

Lecture 9 - Optimization over a Convex Set

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

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

- ▶ C - closed convex subset of \mathbb{R}^n .
- ▶ f - continuously differentiable¹ over C . Not necessarily convex.

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

Lecture 9 - Optimization over a Convex Set

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

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

- ▶ C - closed convex subset of \mathbb{R}^n .
- ▶ f - continuously differentiable¹ over C . Not necessarily convex.

Definition of 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

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

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

Stationarity as a Necessary Optimality Condition

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.

Stationarity as a Necessary Optimality Condition

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 $\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) < 0$.

Stationarity as a Necessary Optimality Condition

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 $\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) < 0$.
- ▶ Thus, $f'(\mathbf{x}^*; \mathbf{d}) < 0$ where $\mathbf{d} = \mathbf{x} - \mathbf{x}^*$.

Stationarity as a Necessary Optimality Condition

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 $\nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) < 0$.
- ▶ Thus, $f'(\mathbf{x}^*; \mathbf{d}) < 0$ where $\mathbf{d} = \mathbf{x} - \mathbf{x}^*$.
- ▶ Therefore $\exists \varepsilon \in (0, 1)$ s.t. $f(\mathbf{x}^* + t\mathbf{d}) < f(\mathbf{x}^*) \forall t \in (0, \varepsilon)$.

Stationarity as a Necessary Optimality Condition

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 $\nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) < 0$.
- ▶ Thus, $f'(\mathbf{x}^*; \mathbf{d}) < 0$ where $\mathbf{d} = \mathbf{x} - \mathbf{x}^*$.
- ▶ Therefore $\exists \varepsilon \in (0, 1)$ s.t. $f(\mathbf{x}^* + t\mathbf{d}) < f(\mathbf{x}^*) \forall t \in (0, \varepsilon)$.
- ▶ Since $\mathbf{x}^* + t\mathbf{d} = (1 - t)\mathbf{x}^* + t\mathbf{x} \in C \forall t \in (0, \varepsilon)$, we conclude that \mathbf{x}^* is *not* a local optimum point of (P). Contradiction.

Examples

- ▶ $C = \mathbb{R}^n$.

Examples

- ▶ $C = \mathbb{R}^n$.
- ▶ \mathbf{x}^* is a stationary point of (P) iff

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

Examples

- ▶ $C = \mathbb{R}^n$.
- ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.

Examples

- ▶ $C = \mathbb{R}^n$.
- ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
- ▶ Suppose that (*) holds.

Examples

- ▶ $C = \mathbb{R}^n$.
- ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
- ▶ Suppose that (*) holds.
- ▶ Plugging $\mathbf{x} = \mathbf{x}^* - \nabla f(\mathbf{x}^*)$ in the above implies $-\|\nabla f(\mathbf{x}^*)\|^2 \geq 0$.

Examples

- ▶ $C = \mathbb{R}^n$.
- ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
- ▶ Suppose that (*) holds.
- ▶ Plugging $\mathbf{x} = \mathbf{x}^* - \nabla f(\mathbf{x}^*)$ in the above implies $-\|\nabla f(\mathbf{x}^*)\|^2 \geq 0$.
- ▶ Thus, $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Examples

- ▶ $C = \mathbb{R}^n$.
 - ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
 - ▶ Suppose that (*) holds.
 - ▶ Plugging $\mathbf{x} = \mathbf{x}^* - \nabla f(\mathbf{x}^*)$ in the above implies $-\|\nabla f(\mathbf{x}^*)\|^2 \geq 0$.
 - ▶ Thus, $\nabla f(\mathbf{x}^*) = \mathbf{0}$.
- ▶ $C = \mathbb{R}_+^n$.

Examples

- ▶ $C = \mathbb{R}^n$.
 - ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
 - ▶ Suppose that (*) 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 $\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.

Examples

- ▶ $C = \mathbb{R}^n$.
 - ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
 - ▶ Suppose that (*) 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 $\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*)^T \mathbf{x} - \nabla f(\mathbf{x}^*)^T \mathbf{x}^* \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.

Examples

- ▶ $C = \mathbb{R}^n$.
 - ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
 - ▶ Suppose that (*) 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 $\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*)^T \mathbf{x} - \nabla f(\mathbf{x}^*)^T \mathbf{x}^* \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*) \geq \mathbf{0}$ and $\nabla f(\mathbf{x}^*)^T \mathbf{x}^* \leq 0$.

