_k) \to 0$ as $k \to \infty$, where*

$$d = \begin{cases} 1/\bar{t} & \text{constant stepsize,} \\ 1/s & \text{backtracking.} \end{cases}$$

See the proof of Theorem 9.14 in the textbook.

- It is easy to see that this result implies that any limit point of the sequence is a stationary point of the problem.
- Rate of convergence of gradient mapping norms can be derived (similar to GD for unconstrained problem).

Theorem (rate of convergence of the sequence of function values)

Consider the problem

$$(P) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & \mathbf{x} \in C, \end{array}$$

where C is a nonempty closed and convex set, and $f \in C_L^{1,1}(C)$ is convex over C. Let $\{\mathbf{x}_k\}_{k \geq 0}$ be generated by GPM for solving (P) with a constant stepsize $t_k = \bar{t} \in (0, \frac{1}{L}]$. Assume the set of optimal solutions X^* is nonempty, and let f^* be the optimal value of (P). Then,

1. *for any $k \geq 0$ and $\mathbf{x}^* \in X^*$,*
$$2\bar{t}(f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*)) \leq \|\mathbf{x}_k - \mathbf{x}^*\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2,$$
2. *for any $n \geq 1$:*
$$f(\mathbf{x}_n) - f^* \leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|^2}{2\bar{t}n}.$$

Proof. In class

Theorem (convergence of the sequence generated by the gradient projection method)

Under the same setting of the previous theorem, the sequence $\{\mathbf{x}_k\}_{k\geq 0}$ *generated by the gradient projection method with a constant stepsize* $t_k = \bar{t} \in (0, \frac{1}{L}]$ *converges to an optimal solution.*

Proof. In class

Outline

The sparsity constrained problem is given by

$$(S) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \|\mathbf{x}\|_0 \le s, \end{array}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is a lower-bounded continuously differentiable function.
- $s > 0$ is an integer smaller than n
- $\|\mathbf{x}\|_0$ is the l_0 norm of $\mathbf{x}$, which counts the number of nonzero components in $\mathbf{x}$.
- We do not assume that f is a convex function. The constraint set is of course not convex.

Notation.

- $\mathbf{I}_1(\mathbf{x}) \equiv \{i : x_i \neq 0\}$ - the support set.
- $\mathbf{I}_0(\mathbf{x}) \equiv \{i : x_i = 0\}$ - the off-support set.
- $C_s = \{\mathbf{x} : \|\mathbf{x}\|_0 \le s\}$.
- For a vector $\mathbf{x} \in \mathbb{R}^n$ and $i \in \{1, 2, \cdots, n\}$, the i-th largest absolute value component in $\mathbf{x}$ is denoted by $M_i(\mathbf{x})$.

Definition

A vector $\mathbf{x}^* \in C_s$ is called a basic feasible (BF) vector of (P) if:

1. when $\|\mathbf{x}^*\|_0 < s$, $\nabla f(\mathbf{x}^*) = 0$;
2. when $\|\mathbf{x}^*\|_0 = s$, $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0$ for all $i \in \mathbf{I}_1(\mathbf{x}^*)$

A Fundamental Necessary Optimality Condition - Basic Feasibility

Theorem (BF is a necessary optimality condition)

Let $\mathbf{x}^*$ *be an optimal solution of (P). Then* $\mathbf{x}^*$ *is a BF vector.*

Proof.

- If $\|\mathbf{x}^*\|_0 < s$, then for any $i \in \{1, 2, \cdots, n\}$

$$0 \in \operatorname{argmin}\{g(t) \equiv f(\mathbf{x}^* + t\mathbf{e}_i)\}$$

  Otherwise there would exist a t_0 for which $f(\mathbf{x}^* + t\mathbf{e}_i) < f(\mathbf{x}^*)$, which is a contradiction to the optimality of $\mathbf{x}^*$.

