بسم الله الرحمن الرحیم

پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

جواب‌های مینیمم نرم در بهینه‌سازی محدب

نگارش
علیرضا تیره‌کار

استاد راهنما
دکتر مجید سلیمانی دامنه

پایان‌نامه برای دریافت درجه کارشناسی
در رشته ریاضیات و کاربردها

تیر ۱۳۹۶

تقدیم به

پدرم که عالمانه به من آموخت تا چگونه در عرصه زندگی، ایستادگی را تجربه نمایم
و به مادرم، دریای بی‌کران فداکاری و عشق که وجودم برایش همه رنج بود و وجودش برایم همه مهر.

سپاس گزاری...

تشکر و سپاس بیکران دارم خدمت استاد گرانقدرم، جناب آقای دکتر سلیمانی.

علیرضا تیره کار

تیر ۱۳۹۶

چکیده

دراین پروژه یک مسأله‌ی بهینه‌سازی با تابع هدف قویاً محدب و ناحیه شدنی بسته و محدب در نظر می‌گیریم، که این ناحیه شدنی مجموعه جواب‌های بهینه‌ی یک مسأله‌ی بهینه‌سازی دیگر است. یک الگوریتم بر پایه گرادیان را برای حل این مسأله مورد بررسی قرار می‌دهیم. در این راستا از برخی ابزارها، مانند عملگر تصویر متعامد، تابع فاصله، روش‌های برش و خواص تابع لیپ‌شیتز بهره خواهیم برد و بدنبال آن هستیم تا همگرایی و نرخ همگرایی الگوریتم، مذکور را نیز مورد مطالعه قرار دهیم.

واژگان کلیدی: بهینه‌سازی دو سطحی[1]، بهینه‌سازی محدب، جواب حداقل نرم، تحدب قوی

[1] Bilevel

فهرست مطالب

فصل ۱

تحدب

در این فصل، به بیان برخی مفاهیم و قضایا، که در فهم بهتر مطالب فصل‌های آتی مفید خواهند بود، می‌پردازیم.

۱.۱ تابع محدب قوی

تعریف ۱.۱.۱ (**تابع محدب قوی**). [۳] *فرض کنید C مجموعه‌ای محدب روی $\mathbb{R}^n$ باشد. تابع f را یک* ***تابع قویاً محدب*** *نامیم، هرگاه تابع $f(x) - \frac{\delta}{2}\|x\|^2$ برای یک $\delta > 0$ روی C محدب باشد، اسکالر δ را* ***پارامتر تحدب قوی*** *می‌نامیم.*

در ادامه به بیان چند قضیه در رابطه با توابع قویاً محدب می‌پردازیم.

قضیه ۲.۱.۱. *تابع f روی مجموعه‌ی محدب C قویاً محدب با پارامتر σ می‌باشد، اگر و تنها اگر داشته باشیم*

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}\lambda(1-\lambda)\|x-y\|^2 \qquad (1.1)$$

برای هر $x, y \in C$ و $\lambda \in [0, 1]$.

برهان. به [۳] رجوع کنید. $\square$

قضیه ۳.۱.۱. *هر تابع قویاً محدب روی C، تابعی اکیداً محدب روی C می‌باشد.*

برهان. به [۳] رجوع کنید. $\square$

قضیه ۴.۱.۱. *اگر تابع f روی مجموعه‌ی محدب C بطور پیوسته مشتق‌پذیر باشد، آنگاه f قویاً محدب (با پارامتر σ) روی C می‌باشد، اگر و تنها اگر داشته باشیم*

$$f(y) \geqslant f(x) + \nabla f(x)^T(y-x) + \frac{\sigma}{۲}\|x-y\|^{۲}, \quad \forall x, y \in C \tag{۲.۱}$$

برهان. به [۳] رجوع کنید. $\square$

قضیه ۵.۱.۱. *اگر تابع f روی مجموعه‌ی محدب C بطور پیوسته مشتق‌پذیر باشد، آنگاه f قویاً محدب (با پارامتر σ) روی مجموعه‌ی C باشد، اگر و تنها اگر*

$$(\nabla f(x) - \nabla f(y))^T(x-y) \geqslant \sigma\|x-y\|^{۲}.$$

برهان. به [۳] رجوع کنید. $\square$

قضیه ۶.۱.۱. *فرض کنید تابع f روی مجموعه‌ی محدب C دوبار پیوسته مشتق‌پذیر باشد. آنگاه f تابعی قویاً محدب (با پارامتر σ) روی C می‌باشد، اگر و تنها اگر داشته باشیم*

$$\nabla^{۲} f(x) \geqslant \sigma I, \quad \forall x \in C \tag{۳.۱}$$

برهان. به [۳] رجوع کنید. $\square$

۲.۱ بهینه‌سازی محدب

ابتدا مسأله‌ی برنامه‌ریزی محدب را بصورت زیر تعریف می‌کنیم.

تعریف ۱.۲.۱. [۳] *مسأله‌ی زیر که در آن $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ تابعی محدب و $C \subseteq \mathbb{R}^n$ مجموعه‌ای ناتهی، بسته و محدب می‌باشد را یک* ***مسأله‌ی برنامه‌ریزی محدب*** *(CP) می‌گوییم.*

$$\begin{aligned} &\min f(x) \\ &s.t. \quad x \in C \end{aligned} \tag{۴.۱}$$

قضیه ۱.۲.۲. [۳] *هر مینیمم موضعی یک مسأله‌ی برنامه‌ریزی محدب، یک مینیمم سراسری برای آن مسأله می‌باشد.*

برهان. فرض کنید $\bar{x}$ یک مینیمم موضعی برای مسأله‌ی برنامه‌ریزی محدب (۱.۴) باشد. پس $\varepsilon > ۰$ وجود دارد، بطوریکه $f(\bar{x}) \leqslant f(x)$ برای هر $x \in B(\bar{x}; \varepsilon) \cap S$. به برهان خلف، فرض کنید

$$\exists x^* \in C; \quad f(x^*) < f(\bar{x}).$$

تعریف می‌کنیم

$$x^\circ := \bar{x} + \lambda(x^* - \bar{x})$$

که در آن

$$\lambda \in \left(۰, \min\left\{۱, \frac{\varepsilon}{۲\|x^* - \bar{x}\|}\right\}\right).$$

چون C محدب است، داریم

$$x^\circ = \lambda x^* + (۱ - \lambda)\bar{x} \in C$$

به علاوه

$$\|x^\circ - \bar{x}\| = \lambda\|x^* - \bar{x}\| < \frac{\varepsilon}{۲\|x^* - \bar{x}\|}\|x^* - \bar{x}\| = \frac{\varepsilon}{۲}$$

پس $x^\circ \in B(\bar{x}, \varepsilon) \cap C$. بنابراین $f(\bar{x}) \leqslant f(x^\circ)$. از طرفی داریم

$$\begin{aligned} f(\bar{x}) \leqslant f(x^\circ) &= f(\lambda x^* + (۱ - \lambda)\bar{x}) \\ &\leqslant \lambda f(x^*) + (۱ - \lambda) f(\bar{x}) \\ &< \lambda f(\bar{x}) + (۱ - \lambda) f(\bar{x}) \\ &= f(\bar{x}) \end{aligned} \tag{۱.۵}$$

در نتیجه

$$f(\bar{x}) < f(\bar{x}).$$

که تناقض است. پس فرض خلف باطل است و حکم قضیه ثابت می‌شود. □

قضیه ۱.۲.۳. *در یک مسأله‌ی برنامه‌ریزی محدب، اگر f اکیداً محدب باشد، آنگاه هر جواب بهینه‌ی موضعی، یک جواب بهینه‌ی سراسری یکتا می‌باشد.*

