

Convergence Analysis of SGD in Convex Optimization Problems

by Hossain Rahnama

Sep 13, 2026

• یادگیری از نگاه بهینه‌سازی

$$\begin{array}{llll}
 \min_{h \in \mathcal{H}} R(h) := \mathbb{E}(\underbrace{\ell(h, (x, y))}_{\text{Loss Function}}) & \xrightarrow{\text{Hypothesis Space}} & \min_{h \in \mathcal{H}} R(h) & \xrightarrow{\text{Data}} \min_{h \in \mathcal{H}} \Phi(h) \\
 & & & \Phi(h) = \underbrace{\frac{1}{N} \sum_{i=1}^N \ell(h, (x_i, y_i))}_{\text{Sample}} + \underbrace{\lambda \Omega(h)}_{\text{Regularization}}
 \end{array}$$

الگوریتم‌های جستجوی خطی

$$\min_{x \in \mathbb{R}^d} f(x)$$

الگوریتم ۱.۴ جستجوی خطی.

ورودی: تابع $f: \mathbb{R}^d \rightarrow \mathbb{R}$

خروجی: تقریبی از یک کمینه‌کننده‌ی موضعی یا سراسری f روی $\mathbb{R}^d$ و مقدار کمینه‌ی متناظر آن.

۱: روال LINE-SEARCH

۲: * مقداردهی اولیه

۳: نقطه‌ی آغازین w_1 ، تعداد تکرار T و پارامترهای لازم دیگر را انتخاب کن.

۴: * حلقه‌ی همگرایی

۵: برای $t = 1, \dots, T$ انجام بده

۶: جهت مناسب d_t را با توجه به f انتخاب کن.

۷: طول گام مناسب s_t را با توجه به f انتخاب کن. دو برابر

۸: قرار بده $w_{t+1} = w_t - s_t d_t$.

۹: * محاسبه‌ی خروجی

۱۰: قرار بده $\bar{w} = \text{out}(w_1, \dots, w_T)$.

۱۱: برگردان $(\bar{w}, f(\bar{w}))$.

• دست‌نویس: (دقیق و ناراضی)، بکنوا و ناکنوا

• تابع out

• نوع حلقه‌ی همگرایی

• نمونه‌ها؟ گسردیان کاهش، گسردیان مزدوج، مختصات کاهش و نسخه‌های تعادلی آن‌ها.

الگوریتم ۲.۴ گرایان کاهشی تصادفی با طول گام ثابت.

- ورودی: تابع محدب $f: \mathbb{R}^d \rightarrow \mathbb{R}$.
خروجی: تقریبی از یک کمینه‌کننده‌ی سراسری f روی $\mathbb{R}^d$ و مقدار کمینه‌ی متناظر آن.
- ۱: **روال** STOCHASTIC-GRADIENT-DESCENT:
 - ۲: * مقداردهی اولیه
 - ۳: نقطه‌ی آغازین w_1 ، تعداد تکرار T و طول گام ثابت $s > 0$ را انتخاب کن.
 - ۴: * حلقه‌ی همگرایی
 - ۵: برای $t = 1, \dots, T$ انجام بده
 - ۶: جهت مناسب d_t را از توزیع $\mathbb{D}_t \in \text{Pb}(\mathbb{R}^d)$ به گونه‌ای انتخاب کن که $\mathbb{E}(d_t | w_t) \in \partial f(w_t)$.
 - ۷: قرار بده $w_{t+1} = w_t - s d_t$.
 - ۸: * محاسبه‌ی خروجی
 - ۹: قرار بده $\bar{w} = \frac{1}{T} \sum_{t=1}^T w_t$.
 - ۱۰: برگردان $(\bar{w}, f(\bar{w}))$.

$$\mathbb{E}(d_t | w_t) = g_t = \nabla f(w_t)$$

- میانگین جهت تصادفی برابر گرایان باشد. تصمیم؟
- معیار همگرایی؟

گزاره ۱۰.۲۸ در ضمیمه را ببینید.

$$\mathbb{E}(\|d_t | w_t\|) \text{ و } \mathbb{E}(f(w_t) - f(w^*)) \text{ و } \mathbb{E}(\|w_t - w^*\|)$$