- Therefore, we have $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = g'(0) = 0$
- If $\|\mathbf{x}^*\|_0 = s$ then the same argument holds for any $i \in \mathbf{I}_1(\mathbf{x}^*)$

$\square$

Definition

A vector $\mathbf{x}^* \in C_s$ is called an L-stationary point of (S) if it satisfies the relation

$$[NC_L] \qquad \mathbf{x}^* \in P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*))$$

- Note that since C_s is not a convex set, the orthogonal projection operator $P_{C_s}(\cdot)$ is not single-valued.
- Specifically, the members of $P_{C_s}(\mathbf{x})$ are vector consisting of the s components of $\mathbf{x}$ with the largest absolute value and zeros elsewhere.
- In general, there could be more than one choice to the s largest components. For example:

$$P_{C_2}((2,1,1)^\top) = \{(2,1,0)^\top, (2,0,1)^\top\}$$

Explicit Reformulation of L-stationarity

Lemma

For any $L > 0$, $\mathbf{x}^*$ satisfies [NC$_L$] if and only if $\|\mathbf{x}^*\|_0 \leq s$ and

$$\left|\frac{\partial f}{\partial x_i}(\mathbf{x}^*)\right| \begin{cases} \leq LM_s(\mathbf{x}^*) & \text{if } i \in \mathbf{I}_0(\mathbf{x}^*) \\ = 0 & \text{if } i \in \mathbf{I}_1(\mathbf{x}^*) \end{cases} \tag{6}$$

$[NC_L] \Rightarrow$ (6).

- Suppose that $\mathbf{x}^*$ satisfies $[NC_L]$. Note that for any index $j \in \{1, 2, \cdots, n\}$, the j-th component of $P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*))$ is either zero or equal to $\mathbf{x}_j^* - \frac{1}{L}\nabla_j f(\mathbf{x}^*)$.
- Since $\mathbf{x}^* \in P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*))$, it follows that if $i \in \mathbf{I}_1(\mathbf{x}^*)$, then $x_i^* = x_i^* - \frac{1}{L}\frac{\partial f}{\partial x_i}(\mathbf{x}^*)$, so that $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0$.
- If $i \in \mathbf{I}_0(\mathbf{x}^*)$, then $\left| x_i^* - \frac{1}{L}\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \leq M_s(\mathbf{x}^*)$, which combined with the fact that $x_i^* = 0$ implies that $\left| \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \leq L M_s(\mathbf{x}^*)$, and consequently (6) holds true.

$\square$

(6) $\Rightarrow$ [NC_L].

- Suppose that $\mathbf{x}^*$ satisfies (6). If $\|\mathbf{x}^*\|_0 < s$, then $M_s(\mathbf{x}^*) = 0$ and by (6) it follows that $\nabla f(\mathbf{x}^*) = 0$. Therefore, $P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*)) = P_{C_s}(\mathbf{x}^*) = \{\mathbf{x}^*\}$
- If $\|\mathbf{x}^*\|_0 = s$, then $M_s(\mathbf{x}^*) \neq 0$ and $|\mathbf{I}_1(\mathbf{x}^*)| = s$. By (6),

$$\left| x_i^* - \frac{1}{L}\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \begin{cases} = |x_i^*| & \text{if } i \in \mathbf{I}_1(\mathbf{x}^*) \\ \leq M_s(\mathbf{x}^*) & \text{if } i \in \mathbf{I}_0(\mathbf{x}^*) \end{cases}$$

- Therefore, the vector $\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*)$ contains the s components of $\mathbf{x}^*$ with the largest absolute value and all other components are smaller or equal to them, so that [NC_L] holds.

$\square$

Remark: Note that the condition [NC_L] depends on L in contrast to the stationarity condition over convex sets.

When $f \in C^{1,1}_{L_f}$, it is possible to show that an optimal solution of (S) is an L-stationary point for any $L > L(f)$.

Theorem

Suppose that $f \in C^{1,1}_{L_f} \in \mathbb{R}^n$, and that $L > L_f$. Let $\mathbf{x}^*$ be an optimal solution of (S). Then $\mathbf{x}^*$ is an L-stationary point.

See the proof of Theorem 9.22 in the textbook.

Algorithm 4 The IHT method

1: **Input:** a constant $L \geq L_f$.
2: **Initialization:** Choose $\mathbf{x}_0 \in C_s$
3: **General step:** $\mathbf{x}^{k+1} \in P_{C_s}(\mathbf{x}^k - \frac{1}{L}\nabla f(\mathbf{x}^k)), \quad (k = 0, 1, 2, \cdots)$

Theorem (convergence of IHT)

Suppose that $f \in C_{L_f}^{1,1}$ and let $\{\mathbf{x}^k\}_{k\geq 0}$ be the sequence generated by the IHT method with stepsize $\frac{1}{L}$ where $L > L_f$. Then any accumulation point of $\{\mathbf{x}^k\}_{k\geq 0}$ is an L-stationary point.

See the proof of Theorem 9.24 in the textbook.