برهان. مشابه قضیه ۱.۲.۲ است. □

قضیه ۱.۲.۴. [۳] *مجموعه جواب‌های بهینه‌ی یک مسأله‌ی برنامه‌ریزی محدب، مجموعه‌ای محدب است.*

برهان. فرض کنید O مجموعه‌ی جواب‌های بهینه‌ی مسأله‌ی برنامه‌ریزی محدب (۱.۴) باشد. حال فرض کنید

$$\bar{x}, x^{\circ} \in O_p$$

آنگاه داریم

$$\bar{x} \in O_p \Longrightarrow f(\bar{x}) \leqslant f(x), \quad \forall x \in C$$

$$x^{\circ} \in O_p \Longrightarrow f(x^{\circ}) \leqslant f(x), \quad \forall x \in C$$

با در نظر گرفتن $\lambda \in [۰, ۱]$، داریم

$$f(\lambda\bar{x} + (۱ - \lambda)x^{\circ}) \leqslant \lambda f(\bar{x}) + (۱ - \lambda)f(x^{\circ}) \leqslant f(x), \quad \forall x \in C$$

از طرفی چون C محدب است، داریم

$$\lambda\bar{x} + (۱ - \lambda)x^{\circ} \in C$$

پس

$$\lambda\bar{x} + (۱ - \lambda)x^{\circ} \in O_p$$

□

قضیه ۱.۲.۵. *مجموعه جواب‌های بهینه‌ی مسأله‌ی بهینه‌سازی زیر، که در آن C مجموعه‌ای محدب و f تابعی اکیداً محدب است، تهی یا تک عضوی است.*

$$P: \quad \min f(x)$$

$$s.t. \quad x \in C$$

برهان. به برهان خلف، فرض می‌کنیم مجموعه‌ی جواب‌های بهینه، دارای دو عضو بصورت $x, y \in O_P$ باشد. آنگاه داریم

$$f(x) = f(y)$$

$$f(\frac{1}{2}x + \frac{1}{2}y) < \frac{1}{2}f(x) + \frac{1}{2}f(y) = f(x)$$

در حالی که می‌دانیم، $k := \frac{1}{2}x + \frac{1}{2}y \in C$، پس نامساوی اکید فوق با مینیمم بودن $f(x)$ در تناقض است. $\square$

فصل ۲

تکنیک‌ها و الگوریتم‌های مربوطه

۱.۲ تابع تصویر متعامد

تعریف ۱.۱.۲. [۳] *فرض کنیم $S \subseteq \mathbb{R}^n$ ناتهی، بسته و محدب باشد، آنگاه تابع فاصله‌ی متناظر با S را بصورت زیر تعریف می‌کنیم.*

$$d_S : \mathbb{R}^n \longrightarrow \mathbb{R}$$
$$d_S(x) = \inf_{y \in S} \|x - y\| \tag{۱.۲}$$

بدیهی است اگر $x \in S$ باشد، آنگاه $d_S(x) = 0$.

در واقع d_S مینیمم فاصله‌ی x تا اعضای S می‌باشد. به برداری که این مینیمم در آن رخ می‌دهد، **تصویر متعامد** x روی S می‌گوییم و با نماد $P_S(x)$ نمایش می‌دهیم.

تعریف ۲.۱.۲. [۳] *برای $S \subseteq \mathbb{R}^n$ ناتهی، بسته و محدب، **عملگر تصویر متعامد** به صورت زیر تعریف می‌گردد.*

$$P_S : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$
$$P_S(x) = arg \min_{y \in S} \|x - y\|^2 \tag{۲.۲}$$

بدیهی است اگر $x \in S$ باشد، آنگاه $P_S(x) = x$.

عملگر تصویر متعامد دارای خواص زیر است.

قضیه ۳.۱.۲. *برای هر $x \in \mathbb{R}^n$ و $y \in S$ داریم*

$$\langle x - P_S(x), y - P_S(x) \rangle \leqslant 0 \qquad (۳.۲)$$

برهان. طبق تعریف ۲.۱.۲، بردار $P_S(x)$ جواب بهینه‌ی مسأله‌ی $\min_{y\in S} \|y-x\|^2$ می‌باشد. این مسأله یک CP می‌باشد، بنابراین هر جواب بهینه‌ی آن ایستای مقید است. پس با قرار دادن

$$h(y) = \|y-x\|^2$$

داریم

$$\nabla h(P_S(x))^T (y - P_S(x)) \geqslant 0, \quad \forall y \in S$$
$$\xrightarrow{\nabla h(y) = 2(y-x)} 2(P_S(x) - x)^T (y - P_S(x)) \geqslant 0, \quad \forall y \in S$$
$$\Longrightarrow \langle (x - P_S(x)), (y - P_S(x)) \rangle \leqslant 0, \quad \forall y \in S$$

□

قضیه ۴.۱.۲. *برای هر $x, y \in \mathbb{R}^n$ داریم*

$$\langle P_S(x) - P_S(y), x - y \rangle \geqslant \|P_S(x) - P_S(y)\|^2 \qquad (۴.۲)$$

برهان. بنابر قضیه قبل

$$(x - P_S(x))^T (z - P_S(x)) \leqslant 0, \quad \forall z \in C$$

با قرار دادن $Z = P_S(y)$، داریم

$$(x - P_S(x))^T (P_S(y) - P_S(x)) \leqslant 0$$

با قرار دادن $Z = P_S(x)$، داریم

$$(y - P_S(y))^T (P_S(x) - P_S(y)) \leqslant 0$$

با جمع دو رابطه‌ی اخیر داریم

$$(-x+P_S(x)+y-P_S(y))^T(P_S(x)-P_S(y)) \leqslant ۰$$
$$\Longrightarrow (x-y)^T(P_S(x)-P_S(y)) \geqslant (P_S(x)-P_S(y))^T(P_S(x)-P_S(y))$$
$$\Longrightarrow \langle P_S(x)-P_S(y), x-y\rangle \geqslant \|P_S(x)-P_S(y)\|^۲.$$

□

قضیه ۵.۱.۲. *برای هر $x, y \in \mathbb{R}^n$ داریم*

$$\|P_S(x)-P_S(y)\| \leqslant \|x-y\| \tag{۵.۲}$$

برهان. اگر $\|P_S(x)-P_S(y)\| = ۰$ باشد، حکم بدیهی است. بنابراین فرض کنید

$$\|P_S(x)-P_S(y)\| \neq ۰.$$

بنابر قضیه قبل، داریم

$$\|P_S(x)-P_S(y)\|^۲ \leqslant (P_S(x)-P_S(y))^T(x-y)$$
$$\overset{\text{کوشی شوارتز}}{\Longrightarrow} \|P_S(x)-P_S(y)\|^۲ \leqslant \|P_S(x)-P_S(y)\|\|x-y\|$$
$$\overset{\div\|P_S(x)-P_S(y)\|}{\Longrightarrow} \|P_S(x)-P_S(y)\| \leqslant \|x-y\|$$