Examples

- ▶ $C = \mathbb{R}^n$.
 - ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
 - ▶ Suppose that (*) 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 $\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*)^T \mathbf{x} - \nabla f(\mathbf{x}^*)^T \mathbf{x}^* \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*) \geq \mathbf{0}$ and $\nabla f(\mathbf{x}^*)^T \mathbf{x}^* \leq 0$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*) \geq \mathbf{0}$ and $x_i^* \frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, n$.

Examples

- ▶ $C = \mathbb{R}^n$.
 - ▶ \mathbf{x}^* is a stationary point of (P) iff

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

- ▶ 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 (*) is satisfied.
- ▶ Suppose that (*) 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 $\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*)^T \mathbf{x} - \nabla f(\mathbf{x}^*)^T \mathbf{x}^* \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*) \geq \mathbf{0}$ and $\nabla f(\mathbf{x}^*)^T \mathbf{x}^* \leq 0$.
 - ▶ $\Leftrightarrow \nabla f(\mathbf{x}^*) \geq \mathbf{0}$ and $x_i^* \frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, n$.
 - ▶ \Leftrightarrow

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

Explicit Stationarity Condition

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

Stationarity in Convex Optimization

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

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

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

Proof.

Stationarity in Convex Optimization

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

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

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).

Stationarity in Convex Optimization

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

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

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

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \geq f(\mathbf{x}^*),$$

Stationarity in Convex Optimization

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

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

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

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \geq f(\mathbf{x}^*),$$

- ▶ establishing the optimality of \mathbf{x}^* .

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

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

Proof.

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

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

Proof.

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

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

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

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

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

$$\nabla g(\mathbf{z})^T(\mathbf{y} - \mathbf{z}) \geq 0 \text{ for all } \mathbf{y} \in C,$$

which is the same as (1).

Properties of the Orthogonal Projection: (Firm) Nonexpansivness

Theorem. Let C be a nonempty closed and convex set. Then

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

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

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

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

Properties of the Orthogonal Projection: (Firm) Nonexpansiveness

Theorem. Let C be a nonempty closed and convex set. Then

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

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

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

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

Proof.

- ▶ For any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in C$:

$$(\mathbf{x} - P_C(\mathbf{x}))^T (\mathbf{y} - P_C(\mathbf{x})) \leq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in C \quad (4)$$

Substituting $\mathbf{x} = \mathbf{v}, \mathbf{y} = P_C(\mathbf{w})$, we have

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

Proof Contd.

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

$$(\mathbf{w} - P_C(\mathbf{w}))^T (P_C(\mathbf{v}) - P_C(\mathbf{w})) \leq 0. \quad (6)$$

Adding the two inequalities (5) and (6),

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

and hence,

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

Proof Contd.

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

$$(\mathbf{w} - P_C(\mathbf{w}))^T (P_C(\mathbf{v}) - P_C(\mathbf{w})) \leq 0. \quad (6)$$

Adding the two inequalities (5) and (6),

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

and hence,

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

- ▶ 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

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

which combined with (2) yields the inequality

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

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

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

if and only if

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

Proof.

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

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

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

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

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

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

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

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

- ▶ Equivalent to

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

namely to stationarity.

The Gradient Mapping

- ▶ It is convenient to define **the gradient mapping** as

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

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**.

The Gradient Projection Method

The Gradient Projection Method

Input: $\varepsilon > 0$ - tolerance parameter.

Initialization: pick $\mathbf{x}_0 \in C$ arbitrarily.

General step: for any $k = 0, 1, 2, \dots$ execute the following steps:

- (a) pick a stepsize t_k by a line search procedure.
- (b) set $\mathbf{x}_{k+1} = P_C(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k))$.
- (c) if $\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \leq \varepsilon$, then STOP and \mathbf{x}_{k+1} is the output.

- ▶ 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

The Gradient Projection Method with Constant Stepsize

Input: $\varepsilon > 0$ - tolerance parameter. $L > 0$ - an upper bound on the Lipschitz constant of ∇f .

Initialization: pick $\mathbf{x}_0 \in C$ arbitrarily. $\bar{\tau} > 0$ - constant stepsize.

General step: for any $k = 0, 1, 2, \dots$ execute the following steps:

- (a) set $\mathbf{x}_{k+1} = P_C(\mathbf{x}_k - \bar{\tau}\nabla f(\mathbf{x}_k))$.
- (b) if $\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \leq \varepsilon$, then STOP and \mathbf{x}_{k+1} is the output.

GPM with Backtracking

Gradient Projection Method with Backtracking

Initialization. Take $\mathbf{x}_0 \in C$ and $s > 0, \alpha \in (0, 1), \beta \in (0, 1)$.