موردی بر آنالیز محدب

تعریف ۱۶.۱ (زیرگرادیان و زیردیفرانسیل^۲). فرض کنید $A \subseteq \mathbb{R}^n$ و $f: A \rightarrow \mathbb{R}$ یک تابع دلخواه باشد و $x \in A$ بردار $v \in \mathbb{R}^n$ را یک زیرگرادیان تابع f در نقطه‌ی x گوئیم، اگر برای هر $y \in A$ داشته باشیم:

$$f(y) \geq f(x) + v^T(y - x).$$

مجموعه‌ی تمامی زیرگرادیان‌های تابع f در نقطه‌ی x را با $\partial f(x)$ نمایش می‌دهیم و آن را زیردیفرانسیل تابع در این نقطه می‌نامیم. همچنین، تعریف می‌کنیم $\text{dom}(\partial f) := \{x \in A \mid \partial f(x) \neq \emptyset\}$ که مجموعه‌ی تمام نقاطی است که زیردیفرانسیل ناتهی دارند.

مثال ۵.۱. اگر $C \subseteq \mathbb{R}^n$ یک مجموعه‌ی محدب باز و $f: C \rightarrow \mathbb{R}$ یک تابع محدب مشتق‌پذیر در $x \in C$ باشد، با توجه به بخش دوم گزاره ۱۴.۱ می‌دانیم که $\nabla f(x)$ یک زیرگرادیان در نقطه‌ی x است.

گزاره ۲۷.۱. فرض کنید $A \subseteq \mathbb{R}^n$ و $f: A \rightarrow \mathbb{R}$ یک تابع دلخواه باشد و $x \in \text{int}(A)$ نقطه‌ای است که $\partial f(x) \neq \emptyset$. اگر تابع f در x مشتق‌پذیر باشد، آن‌گاه $\partial f(x) = \{\nabla f(x)\}$.

گزاره ۱۴.۱. فرض کنید $C \subseteq \mathbb{R}^n$ یک مجموعه‌ی محدب باز و تابع $f: C \rightarrow \mathbb{R}$ روی C مشتق‌پذیر باشد. در این صورت گزاره‌های زیر معادل‌اند.

- تابع f محدب است.
- گراف f همواره در یک سمت صفحه‌ی مماس بر آن قرار می‌گیرد، یعنی برای هر $x, y \in C$ داریم:

$$f(x) \geq f(y) + \nabla f(y)^T(x - y).$$

- گرایان شرط یکنوایی را ارضا می‌کند، یعنی برای هر $x, y \in C$ داریم:

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq 0.$$

مروری بر آنالیز محدب

تعریف ۱۷.۱ (تابع لیپشیتز^۱). مجموعه‌ی $A \subseteq \mathbb{R}^n$ و تابع $f: A \rightarrow \mathbb{R}$ را در نظر بگیرید. تابع f روی A لیپشیتز است اگر عدد حقیقی $\rho > 0$ وجود داشته باشد به طوری که برای هر $x, y \in A$ داشته باشیم $|f(y) - f(x)| \leq \rho \|y - x\|$. عدد ρ را ثابت لیپشیتز می‌نامند. تابع f روی A لیپشیتز موضعی^۲ است اگر برای هر $x \in A$ عدد $\delta > 0$ وجود داشته باشد که تابع f روی گوی باز $B_\delta(x)$ لیپشیتز باشد.

- مثال. تابع خطی $f(x) = ax$ روی $\mathbb{R}$. تابع $f(x) = x^2$ روی $\mathbb{R}$. روی $[1, 1]$ ؟
- ارتباط با نیمه‌محدب.

گزاره ۳۳.۱ (ویژگی‌های اساسی توابع محدب). فرض کنید $C \subseteq \mathbb{R}^n$ یک مجموعه‌ی محدب باز و $f: C \rightarrow \mathbb{R}$ تابعی محدب است.

- برای هر $x \in C$ زیردیفرانسیل $\partial f(x)$ ناتهی، محدب و فشرده است.
- تابع f روی C به طور موضعی لیپشیتز است.
- برای هر $x \in C$ ثابت‌های $\rho > 0$ و $\delta > 0$ وجود دارند به طوری که برای هر $y \in B_\delta(x)$ داریم $\partial f(y) \subseteq B_\rho(0)$. اینجا ρ همان ثابت مربوط به لیپشیتز موضعی بودن تابع f در نقطه‌ی x است.

• تابع محدب است اگر و تنها اگر...؟

مروری بر آنالیز محدب

گزاره ۳.۴. مجموعه‌ی محدب $C \subseteq \mathbb{R}^n$ و تابع محدب $f: C \rightarrow \mathbb{R}$ را در نظر بگیرید. f روی C تابعی ρ -لیپشیتز است اگر و تنها اگر برای هر $x \in C$ داشته باشیم $\partial f(x) \subseteq B_\rho(0)$.