□

مثال ۶.۱.۲. *فرض کنید $C = B[\bar{x}, r]$، $x \in \mathbb{R}^n$ و $r > ۰$ داده شده‌اند. می‌خواهیم یک رابطه برای عملگر تصویر متعامد P_C پیدا کنیم. داریم:*

$$h(y) := \|y-x\|^۲$$

و

$$(P): \quad \min \|y-x\|^۲$$
$$s.t. \quad \|y-\bar{x}\| \leqslant r$$

پس

$$if \quad x \in C \Longrightarrow P_C(x) = x$$

$$if \quad x \notin C \Longrightarrow \|\bar{x} - x\| > r \Longrightarrow \text{مسأله را برای این حالت حل می‌کنیم}$$

ثابت می‌کنیم، مینیمم برای مسأله‌ی P در حالت $\|y - \bar{x}\| = r$ رخ می‌دهد. فرض کنید y^* جواب بهینه است. اگر

$$\|y^* - \bar{x}\| < r$$

آنگاه

$$\nabla h(y^*) = ۰$$

پس $x = y^*$، که در تناقض است، زیرا $y^* \in C$ ولی $x \notin C$. بنابراین داریم $\|y^* - \bar{x}\| = r$ و به مسأله‌ی زیر می‌رسیم:

$$(q): \quad \min \|y - x\|^۲$$
$$s.t. \quad \|y - \bar{x}\| = r$$

$$\begin{aligned} h(y) &= \|y - \bar{x} + \bar{x} - x\|^۲ \\ &= \|y - \bar{x}\|^۲ + \|\bar{x} - x\|^۲ + ۲(y - \bar{x})^T(\bar{x} - x) \\ &= r^۲ + \|x - \bar{x}\|^۲ + ۲(y - \bar{x})^T(\bar{x} - x) \end{aligned}$$

که $r^۲ + \|x - \bar{x}\|^۲$ ثابت می‌باشد، پس به دنبال بهینه کردن $۲(y - \bar{x})^T(\bar{x} - x)$ می‌باشیم. بنابراین داریم

$$(Z): \quad \min ۲(y - \bar{x})^T(\bar{x} - x)$$
$$s.t. \quad \|y - \bar{x}\| = r$$

بنابر نامساوی کوشی‌شواتز داریم

$$۲(y - \bar{x})^T(\bar{x} - x) \geqslant -۲\|y - \bar{x}\|\|\bar{x} - x\| = -۲r\|\bar{x} - x\|$$

پس $-۲r\|\bar{x}-x\|$ یک کران پایین برای مقدار بهینه‌ی مسأله‌ی Z است. اکنون مقدار تابع هدف در $\bar{x}+\frac{r(x-\bar{x})}{\|x-\bar{x}\|}$ را محاسبه می‌کنیم و داریم

$$۲(\bar{x}+\frac{r(x-\bar{x})}{\|x-\bar{x}\|}-\bar{x})^T(\bar{x}-x)=-۲r\frac{\|\bar{x}-x\|^۲}{\|\bar{x}-x\|}=-۲r\|\bar{x}-x\|$$

پس $\bar{x}+r\frac{(x-\bar{x})}{\|x-\bar{x}\|}$ یک جواب بهینه‌ی مسأله‌ی (Z) است.

۲.۲ شرایط KKT

تعریف ۱.۲.۲. [۳] ***نقاط** KKT* *: در مسأله‌ی مینیمم‌سازی زیر*

$$\begin{aligned} &\min f(x) \\ &s.t. \quad g_i(x)\leqslant ۰ \quad i=۱,۲,\cdots,m \\ &\qquad h_j(x)=۰ \quad j=۱,۲,\cdots,p \end{aligned} \tag{۶.۲}$$

هرگاه توابع $f,g_۱,g_۲,\cdots,g_m,h_۱,\cdots,h_p$ روی $\mathbb{R}^n$ بطور پیوسته مشتق‌پذیر باشند، نقطه‌ی شدنی x^* را یک نقطه‌ی KKT گوییم، هرگاه وجود داشته باشند $\lambda_۱,\cdots,\lambda_m\geqslant ۰$ و $\mu_۱,\mu_۲,\cdots,\mu_p\in\mathbb{R}$، بطوریکه

$$\nabla f(x^*)+\sum_{i=۱}^{m}\lambda_i\nabla g_i(x^*)+\sum_{j=۱}^{p}\mu_j\nabla h_j(x^*)=۰$$

$$\lambda_i g_i(x^*)=۰ \quad i=۱,۲,\ldots,m$$

قضیه ۲.۲.۲ (کافی بودن شرایط KKT برای حل مسائل برنامه‌ریزی محدب). [۳] *فرض کنید x^* یک جواب شدنی مسأله‌ی (۶.۲) باشد. همچنین f و $g_۱,\cdots,g_m$ توابعی محدب و بطور پیوسته مشتق‌پذیر روی $\mathbb{R}^n$ باشند و بعلاوه $h_۱,\cdots,h_p$ توابعی آفینی باشند.*

فرض کنید ضرایب $\lambda_۱,\cdots,\lambda_m\geqslant ۰$ و $\mu_۱,\cdots,\mu_p\in\mathbb{R}$ وجود دارند، بطوریکه

$$\begin{aligned} &\nabla f(x^*)+\sum_{i=۱}^{m}\lambda_i\nabla g_i(x^*)+\sum_{j=۱}^{p}\mu_j\nabla h_j(x^*)=۰ \\ &\lambda_i g_i(x^*)=۰ \quad i=۱,\cdots,m \end{aligned} \tag{۷.۲}$$

آنگاه x^* یک جواب بهینه‌ی مسأله‌ی (۲.۶) است.

برهان. نشان می‌دهیم به ازای $x \in \mathbb{R}^n$، $f(x) \geqslant f(x^*)$. تابع $S(.)$ را بصورت زیر تعریف می‌کنیم.

$$S(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

از آنجاییکه تابع S، مجموع توابع محدب می‌باشد، پس خود یک تابع محدب است. از طرفی داریم x^* مینیمم کننده‌ی S روی $\mathbb{R}^n$ می‌باشد. بنابراین

$$\begin{aligned} f(x^*) &= f(x^*) + \sum_{i=1}^{m} \lambda_i g_i(x^*) + \sum_{j=1}^{p} \mu_j h_j(x^*) \\ &= S(x^*) \\ &\leqslant S(x) \\ &= f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \\ &\leqslant f(x) \quad (\lambda_i \geqslant 0, g_i(x) \leqslant 0, h_j(x) = 0) \\ &\Longrightarrow f(x^*) \leqslant f(x) \end{aligned}$$

بنابراین x^* یک جواب بهینه‌ی مسأله‌ی مورد نظر است. $\square$

قضیه ۳.۲.۲ (لازم بودن شرایط KKT تحت شرایط اسلاتر). [۳] *فرض کنید x^* یک جواب بهینه مسئله‌ی زیر باشد:*

$$\begin{aligned} (P): \quad & \min f(x) \\ & s.t. \quad g_i(x) \leqslant 0 \quad i = 1, 2, \ldots, m \end{aligned}$$

همچنین $f, g_1, \ldots, g_m$ توابعی محدب و بطور پیوسته مشتق‌پذیر روی $\mathbb{R}^n$ باشند. فرض کنید $\hat{x} \in \mathbb{R}^n$ وجود دارد بطوریکه

$$g_i(\hat{x}) < 0, \quad i = 1, 2, \ldots, m$$

آنگاه $\hat{x}$ یک نقطه KKT است.