General Step ($k \geq 1$)

- ▶ 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 $t_k := \beta t_k$.

- ▶ Set $\mathbf{x}_{k+1} = P_C(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k))$

Stopping Criteria $\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \leq \varepsilon$.

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. $G_d(\mathbf{x}_k) \rightarrow 0$ as $k \rightarrow \infty$, 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 book

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. $G_d(\mathbf{x}_k) \rightarrow 0$ as $k \rightarrow \infty$, 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 book

- ▶ It is easy to see that this result implies that any limit point of the sequence is a stationary point of the problem.
- ▶ When f is convex, it is possible to show that the sequence converges to a global optimal solution.

Sparsity Constrained Problems

The sparsity constrained problem is given by

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

- ▶ $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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.

Sparsity Constrained Problems

The sparsity constrained problem is given by

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

- ▶ $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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.

- ▶ $I_1(\mathbf{x}) \equiv \{i : x_i \neq 0\}$ - the support set.
- ▶ $I_0(\mathbf{x}) \equiv \{i : x_i = 0\}$ - the off-support set.
- ▶ $C_s = \{\mathbf{x} : \|\mathbf{x}\|_0 \leq s\}$.
- ▶ For a vector $\mathbf{x} \in \mathbb{R}^n$ and $i \in \{1, 2, \dots, n\}$, the i th largest absolute value component in \mathbf{x} is denoted by $M_i(\mathbf{x})$.

A Fundamental Necessary Optimality Condition - Basic Feasibility

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 $i \in I_1(\mathbf{x}^*)$.

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

Proof.

A Fundamental Necessary Optimality Condition - Basic Feasibility

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 $i \in I_1(\mathbf{x}^*)$.

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 $i \in \{1, 2, \dots, n\}$

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

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

A Fundamental Necessary Optimality Condition - Basic Feasibility

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 $i \in I_1(\mathbf{x}^*)$.

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 $i \in \{1, 2, \dots, n\}$

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

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

- ▶ Therefore, we have $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = g'(0) = 0$.

A Fundamental Necessary Optimality Condition - Basic Feasibility

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 $i \in I_1(\mathbf{x}^*)$.

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 $i \in \{1, 2, \dots, n\}$

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

Otherwise there would exist a t_0 for which $f(\mathbf{x}^* + t_0\mathbf{e}_i) < f(\mathbf{x}^*)$, 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 $i \in I_1(\mathbf{x}^*)$.

L-stationarity

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

$$[\text{NC}_L] \quad \mathbf{x}^* \in P_{C_S} \left(\mathbf{x}^* - \frac{1}{L} \nabla f(\mathbf{x}^*) \right).$$

- ▶ 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:

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

Explicit Reformulation of L -stationarity

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

$$\left| \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \begin{cases} \leq LM_s(\mathbf{x}^*) & \text{if } i \in I_0(\mathbf{x}^*), \\ = 0 & \text{if } i \in I_1(\mathbf{x}^*). \end{cases} \quad (7)$$

Proof.

Explicit Reformulation of L -stationarity

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

$$\left| \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \begin{cases} \leq LM_s(\mathbf{x}^*) & \text{if } i \in I_0(\mathbf{x}^*), \\ = 0 & \text{if } i \in I_1(\mathbf{x}^*). \end{cases} \quad (7)$$

Proof. ($[\text{NC}_L] \Rightarrow (7)$).

- ▶ Suppose that \mathbf{x}^* satisfies $[\text{NC}_L]$. Note that for any index $j \in \{1, 2, \dots, n\}$, the j -th component of $P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*))$ is either zero or equal to $x_j^* - \frac{1}{L}\nabla_j f(\mathbf{x}^*)$.

Explicit Reformulation of L -stationarity

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

$$\left| \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \begin{cases} \leq LM_s(\mathbf{x}^*) & \text{if } i \in I_0(\mathbf{x}^*), \\ = 0 & \text{if } i \in I_1(\mathbf{x}^*). \end{cases} \quad (7)$$

Proof. ($[\text{NC}_L] \Rightarrow (7)$).

- ▶ Suppose that \mathbf{x}^* satisfies $[\text{NC}_L]$. Note that for any index $j \in \{1, 2, \dots, n\}$, the j -th component of $P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*))$ is either zero or equal to $x_j^* - \frac{1}{L}\nabla_j f(\mathbf{x}^*)$.
- ▶ Since $\mathbf{x}^* \in P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*))$, it follows that if $i \in I_1(\mathbf{x}^*)$, then $x_i^* = x_i^* - \frac{1}{L}\frac{\partial f}{\partial x_i}(\mathbf{x}^*)$, so that $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0$.