برهان. فرض کنید f تابعی ρ -لیپشیتز باشد.

$$\begin{aligned} v \in \partial f(x) &\Rightarrow \forall y \in C, f(y) \geq f(x) + v^T(y-x) \\ &\Rightarrow v^T(y-x) \leq f(y) - f(x) \leq \rho \|y-x\| \end{aligned}$$

$$\exists \alpha > 0, y = x + \alpha v \in C \Rightarrow \alpha v^T v \leq \alpha \rho \|v\| \Rightarrow \|v\| \leq \rho \Rightarrow v \in B_\rho(0)$$

باز بودن C

برعکس فرض کنید که زیرگرادیان کران دار است.

$$\begin{aligned} v \in \partial f(x) &\Rightarrow \forall y \in C, f(y) \geq f(x) + v^T(y-x) \\ &\Rightarrow f(x) - f(y) \leq v^T(x-y) \leq \|v\| \|x-y\| \leq \rho \|x-y\| \end{aligned}$$

$$\begin{aligned} w \in \partial f(y) &\Rightarrow \forall x \in C, f(x) \geq f(y) + w^T(x-y) \\ &\Rightarrow f(y) - f(x) \leq w^T(y-x) \leq \|w\| \|y-x\| \leq \rho \|y-x\| \end{aligned}$$

تصغیر برای توابع محدب لیپشیتز

۱.۴. فرض کنید $f: \mathbb{R}^d \rightarrow \mathbb{R}$ تابعی محدب باشد و $w^* \in \arg \min(f, \mathbb{R}^d)$. دنباله‌ی نقاط تصادفی $(w_1, \dots, w_T)$ با $\bar{w} = 1/T \sum_{t=1}^T w_t$ و جهت‌های تصادفی $(d_1, \dots, d_T)$ را در نظر بگیرید به طوری که برای هر $t \in [T]$ داشته باشیم $\mathbb{E}(d_t | w_t) \in \partial f(w_t)$ در این صورت، داریم:

$$\mathbb{E}(f(\bar{w}) - f(w^*)) \leq \frac{1}{T} \sum_{t=1}^T \mathbb{E}((w_t - w^*)^\top d_t).$$

$$\mathbb{E}\left(f\left(\frac{1}{T} \sum_{t=1}^T w_t\right) - f(w^*)\right) \leq \mathbb{E}\left(\sum_{t=1}^T \frac{1}{T} f(w_t) - f(w^*)\right) = \frac{1}{T} \sum_{t=1}^T \mathbb{E}(f(w_t) - f(w^*)) \quad \text{برگزن.}$$

$$f(w^*) \geq f(w_t) + \mathbb{E}(d_t | w_t)^\top (w^* - w_t) \implies \mathbb{E}(f(w_t) - f(w^*)) \leq \mathbb{E}(\mathbb{E}(d_t | w_t)^\top (w_t - w^*))$$

$$\mathbb{E}(f(w_t) - f(w^*)) \leq \mathbb{E}(\mathbb{E}((w_t - w^*)^\top d_t | w_t)) = \mathbb{E}((w_t - w^*)^\top d_t)$$

تصغیر برای توابع محدب لیپشیتز

تعریف ۱.۵. فرض کنید $x: \Omega \rightarrow \mathbb{R}^n$ یک بردار تصادفی با $\mathbb{E}(x) = m$ است. ماتریس کوواریانس متناظر با x را به صورت زیر تعریف کنید:

$$\text{cov}(x) := \mathbb{E}((x - m)(x - m)^\top).$$

درایه‌ی (i, j) در این ماتریس $\text{cov}(x_i, x_j) := \mathbb{E}((x_i - m_i)(x_j - m_j))$ است و بنابراین عناصر روی قطر آن واریانس درایه‌های بردار x ، یعنی $\text{var}(x_i)$ هستند. جمع عناصر قطری این ماتریس را به صورت زیر نمایش دهید:

$$\text{tr}(\text{cov}(x)) := \sum_{i=1}^n \text{var}(x_i).$$

لم ۸.۵. فرض کنید $x: \Omega \rightarrow \mathbb{R}^n$ یک بردار تصادفی است. در این صورت، $\text{tr}(\text{cov}(x)) = \mathbb{E}(\|x\|^2) - \|\mathbb{E}(x)\|^2$.