برهان. هنگامی‌که x^* یک جواب بهینه‌ی مسئله (P) باشد، شرایط فریزجان برقرار می‌باشند. بنابراین وجود دارند $\tilde{\lambda}_0, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_m \geqslant 0$ که همگی صفر نیستند و داریم:

$$\begin{aligned} &\tilde{\lambda}_0 \nabla f(x^*) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla g_i(x^*) = 0 \\ &\tilde{\lambda}_i g_i(x^*) = 0 \quad i = 1, \ldots, m \end{aligned} \tag{۸.۲}$$

اکنون تنها چیزی که باید ثابت کنیم این است که $\tilde{\lambda}_0 > 0$ و با درنظر گرفتن

$$\lambda_i = \frac{\tilde{\lambda}_i}{\tilde{\lambda}_0} \quad i = 1, \ldots, m$$

شرایط KKT برقرار می‌شوند. برای اثبات $\tilde{\lambda}_0 > 0$ فرض کنیم $\tilde{\lambda}_0 = 0$، آنگاه

$$\sum_{i=1}^{m} \tilde{\lambda}_i \nabla g_i(x^*) = 0$$

بنابر نامساوی گرادیان برای هر $i = 1, \ldots, m$ داریم:

$$0 > g_i(x^*) + \nabla g_i(x^*)^T (\hat{x} - x^*)$$

بنابراین داریم:

$$0 > \sum_{i=1}^{m} \tilde{\lambda}_i g_i(x^*) + \left(\sum_{i=1}^{m} \tilde{\lambda}_i \nabla g_i(x^*) \right)^T (\hat{x} - x^*) \tag{۹.۲}$$

که همه‌ی $\tilde{\lambda}_i$ ها صفر می‌باشند که با توجه به رابطه‌ی (۹.۲) داریم $0 > 0$. بنابراین فرض خلف باطل و حکم ثابت می‌شود. □

۳.۲ فاصله‌ی بِرگمن

فرض کنید $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ یک تابع محدب اکید باشد. ***فاصله‌ی برگمن***[1] *این تابع بصورت زیر تعریف می‌گردد.*

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y \rangle \tag{۱۰.۲}$$

فاصله‌ی بِرگمن دارای دو خاصیت پایه‌ای زیر می‌باشد [۳].

[1] Bergman distance

۱. $\forall x,y \in \mathbb{R}^n; x \neq y \quad D_h(x,y) > 0$،

۲. $\forall x,y \in \mathbb{R}^n; x = y. \Longleftrightarrow D_h(x,y) = 0$

برای اثبات ویژگی (۱) به ترتیب زیر عمل می‌کنیم.

$$D_h(x,y) := h(x) - h(y) - \nabla h(y)(x-y)$$

بنابر محدب اکید بودن تابع h داریم

$$h(x) - h(y) - \nabla h(y)^T(x-y) > 0$$

$$\Rightarrow D_h(x-y) > 0$$

برای اثبات ویژگی (۲) نیز بترتیب زیر عمل می‌کنیم. اگر $x = y$، آنگاه

$$D_h(x,y) = h(x) - h(x) - \nabla h(x)(x-x) = 0$$

حال اگر برعکس داشته باشیم $D_h(x,y) = 0$، بنابر قسمت (۱)، داریم $x = y$. بعلاوه همانطور که قبلاً دیدیم، اگر تابع h قویاً محدب با پارامتر $\sigma > 0$ باشد، داریم

$$D_h(x,y) \geqslant \frac{\sigma}{2}\|x-y\|^2$$

حال اگر ω که در بخش قبل معرفی شد یک تابع قویاً محدب با مرکز پروکسی a باشد، آنگاه برای هر $x \in \mathbb{R}^n$ داریم

$$\omega(x) = D_\omega(x,a) \geqslant \frac{\sigma}{2}\|x-a\|^2 \tag{۱۱.۲}$$

همچنین برای هر $x,y \in \mathbb{R}^n$ داریم

$$D_\omega(x,y) \geqslant \frac{\sigma}{2}\|x-y\|^2 \tag{۱۲.۲}$$

لم ۱.۳.۲. [۳] فرض کنید $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ یک تابع قویاً محدب با پارامتر $\delta > 0$ باشد، آنگاه برای هر $x,y,z \in \mathbb{R}^n$ داریم

$$D_h(x,y) - D_h(y,z) - D_h(x,z) = \langle x-y, \nabla h(z) - \nabla h(y)\rangle \tag{۱۳.۲}$$

برهان.

$$D_h(x,y) = h(x) - h(y) - \nabla h(y)^T(x-y)$$

$$+$$

$$D_h(y,z) = h(y) - h(z) - \nabla h(z)^T(y-z)$$

$$-$$

$$D_h(x,z) = h(x) - h(z) - \nabla h(z)^T(x-z)$$

$$\equiv \langle x-y, \nabla h(z) - \nabla h(y)\rangle$$

$\square$

یکی دیگر از خواص مهم نگاشت گرادیان، ویژگی یکنوایی آن با نسبت M می‌باشد [۸].

۴.۲ ترازهای برشی

مسئله‌ی زیر را به گونه‌ای در نظر بگیرد، که در آن f تابعی قویاً محدب و بطور پیوسته مشتق‌پذیر است و بعلاوه $\emptyset \neq X \subseteq \mathbb{R}^n$. مجموعه جواب‌های بهینه‌ی این مسئله را با X^* نمایش می‌دهیم.

$$\begin{aligned} &\min f(x) \\ &s.t. \quad x \in X \end{aligned} \tag{۱۴.۲}$$

مفهوم صفحه‌ی برشی، یک مفهوم پایه‌ای در الگوریتم‌های بهینه‌سازی مانند روش‌های برش‌های بیضوی می‌باشد. به طور مثال ابتدا فرض کنید مسأله نامقید است و نقطه‌ی $x \in \mathbb{R}^n$ را در نظر بگیرید. هدف یافتن ابر صفحه‌ای برای جداسازی نقطه‌ی $x \in X$ از X^* می‌باشد.

برای هر $x \in X$، داریم

$$X^* \subseteq \{z \in \mathbb{R}^n : \nabla f(x)^T(x-z) \geqslant ۰\}.$$

اهمیت نتیجه بالا، استخراج ابرصفحه‌ی زیر از آن می‌باشد

$$\{z \in \mathbb{R}^n : \langle \nabla f(x), x - z \rangle < ۰\}.$$

توجه داریم که x متعلق به ابر صفحه‌ی زیر است

$$H = \{z \in \mathbb{R}^n : \langle \nabla f(x), x - z \rangle = ۰\}. \tag{۱۵.۲}$$

از آنجا که $x \in H$ می‌باشد، به آن **برش خنثی** می‌گوییم. حال اگر $x \notin H$ آنگاه به H **برش عمیق** گوییم. برش‌های عمیق هسته‌ی اصلی الگوریتم‌های حداقل نرم گرادیان می‌باشند، که در ادامه درباره‌ی آنها صحبت می‌کنیم.

در قسمت بعدی درباره‌ی چگونگیِ ساخت این برش‌ها در حالت‌های مختلف صحبت خواهیم کرد.

۵.۲ تصویر گرادیان

تعریف ۱.۵.۲. [۲] فرض کنید $M > ۰$.

الف) برای هر $x \in \mathbb{R}^n$، نگاشت تصویر گرادیان بصورت زیر تعریف می‌گردد.

$$T_M(x) \equiv P_X(x - \frac{۱}{M}\nabla f(x)) \tag{۱۶.۲}$$

ب) نگاشت گرادیان مقید بصورت زیر تعریف می‌گردد.

$$G_M(x) \equiv M(x - T_M(x)) = M[x - P_X(x - \frac{۱}{M}\nabla f(x)] \tag{۱۷.۲}$$

در حالت غیر مقید یعنی زمانی که $X = \mathbb{R}^n$ می‌باشد، اپراتور تصویر متعامد یک اپراتور همانی است، پس

۱. نگاشت تصویر گرادیان T_M برابر است با $I - \frac{۱}{M}\nabla f$

۲. نگاشت گرادیان، G_M برابر است با ∇f.