Explicit Reformulation of L -stationarity

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

$$\left| \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \begin{cases} \leq LM_s(\mathbf{x}^*) & \text{if } i \in I_0(\mathbf{x}^*), \\ = 0 & \text{if } i \in I_1(\mathbf{x}^*). \end{cases} \quad (7)$$

Proof. ($[\text{NC}_L] \Rightarrow (7)$).

- ▶ Suppose that \mathbf{x}^* satisfies $[\text{NC}_L]$. Note that for any index $j \in \{1, 2, \dots, n\}$, the j -th component of $P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*))$ is either zero or equal to $x_j^* - \frac{1}{L}\nabla_j f(\mathbf{x}^*)$.
- ▶ Since $\mathbf{x}^* \in P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*))$, it follows that if $i \in I_1(\mathbf{x}^*)$, 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 $i \in I_0(\mathbf{x}^*)$, 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 $\left| \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \leq LM_s(\mathbf{x}^*)$, and consequently (7) holds true.

Proof Contd.

((7) \Rightarrow [NC_L]).

- ▶ Suppose that \mathbf{x}^* satisfies (7). If $\|\mathbf{x}^*\|_0 < s$, then $M_s(\mathbf{x}^*) = 0$ and by (7) it follows that $\nabla f(\mathbf{x}^*) = 0$. Therefore, $P_{C_s}(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*)) = P_{C_s}(\mathbf{x}^*) = \{\mathbf{x}^*\}$.

Proof Contd.

((7) \Rightarrow [NC_L]).

- ▶ Suppose that \mathbf{x}^* satisfies (7). If $\|\mathbf{x}^*\|_0 < s$, then $M_s(\mathbf{x}^*) = 0$ and by (7) 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 $|I_1(\mathbf{x}^*)| = s$. By (7)

$$\left| x_i^* - 1/L \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \begin{cases} = |x_i^*| & i \in I_1(\mathbf{x}^*) \\ \leq M_s(\mathbf{x}^*) & i \in I_0(\mathbf{x}^*). \end{cases}$$

Proof Contd.

((7) \Rightarrow [NC_L]).

- ▶ Suppose that \mathbf{x}^* satisfies (7). If $\|\mathbf{x}^*\|_0 < s$, then $M_s(\mathbf{x}^*) = 0$ and by (7) 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 $|I_1(\mathbf{x}^*)| = s$. By (7)
$$\left| x_i^* - \frac{1}{L} \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \begin{cases} = |x_i^*| & i \in I_1(\mathbf{x}^*) \\ \leq M_s(\mathbf{x}^*) & i \in I_0(\mathbf{x}^*). \end{cases}$$
- ▶ 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.

Proof Contd.

((7) \Rightarrow [NC_L]).

- ▶ Suppose that \mathbf{x}^* satisfies (7). If $\|\mathbf{x}^*\|_0 < s$, then $M_s(\mathbf{x}^*) = 0$ and by (7) 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 $|I_1(\mathbf{x}^*)| = s$. By (7)
$$\left| x_i^* - 1/L \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \right| \begin{cases} = |x_i^*| & i \in I_1(\mathbf{x}^*) \\ \leq M_s(\mathbf{x}^*) & i \in I_0(\mathbf{x}^*). \end{cases}$$
- ▶ 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.

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

L -Stationarity as a Necessary Optimality Condition

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

Theorem. Suppose that $f \in C_{L(f)}^{1,1} \in \mathbb{R}^n$, 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 book.

The Iterative Hard-Thresholding (IHT) Method

The IHT method

Input: a constant $L \geq L(f)$.

- **Initialization:** Choose $\mathbf{x}_0 \in C_S$.
- **General step :** $\mathbf{x}^{k+1} \in P_{C_S}(\mathbf{x}^k - \frac{1}{L}\nabla f(\mathbf{x}^k))$, $(k = 0, 1, 2, \dots)$

The Iterative Hard-Thresholding (IHT) Method

The IHT method

Input: a constant $L \geq L(f)$.

- **Initialization:** Choose $\mathbf{x}_0 \in C_S$.
- **General step :** $\mathbf{x}^{k+1} \in P_{C_S}(\mathbf{x}^k - \frac{1}{L}\nabla f(\mathbf{x}^k))$, $(k = 0, 1, 2, \dots)$

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 book