$$\begin{aligned} \text{tr}(\text{cov}(x)) &= \sum_{i=1}^n \text{var}(x_i) = \sum_{i=1}^n (\mathbb{E}(x_i^2) - \mathbb{E}(x_i)^2) = \sum_{i=1}^n \mathbb{E}(x_i^2) - \sum_{i=1}^n \mathbb{E}(x_i)^2 \\ &= \mathbb{E}\left(\sum_{i=1}^n x_i^2\right) - \sum_{i=1}^n (\mathbb{E}(x_i))^2 = \mathbb{E}(\|x\|^2) - \|\mathbb{E}(x)\|^2 \quad \text{برگزن.} \end{aligned}$$

همگرایی برای توابع محدب لیسپشیتز

لم ۲.۴. دنباله‌ی جهت‌های تصادفی $(d_1, \dots, d_T)$ و طول گام‌های $(s_1, \dots, s_T)$ را در نظر بگیرید که برای هر $t \in [T]$ داریم $s_t > 0$. از روی آن‌ها دنباله‌ی نقاط تصادفی $(w_1, \dots, w_T)$ را می‌سازیم به طوری که برای هر $t \in [T]$ داریم $w_{t+1} = w_t - s_t d_t$. در این صورت نامساوی زیر برقرار است:

$$\mathbb{E}((w_t - w^*)^\top d_t) \leq \frac{1}{2s_t} \mathbb{E}(\|w_t - w^*\|^2 - \|w_{t+1} - w^*\|^2) + \frac{s_t}{2} (\text{tr}(\text{cov}(d_t)) + \mathbb{E}(\|d_t\|^2)).$$

برهان.

$$2a^\top b = \|a\|^2 + \|b\|^2 - \|a-b\|^2$$

$$(w_t - w^*)^\top d_t = \frac{1}{s_t} (w_t - w^*)^\top (s_t d_t) = \frac{1}{2s_t} (\|w_t - w^*\|^2 + \|s_t d_t\|^2 - \|w_t - w^* - s_t d_t\|^2)$$

$$\mathbb{E}((w_t - w^*)^\top d_t) = \frac{1}{2s_t} \mathbb{E}(\|w_t - w^*\|^2 - \|w_{t+1} - w^*\|^2) + \frac{s_t}{2} \mathbb{E}(\|d_t\|^2)$$

می‌توان فرض کرد $\mathbb{E}(\|d_t\|^2) \leq \frac{rc}{\sqrt{T}}$ و شرط لیسپشیتز را حذف کرد

همگرایی برای توابع محدب لیسپشیتز

گزاره ۴.۴. فرض کنید $f: \mathbb{R}^d \rightarrow \mathbb{R}$ تابعی محدب و ρ -لیسپشیتز است. همچنین، بپذیرید $w^* \in \arg \min(f, \mathbb{R}^d)$ به طوری که $\|w_1 - w^*\| \leq c$. فرض کنید عدد $\sigma > 0$ وجود دارد که برای هر $t \in [T]$ داریم $\text{tr}(\text{cov}(d_t)) \leq \sigma^2$. تعریف کنید $r := \sqrt{\rho^2 + \sigma^2}$. پس از T تکرار حلقه‌ی همگرایی در الگوریتم ۲.۴ با طول گام ثابت $s = c/(r\sqrt{T})$ داریم:

$$\mathbb{E}(f(\bar{w}) - f(w^*)) \leq \frac{rc}{\sqrt{T}}. \quad (5.4)$$

پس برای هر $\epsilon > 0$ ، عدد $M = r^2 c^2 / \epsilon^2$ تضمین می‌کند که برای هر تعداد تکرار $T \geq M$ داریم:

$$\mathbb{E}(f(\bar{w}) - f(w^*)) \leq \epsilon.$$

برهان. تعریف کنید $J(\bar{w}) := \mathbb{E}(f(\bar{w}) - f(w^*))$.