توجه داریم که

$$x \in X^* \Longleftrightarrow G_M(x) = ۰$$

برهان. فرض کنیم $x^* \in X^*$. آنگاه $x^* = P_X(x^* - \frac{۱}{M}\nabla f(x^*))$. پس

$$(x^* - \frac{۱}{M}\nabla f(x^*) - x^*)^T(x - x^*) \leqslant ۰, \quad \forall x \in X$$
$$\Longleftrightarrow -\frac{۱}{M}\nabla f(x^*)^T(x - x^*) \leqslant ۰, \quad \forall x \in X$$
$$\overset{M>۰}{\Longleftrightarrow} \nabla f(x^*)^T(x - x^*) \geqslant ۰, \quad \forall x \in X$$

$\Longleftrightarrow$ x^* ایستای مقید برای P است $\Leftrightarrow x^* \in X^*$

□

نیم فضای متناظر با برش‌های عمیق بصورت زیر تعریف می‌گردد.

$$Q_{M,\alpha,x} \equiv \{z \in \mathbb{R}^n : \langle G_M(x), x - z\rangle \geqslant \frac{۱}{\alpha M}\|G_M(x)\|^۲\} \qquad (۱۸.۲)$$

بطوریکه مقدار α و M به شرایط مسأله بستگی دارد. بدیهی است که در حالت نامقید (یعنی زمانی که $X = \mathbb{R}^n$ باشد) داریم $G_M(x) \equiv \nabla f(x)$ و (۱۸.۲) بصورت زیر تعریف می‌گردد.

$$Q_{M,\alpha,x} \equiv \{z \in \mathbb{R}^n : \langle \nabla f(x), x - z\rangle \geqslant \frac{۱}{\alpha M}\|\nabla f(x)\|^۲\} \qquad (۱۹.۲)$$

حال شرایط مسأله را به دو حالت تقسیم می‌کنیم.

۱. ثابت لیپ‌شیتز L معلوم باشد.

۲. ثابت لیپ‌شیتز L معلوم نباشد.

۱.۵.۲ ثابت لیپ‌شیتز L معلوم باشد

در حالت نامقید ($X = \mathbb{R}^n$) می‌توانیم با توجه به [۱۸] از رابطه‌ی زیر استفاده کنیم.

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geqslant \frac{۱}{L}\|\nabla f(x) - \nabla f(y)\|^۲, \quad \forall x, y \in \mathbb{R}^n \qquad (۲۰.۲)$$

با قراردادن $y = x^*$ بطوریکه $x^* \in X^*$ در رابطه‌ی (۲۰.۲)، با توجه به $\nabla f(x^*) = 0$، به رابطه‌ی زیر می‌رسیم

$$\langle \nabla f(x), x - x^* \rangle \geqslant \frac{1}{L}\|\nabla f(x)\|^2, \quad \forall x \in \mathbb{R}^n, x^* \in X^* \qquad (۲۱.۲)$$

بنابراین برای هر $x \in \mathbb{R}^n$ داریم $X^* \subseteq Q_{L,1,x}$.[۳]

زمانیکه مسأله مقید است و $X \neq \mathbb{R}^n$، رابطه‌ی (۲۱.۲) کمی پیچیده‌تر خواهد شد و داریم $X^* \subseteq Q_{L,\frac{4}{3},x}$. این نتیجه بر پایه‌ی خواص نگاشت گرادیان G_L بدست می‌آید.

لم ۲.۵.۲. [۲] برای هر $x, y \in \mathbb{R}^n$ داریم

$$\langle G_L(x) - G_L(y), x - y \rangle \geqslant \frac{3}{4L}\|G_L(x) - G_L(y)\|^2. \qquad (۲۲.۲)$$

برهان. بنابر (۴.۲)، داریم

$$\langle T_L(x) - T_L(y), (x - \frac{1}{L}\nabla f(x)) - (y - \frac{1}{L}\nabla f(y)) \rangle \geqslant \|T_L(x) - T_L(y)\|^2$$

و چون داریم $T_L = I - \frac{1}{L}G_L$

$$\begin{aligned}\langle (x - \frac{1}{L}G_L(x)) - (y - \frac{1}{L}G_L(y)), (x - \frac{1}{L}\nabla f(x)) - (y - \frac{1}{L}\nabla f(y)) \rangle \\ \geqslant \|(x - \frac{1}{L}G_L(x)) - (y - \frac{1}{L}G_L(y))\|\end{aligned}$$

بنابراین

$$\langle (x - \frac{1}{L}G_L(x)) - (y - \frac{1}{L}G_L(y)), (G_L(x) - \nabla f(x)) - (G_L(y) - \nabla f(y)) \rangle \geqslant 0$$

در نتیجه

$$\begin{aligned}\langle G_L(x) - G_L(y), x - y \rangle &\geqslant \frac{1}{L}\|G_L(x) - G_L(y)\|^2 \\ &+ \langle \nabla f(x) - \nabla f(y), x - y \rangle - \frac{1}{L}\langle G_L(x) - G_L(y), \nabla f(x) - \nabla f(y) \rangle\end{aligned}$$

حال بنابر (۲۰.۲) داریم

$$L\langle G_L(x)-G_L(y), x-y\rangle \geqslant \|G_L(x)-G_L(y)\|^۲ + \|\nabla f(x)-\nabla f(y)\|^۲$$

$$-\langle G_L(x)-G_L(y), \nabla f(x)-\nabla f(y)\rangle$$

$$\xRightarrow{\text{کوشی شوارتز}} L\langle G_L(x)-G_L(y), x-y\rangle \geqslant \|G_L(x)-G_L(y)\|^۲$$

$$+\|\nabla f(x)-\nabla f(y)\|^۲ - \|\nabla G_L(x)-\nabla G_L(y)\|.\|\nabla f(x)-\nabla f(y)\|$$

(۲۳.۲)

$$\alpha := \|G_L(x)-G_L(y)\|, \quad \beta := \|\nabla f(x)-\nabla f(y)\|$$

$$\alpha^۲ + \beta^۲ - \alpha\beta = \frac{۳}{۴}\alpha^۲ + (\frac{\alpha}{۲}-\beta)^۲ \geqslant \frac{۳}{۴}\alpha^۲$$

با توجه به رابطه‌ی (۲۳.۲)، داریم

$$L\langle G_L(x)-G_L(y), x-y\rangle \geqslant \frac{۳}{۴}\|G_L(x)-G_L(y)\|^۲$$

با قرار دادن $y = x^*$ که $x^* \in X^*$ در رابطه‌ی (۲۲.۲)، نتیجه می‌گیریم

$$X^* \subseteq Q_{L,\frac{۳}{۴},x}$$

□

لم زیر خلاصه بحث فوق است.

لم ۳.۵.۲. [۲] *برای هر* $x \in \mathbb{R}^n$ *و* $x^* \in X^*$ *داریم*

$$\langle G_L(x), (x-x^*)\rangle \geqslant \frac{۳}{۴L}\|G_L(x)\|^۲ \qquad (۲۴.۲)$$

که این یعنی

$$X^* \subseteq Q_{L,\frac{۴}{۳},x}$$

اگر $X = \mathbb{R}^n$ *باشد، آنگاه*

$$\langle \nabla f(x), x-x^*\rangle \geqslant \frac{۱}{L}\|\nabla f(x)\|^۲ \qquad (۲۵.۲)$$

که این یعنی

$$X^* \subseteq Q_{L,۱,x}$$

۲.۵.۲ ثابت لیپ‌شیتز L معلوم نباشد

لم ۴.۵.۲. [۲] *فرض کنید $x \in \mathbb{R}^n$ در رابطه‌ی زیر صدق کند*

$$f(T_M(x)) \leqslant f(x) + \langle \nabla f(x), T_M(x) - x\rangle + \frac{M}{۲}\|T_M(x) - x\|^۲ \tag{۲۶.۲}$$

آنگاه برای هر $x^* \in X^*$ داریم

$$\langle G_M(x), x - x^*\rangle \geqslant \frac{۱}{M}\|G_M(x)\|^۲ \tag{۲۷.۲}$$

که این یعنی

$$X^* \subseteq Q_{M,۲,x}$$

برهان. با در نظر گرفتن $x^* \in X^*$ و با توجه به رابطه‌ی (۲۶.۲)، داریم

$$۰ \leqslant f(T_M(x)) - f^* \leqslant f(x) - f^* + \langle \nabla f(x), T_M(x) - x\rangle + \frac{M}{۲}\|T_M(x) - x\|^۲. \tag{۲۸.۲}$$

با توجه به محدب بودن تابع f، داریم

$$f(x) - f(x^*) \leqslant \langle \nabla f(x), x - x^*\rangle$$

که با توجه به رابطه‌ی (۲۸.۲)، داریم

$$۰ \leqslant \langle \nabla f(x), T_M(x) - X^*\rangle + \frac{M}{۲}\|T_M(x) - x\|^۲ \tag{۲۹.۲}$$

حال با توجه به تعریف T_M و رابطه‌ی (۲۳.۲) داریم

$$\langle x - \frac{۱}{M}\nabla f(x) - T_M(x), T_M(x) - x^*\rangle \geqslant ۰$$

با ضریب رابطه‌ی اخیر در M و جمع آن با رابطه‌ی (۲۹.۲)، داریم

$$M\langle x - T_M(x), T_M(x) - x^*\rangle + \frac{M}{۲}\|T_M(x) - x\|^۲ \geqslant ۰$$

با قرار دادن $T_M(x) = x - \frac{1}{M}G_M(x)$ داریم

$$\langle G_M(x), x - x^* \rangle \geqslant \frac{1}{2M}\|G_M(x)\|^2$$

$\square$

لم ۲.۵.۵ (کاهشی). *فرض کنید f تابعی بطور پیوسته مشتق‌پذیر باشد و بعلاوه $f \in C_L^{1,1}$. آنگاه*

$$\forall x, y \in \mathbb{R}^n : \quad f(y) \leqslant f(x) + \langle \nabla f(y), x - y \rangle + \frac{L}{2}\|x - y\|^2 \tag{۲.۳۰}$$

برهان. به [۲] رجوع کنید. $\square$

فصل ۳

بیان مسئله

۱.۳ بیان ریاضی مسئله اصلی

مسأله بهینه‌سازی مقید زیر را در نظر بگیرید:

$$\min f(x) \qquad s.t. \quad x \in X \tag{۱.۳}$$

که در سراسر این پروژه فرضیات زیر را داریم:

الف) X *یک زیرمجموعه ناتهی، بسته و محدب در* $\mathbb{R}^n$ *می‌باشد،*

ب) f *به عنوان تابع هدف، تابعی محدب و روی* $\mathbb{R}^n$ *به طور پیوسته مشتق‌پذیر است. همچنین نگاشت گرادیان این تابع، لیپ‌شیتز با ثابت* L *می‌باشد. یعنی برای هر* $x, y \in \mathbb{R}^n$

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\|$$

ج) معمولاً مسأله‌ی (۱.۳) دارای جواب بهینه دگرین است. در اینجا بدنبال پیدا کردن جواب‌های بهینه‌ی مسأله‌ی (۱.۳) با مینیمم نرم اقلیدسی هستیم. مسأله‌ی زیر را در نظر می‌گیریم:

$$(Q) \quad : \quad \min\{\frac{۱}{۲}\|x\|^{۲} : x \in X^*\}, \tag{۲.۳}$$

که در آن X^* مجموعه جواب‌های بهینه‌ی (۱.۳) است. یک جواب بهینه‌ی (Q) را با x_Q^* نشان می‌دهیم.

برای بدست آوردن جواب مسأله‌ی Q، از مسأله‌ی شناخته شده‌ی تیخونوف استفاده می‌کنیم. بطور دقیق‌تر برای $\varepsilon > 0$ داده شده، مسأله‌ی محدبی بصورت زیر تعریف می‌کنیم:

$$(Q_\varepsilon): \quad \min\{f(x) + \frac{\varepsilon}{2}\|x\| : x \in X\} \tag{۳.۳}$$

چون $f(.) + \frac{\varepsilon}{2}\|.\|$ تابعی محدب اکید است، مسأله‌ی بالا در صورت داشتن جواب بهینه، دارای جواب بهینه‌ی یکتا می‌باشد که آن را با x^ε نشان می‌دهیم. در [۲۰]، تیخونوف نشان داد که در حالت خطی، یعنی زمانی که f خطی باشد و X اشتراک تعدادی متناهی نیم فضا باشد، آنگاه با $\varepsilon \longrightarrow 0^+$ داریم $x^\varepsilon \longrightarrow x_Q^*$، به علاوه برای هر $\varepsilon > 0$ که به اندازه کافی کوچک باشد، بردار x^ε می‌تواند به عنوان تقریبی از جواب حداقل نرم x_Q^* باشد. یک نتیجه‌ی قوی‌تر در حالت خطی آن است که می‌توانیم نشان دهیم، برای هر $\varepsilon > 0$ به قدر کافی کوچک x^ε دقیقا همان x_Q^* است [۱۵].

از نقطه نظر کاربردی، ارتباط بین جواب مسأله‌ی حداقل نرم و مسأله‌ی تیخونوف ایجاد کننده‌ی الگوریتم صریحی برای حل مسأله‌ی Q نیست. همچنین روش مشخص و کارایی برای انتخاب دنباله‌ی مناسب از پارامترهای

$\varepsilon_k \longrightarrow 0^+$ و حل زیرمسأله‌های ایجاد شده وجود ندارد.

اصلی‌ترین فعالیت این پروژه، ساختن و مطالعه‌ی روش مرتبه اول جدیدی برای حل مسأله‌ی Q می‌باشد، که **مسأله‌ی جواب حداقل نرم** یا MNP نامیده می‌شود.

مسأله‌ی MNP شامل یافتن جواب بهینه‌ی مسأله‌ی (۱.۳) است بطوریکه تابع قویاً محدب ω را مینیمم کند:

$$(MNP) \quad : \quad \min\{\omega(x) : x \in X^*\} \tag{۴.۳}$$

فرض می‌کنیم تابع ω یک تابع قویاً محدب با پارامتر $\sigma > 0$ است و بعلاوه این تابع به طور پیوسته مشتق‌پذیر می‌باشد.

با توجه به قویاً محدب بودن تابع ω، مسأله‌ی (MNP) دارای جواب بهینه‌ی یکتایی است که آن را با x_{mn}^*

نشان می‌دهیم.

برای ساده‌سازی، مسأله‌ی (۱.۳) را مسأله‌ی درونی و مسأله‌ی (MNP) را مسأله‌ی بیرونی می‌نامیم. بنابراین ω تابع هدف مسأله‌ی بیرونی می‌باشد.