$$J(\bar{w}) \leq \frac{1}{2sT} \mathbb{E} \left(\sum_{t=1}^T (\|w_t - w^*\|^2 - \|w_{t+1} - w^*\|^2) \right) + \frac{s}{2T} \sum_{t=1}^T \text{tr}(\text{cov}(d_t)) + \frac{s}{2T} \sum_{t=1}^T \mathbb{E}(\|d_t\|^2)$$

$$= \frac{1}{2sT} \mathbb{E}(\|w_1 - w^*\|^2 - \|w_T - w^*\|^2) + \frac{s}{2T} \sum_{t=1}^T \text{tr}(\text{cov}(d_t)) + \frac{s}{2T} \sum_{t=1}^T \mathbb{E}(\|d_t\|^2)$$

$$\leq \frac{1}{2sT} \|w_1 - w^*\|^2 + \frac{s}{2T} \sum_{t=1}^T \text{tr}(\text{cov}(d_t)) + \frac{s}{2T} \sum_{t=1}^T \mathbb{E}(\|d_t\|^2)$$

تکمیل برای توابع محدب لیپشیتز

لدامی برهان.

$$\leq \frac{C^2}{2sT} + \frac{\sigma^2}{2}s + \frac{s}{2T} \sum_{t=1}^T \|E(d_t)\|^2$$

لیپشیتز بودن تابع هدف، گزاره ۳.۴

می دانیم $E(d_t) \in \mathcal{B}_M(0) \subseteq \mathcal{B}_M(0) \subseteq \mathcal{B}_M(0)$

$E(d_t) = E(E(d_t|w_t))$ گزاره ۱۵ در ضمیمه

$$E(d_t) = E(E(d_t|w_t)) \implies \|E(d_t)\| \leq M$$

بنابراین نتیجه می شود که

$$\leq \frac{C^2}{2T} \frac{1}{s} + \frac{(\sigma^2 + M^2)}{2}s \longrightarrow s^* = \frac{C}{\sqrt{T}}$$

نسبت به s بهینه کنیم

جایگزینی در کران بالا نتیجه می دهد

$$\leq \frac{C^2}{2T} \frac{1}{s^*} + \frac{1}{2}s^* = \frac{C^2}{2T} \frac{\sqrt{T}}{C} + \frac{1}{2} \frac{C}{\sqrt{T}} = \frac{1}{2\sqrt{T}} + \frac{1}{2\sqrt{T}} = \frac{1}{\sqrt{T}}$$

ضمیمه: مجموعه ها و توابع محدب

تعریف ۱.۱ (مجموعه ی محدب). مجموعه ی $C \subseteq \mathbb{R}^n$ را محدب گویند اگر برای هر دو نقطه ی دلخواه $x, y \in C$ پاره خط $[x, y]$ در C قرار بگیرد، یعنی $[x, y] \subseteq C$ که $[x, y] := \{\lambda x + (1 - \lambda)y \mid \lambda \in [0, 1]\}$.

تعریف ۴.۱ (تابع محدب). فرض کنید $C \subseteq \mathbb{R}^n$ مجموعه ای دلخواه باشد. تابع $f : C \rightarrow \mathbb{R}$ را محدب می گویند اگر C محدب باشد و برای هر دو نقطه ی متمایز دلخواه $x, y \in C$ و هر عدد حقیقی $\lambda \in (0, 1)$ داشته باشیم:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (۱.۱)$$

تابع f را محدب اکید گویند اگر نامساوی در (۱.۱) به صورت اکید برقرار باشد. تابع f را مقعر گویند اگر تابع قرینه ی آن یعنی $-f$ محدب باشد.

گزاره ۵.۱. فرض کنید $C \subseteq \mathbb{R}^n$ مجموعه ای دلخواه باشد. تابع $f : C \rightarrow \mathbb{R}$ محدب است اگر و تنها اگر برای نقاط متمایز دلخواه $x_1, \dots, x_k \in C$ و اعداد حقیقی $\lambda_1, \dots, \lambda_k \in (0, 1)$ که $\sum_{i=1}^k \lambda_i = 1$ داشته باشیم:

$$f\left(\sum_{i=1}^k \lambda_i x_i\right) \leq \sum_{i=1}^k \lambda_i f(x_i). \quad (۲.۱)$$

گزاره ۶.۱. فرض کنید $C \subseteq \mathbb{R}^n$ و تابع $f : C \rightarrow \mathbb{R}$ محدب باشد. اگر x^* یک کمینه کننده ی موضعی f باشد، آن گاه یک کمینه کننده ی سراسری نیز هست. اگر f محدب اکید باشد، کمینه کننده ی سراسری یکتاست.

گزاره ۷.۱. فرض کنید $C \subseteq \mathbb{R}^n$ یک مجموعه و $f : C \rightarrow \mathbb{R}$ تابعی دلخواه باشد. آن گاه گزاره های زیر معادل اند.