بدیهی است که مسأله‌ی (Q) حالت خاصی از مسأله‌ی (MNP) با انتخاب $\omega(x) = \frac{1}{2}\|x\|^2$ می‌باشد. قرار می‌دهیم:

$$a \equiv arg_{x\in\mathbb{R}^n} \min \omega(x) \tag{۵.۳}$$

بدون کاستن از کلیت مسأله فرض می‌کنیم $\omega(a) = 0$. همچنین، با توجه به قویاً محدب بودن تابع $\omega(.)$، برای هر $x \in \mathbb{R}^n$ داریم

$$\omega(x) \geqslant \frac{\sigma}{2}\|x-a\|^2 \tag{۶.۳}$$

۲.۳ راه‌حل مرحله‌ای

باید توجه کنیم که مسأله‌ی (MNP) را می‌توان به شکل زیر نوشت، که این مسأله هم یک مسأله‌ی بهینه‌سازی محدب است:

$$\begin{aligned} &\min \omega(x) \\ &s.t. \quad f(x) \leqslant f^* \\ &x \in X \end{aligned} \tag{۷.۳}$$

البته مقدار بهینه‌ی مسأله‌ی درونی در ادامه شناخته شده نیست. راه‌حل پیشنهادی از دو مرحله بصورت زیر تشکیل شده است.

مرحله‌ی اول: یافتن مقدار بهینه‌ی مسأله‌ی درونی،

مرحله‌ی دوم : حل مسأله‌ی (۷.۳).

این روش دو مرحله‌ای، دارای مشکلاتی می‌باشد. اول آنکه چون مقدار بهینه‌ی f^* همیشه به طور دقیق بدست نمی‌آید، باعث اشتباه شدن جواب‌های شدنی مسأله‌ی بیرونی می‌شود و یا اینکه باعث می‌گردد مسأله‌ی بیرونی

ناشدنی شود. دوم آنکه حتی اگر بتوانیم مقدار f^* را بطور دقیق بدست بیاوریم، مسأله‌ی (۷.۳) ذاتاً شرایط اسلاتر را برآورده نمی‌کند و این یعنی این تقریب دو مرحله‌ای معمولاً مشکلات عددی ایجاد می‌کند. توجه داریم که عدم برقراری شرط اسلاتر برای مسأله‌ی (۷.۳)، دلیل بر عدم برقراری شرایط KKT در بهینگی نیست.

برای مثال، مرجع [۴] را ببینید که در آن شرایط KKT تحت فرضیات متفاوتی بدست آمده است.

قابل توجه است که مسأله‌ی (MNP) حالت خاصی از مسأله‌ی دو سطحی [۱] می‌باشد. برای مشاهده‌ی مثال‌های بیشتر به [۸] رجوع کنید.

۳.۳ طرح پروژه

این پروژه یک روش مرتبه اول معرفی می‌کند که به آن روش حداقل نرم گرادیان می‌گوییم. هدف حل مسأله‌ی حداقل نرم (MNP) است. برخلاف روش گفته شده در بخش قبل، روش پیشنهاد شده یک الگوریتم تکراری می‌باشد که مسأله‌ی (MNP) را مستقیماً حل می‌کند و بصورت مرحله به مرحله یا دنباله‌ای از مسائل بهینه‌سازی مرتبط با هم نمی‌باشد. بعلاوه، مطالبی پیرامون برخی ابزارهای ریاضی شامل تصویر متعامد، تصویر گرادیان و برش صفحه‌ها آورده‌ایم.

در ادامه، روش گرادیان حداقل نرم معرفی و بررسی شده است. در هر تکرار از این روش، محاسبات زیر را داریم:

۱. مقدار گرادیان برای تابع هدف دورنی،

۲. تصویر متعامد روی مجموعه‌ی شدنی مسأله‌ی درونی،

۳. یک جواب برای مسأله‌ی مینیمم‌سازی تابع هدفِ مسأله‌ی بیرونی، بطوریکه شامل فصل مشترک دو نیم فضا باشد.

نشان می‌دهیم، همگرایی دنباله‌ی تولید شده با روش گفته شده، از مرتبه‌ی $o(\frac{1}{\sqrt{k}})$ است.

[۱] Bilevel Programming

فصل ۴

الگوریتم ارائه‌شده

قبل از آنکه الگوریتم حداقل نرم گرادیان را توضیح دهیم، به معرفی تعدادی نماد می‌پردازیم که برای مینیمم کردن ω روی یک مجموعه‌ی بسته و محدب S بکار می‌روند.

$$\Omega(S) \equiv arg \min_{x \in S} \omega(x) \qquad (۱.۴)$$

با توجه به شرایط بهینگی در مسأله‌ی بالا، داریم

$$\tilde{x} = \Omega(S) \iff \langle \nabla\omega(\tilde{x}), x - \tilde{x} \rangle \geqslant ۰ \quad \forall x \in S \qquad (۲.۴)$$

هرگاه $\omega(x) = \frac{۱}{۲}\|x - a\|^۲$ آنگاه $\Omega(s) = P_S(a)$. اکنون آماده‌ایم تا الگوریتم را با ثابت لیپ‌شیتز داده شده‌ی L معرفی کنیم.

۱.۴ روش حداقل نرم گرادیان (با ثابت لیپ شیتز)

- ***ورودی :*** Lـ ثابت لیپ‌شیتز برای ∇f
- ***مقدار آغازین:*** $x_۰ = a$
- ***مرحله‌ی kـ ام*** $k = ۱, ۲, \cdots$

$$x_k = \Omega(Q^k \cap \omega^k)$$

بطوریکه

$$Q^k = Q_{L,\beta,x_{k-1}}$$
$$\omega^k = \{z \in \mathbb{R}^n : \langle \nabla\omega(x_{k-1}), z - x_{k-1}\rangle \geqslant 0\}$$

برای β داریم

$$\beta = \begin{cases} \frac{4}{3} & X \neq \mathbb{R}^n \\ 1 & X = \mathbb{R}^n \end{cases}$$

هرگاه ثابت لیپ‌شیتز داده نشده باشد، باید از روش $backtracking$ استفاده کنیم.

۲.۴ روش حداقل نرم گرادیان (بدون ثابت لیپ‌شیتز)

- ***ورودی :*** $L_0 > 0$ و $\eta > 1$
- ***مقدار آغازین:*** $x_0 = a$
- ***مرحله‌ی kـ ام*** $k = 1, 2, \cdots$ *کوچکترین عدد طبیعی نامنفی i_k را چنان پیدا کنید که*

$$\bar{L} = \eta^{i_k} L_{k-1}$$

و اگر نامساوی زیر برقرار باشد

$$\begin{aligned} f(T_{\bar{L}} - (x_{k-1})) \leqslant & f(x_{k-1}) + \langle \nabla f(x_{k-1}), T_{\bar{L}}(x_{k-1}) - x_{k-1}\rangle \\ & + \frac{\bar{L}}{2}\|T_{\bar{L}}(x_{k-1}) - x_{k-1}\|^2 \end{aligned}$$

قرار دهید $L_k = \bar{L}$

قرار دهید:

$$x_k = \Omega(Q^k \cap W^k), \tag{۳.۴}$$

که در آن

$$Q^k = Q_{L,\beta,x_{k-1}}, \qquad W^k = \{z \in \mathbb{R}^n : \langle \nabla\omega(x_{k-1}), z - x_{k-1}\rangle \geqslant 0\}, \tag{۴.۴}$$