- تابع f محدب است.
- مجموعه ی $\text{epi}(f)$ محدب است.
- مجموعه ی $\text{spi}(f)$ محدب است.

$$\text{epi}(f) = \{(x, y) \in C \times \mathbb{R} : y \geq f(x)\}$$

ای گراف اکید

ضمیمه امید ریاضی شرطی و نامساوی ینسن

• فرض کنید که $(\Omega, \mathcal{F}, \mathbb{P})$ یک فضای احتمال باشد.

گزاره ۹.۱. بردارهای تصادفی $x : \Omega \rightarrow \mathbb{R}^d$ و $y : \Omega \rightarrow \mathbb{R}^d$ همراه با فرض های $\mathbb{E}(\|x\|) < +\infty$ و $\mathbb{E}(\|y\|) < +\infty$ را در نظر بگیرید و فرض کنید $u : \Omega \rightarrow \mathbb{R}^k$ بردار تصادفی دلخواه دیگری باشد. در این صورت، گزاره های زیر برقرارند.

- برای هر $a \in \mathbb{R}$ داریم $\mathbb{E}(ax|u) = a\mathbb{E}(x|u)$.
- $\mathbb{E}(x+y|u) = \mathbb{E}(x|u) + \mathbb{E}(y|u)$.
- فرض کنید $d=1$. اگر $x \leq y$ ، آن گاه $\mathbb{E}(x|u) \leq \mathbb{E}(y|u)$.
- همواره داریم $\mathbb{E}(\mathbb{E}(x|u)) = \mathbb{E}(x)$ که به قاعده ی امید ریاضی کل معروف است.
- $\mathbb{E}(y^T x|y) = y^T \mathbb{E}(x|y)$.

• فرض کنید x و u مستقل هستند. تابع $f : \mathbb{R}^k \times \mathbb{R}^d \rightarrow \mathbb{R}$ را در نظر بگیرید به طوری که شرط $\mathbb{E}(|f(u, x)|) < +\infty$ برآورده شود. در این صورت، داریم $\mathbb{E}(f(u, x)|u) = \mathbb{E}_x(f(u, x))$.

گزاره ۱۰.۱. متغیر تصادفی $x : \Omega \rightarrow \mathbb{R}^d$ را در نظر بگیرید. فرض کنید $\mathbb{E}(\|x\|) < +\infty$ و $\text{rng}(x) \subseteq C$ که C یک مجموعه ی محدب است. همچنین، $f : C \rightarrow \mathbb{R}$ تابعی محدب و اندازه پذیر است. در این صورت، روابط زیر برقرارند.

- $\mathbb{E}(f(x)) < +\infty$ و $\mathbb{E}(x) \in C$.
- $f(\mathbb{E}(x)) \leq \mathbb{E}(f(x))$. این رابطه به نامساوی ینسن معروف است.
- اگر $y : \Omega \rightarrow \mathbb{R}^d$ متغیری تصادفی باشد که $\mathbb{E}(\|y\|) < +\infty$ ، آن گاه تقریباً همه جا داریم $f(\mathbb{E}(x|y)) \leq \mathbb{E}(f(x)|y)$.

ضمیمه بهینه ی سراسری و زیرگرادیان

تعریف ۵.۱ (کمینه کننده ی سراسری و موضعی). فرض کنید $A \subseteq \mathbb{R}^n$ و $f : A \rightarrow \mathbb{R}$ تابعی دلخواه باشد. اگر $\forall x \in A, f(x^*) \leq f(x)$ ، آن گاه گوئیم x^* یک کمینه کننده ی سراسری f است. همچنین، اگر $\exists r > 0, \forall x \in B_r(x^*) \cap A, f(x^*) \leq f(x)$ ، آن گاه گوئیم x^* یک کمینه کننده ی موضعی f است.

گزاره ۲۸.۱. فرض کنید $A \subseteq \mathbb{R}^n$ و $f : A \rightarrow \mathbb{R}$ یک تابع دلخواه است. x^* یک کمینه کننده ی سراسری f است اگر و تنها اگر $0 \in \partial f(x^*)$.

برهان. با توجه به تعاریف، گزاره های زیر معادل هستند

$$0 \in \partial f(x^*) \iff \forall z \in A, f(z) \geq f(x^*) + 0^T(z - x^*) \iff \forall z \in A, f(z) \geq f(x^*)$$