اکنون برای طول گام ثابت قرار می‌دهیم:

$$L_k = L, \quad \forall k \quad \eta = 1 \tag{۵.۴}$$

در این تعریف نیم فضای Q^k را در طول گام ثابت و همچنین در backtracking می‌توانیم بصورت زیر تعریف کنیم:

$$Q^k = Q_{L_k,\beta,x_{k-1}} \tag{۶.۴}$$

که داریم

$$\beta = \begin{cases} \frac{4}{3} & X \neq \mathbb{R}^n \text{ ثابت لیپ‌شیتز مشخص} \\ 1 & X = \mathbb{R}^n \text{ ثابت لیپ‌شیتز مشخص} \\ 2 & X = \mathbb{R}^n \text{ ثابت لیپ‌شیتز نامشخص} \end{cases} \tag{۷.۴}$$

از طرفی می‌توانیم نشان دهیم، هنگامیکه داریم

$$L_0 \leqslant L_k \leqslant \eta L \quad k = 0, 1, 2, \cdots \tag{۸.۴}$$

می‌توانیم نشان دهیم

$$\|G_{L_0}(x)\| \leqslant \|G_{L_k}(x)\| \leqslant \|G_{\eta L}(x)\| \tag{۹.۴}$$

برهان. برای اثبات عبارت فوق نشان می‌دهیم

$$L_1 \geqslant L_2 \Longrightarrow \|G_{L_1}(x)\| \geqslant \|G_{L_2}(x)\|$$

بنابر قضیه داریم که

$$\langle \nu - P_C(\nu), P_C(\nu) - \omega\rangle \geqslant 0$$

قرار می‌دهیم

$$\nu = x - \frac{۱}{L_۱}\nabla f(x)$$
$$\omega = P_C(x - \frac{۱}{L_۲}\nabla f(x))$$

بنابر نامساوی گفته شده، داریم

(۱۰.۴)
$$\langle x - \frac{۱}{L_۱}\nabla f(x) - P_C(x - \frac{۱}{L_۱}\nabla f(x)), P_C(x - \frac{۱}{L_۱}\nabla f(x)) - P_C(x - \frac{۱}{L_۲}\nabla f(x))\rangle \geqslant ۰$$

$$\Longrightarrow \langle \frac{۱}{L_۱}G_{L_۱}(x) - \frac{۱}{L_۱}\nabla f(x), \frac{۱}{L_۲}G_{L_۲}(x) - \frac{۱}{L_۱}G_{L_۱}(x)\rangle \geqslant ۰ \qquad (۱۱.۴)$$

با تعویض جای $L_۱$ و $L_۲$ داریم

$$\langle \frac{۱}{L_۲}G_{L_۲}(x) - \frac{۱}{L_۲}\nabla f(x), \frac{۱}{L_۱}G_{L_۱}(x) - \frac{۱}{L_۲}G_{L_۲}(x)\rangle \geqslant ۰ \qquad (۱۲.۴)$$

اکنون نامساوی (۱۱.۴) را در $L_۱$ و نامساوی (۱۲.۴) را در $L_۲$ ضرب می‌کنیم سپس با هم جمع می‌کنیم، داریم

$$\langle G_{L_۱}(x) - G_{L_۲}(x), \frac{۱}{L_۲}G_{L_۲}(x) - \frac{۱}{L_۱}G_{L_۱}(x)\rangle \geqslant ۰$$

اکنون داریم

$$\frac{۱}{L_۱}\|G_{L_۱}(x)\|^۲ + \frac{۱}{L_۲}\|G_{L_۲}(x)\|^۲ \leqslant (\frac{۱}{L_۱} + \frac{۱}{L_۲})G_{L_۱}(x)^T G_{L_۲}(x)$$

با استفاده از نامساوی کوشی شوارتز داریم

$$\frac{۱}{L_۱}\|G_{L_۱}(x)\|^۲ + \frac{۱}{L_۲}\|G_{L_۲}(x)\|^۲ \leqslant (\frac{۱}{L_۱} + \frac{۱}{L_۲})\|G_{L_۱}(x)\|\|G_{L_۲}(x)\| \qquad (۱۳.۴)$$

اگر داشته باشیم $G_{L_۱}(x) = ۰$، آنگاه بنابر رابطه اخیر داریم $G_{L_۲}(x) = ۰$. اکنون فرض می‌کنیم $G_{L_۲}(x) \neq ۰$ آنگاه تعریف می‌کنیم

$$t = \frac{\|G_{L_۱}(x)\|}{\|G_{L_۲}(x)\|}$$

با استفاده از رابطه‌ی (۱۳.۴) داریم

$$\begin{aligned}
&\frac{1}{L_1}t^2 - (\frac{1}{L_1} + \frac{1}{L_2})t + \frac{1}{L_2} \leqslant 0 \\
&\Longrightarrow \quad 1 \leqslant t \leqslant \frac{L_1}{L_2} \\
&\Longrightarrow \quad 1 \leqslant \frac{\|G_{L_1}(x)\|}{\|G_{L_2}(x)\|} \\
&\Longrightarrow \|G_{L_1}(x)\| \geqslant \|G_{L_2}(x)\|
\end{aligned}$$

□

مثال ۱.۲.۴. *[۲] ω را با نرم اقلیدسی بصورت $\omega = \frac{1}{2}\|.\|^2$ تعریف می‌کنیم و داریم $\Omega(S) = P_S$ و برای گام اصلی داریم*

$$x_k = \Omega(Q^k \cap \omega^k)$$

تا اینکه تصویر متعامد روی فصل مشترک دو نیم فضا را بیابیم. بطوریکه تصویر متعامد روی فصل مشترک نیم‌فضاها را بصورت زیر داریم

$$T = \{x \in \mathbb{R}^n : \langle a_1, x\rangle \leqslant b_1, \langle a_2, x\rangle \leqslant b_2\} \quad \forall a_1, a_2 \in \mathbb{R}^n, \forall b_1, b_2 \in \mathbb{R}$$

و برای $P_T(x)$ داریم

$$P_T(x) = \begin{cases} x & : \ \alpha \leqslant 0, \beta \leqslant 0 \\ x - (\frac{\beta}{\nu})a_2 & : \ \alpha \leqslant \pi(\frac{\beta}{\nu}), \beta > 0 \\ x - (\frac{\alpha}{\mu})a_1 & : \ \beta \leqslant \pi(\frac{\alpha}{\mu}), \alpha > 0 \\ x + (\frac{\alpha}{\rho})(\pi a_2 - \nu a_1) + (\frac{\beta}{\rho})(\pi a_1 - \mu a_2) & : \ otherwise \end{cases}$$

بطوریکه داریم

$$\pi = \langle a_1, a_2\rangle, \mu = \|a_1\|^2, \nu = \|a_2\|^2, \rho = \mu\nu - \pi^2$$

$$\alpha = \langle a_1, x\rangle - b_1, \quad \beta = \langle a_2, x\rangle - b_2$$

لم ۲.۲.۴. [۲] *فرض کنید دنباله‌ی $\{x_k\}_{k\geqslant 0}$ به وسیله‌ی روش حداقل نرم گرادیان توسط یک طول گام ثابت یا $backtracking$ تولید شده باشد. آنگاه*

$$X^* \subseteq Q^k \cap \omega^k \quad \forall k = 1, 2, \cdots \tag{۱۴.۴}$$

برهان. با توجه به لم‌های ۳.۵.۲ و ۴.۵.۲، برای هر $k = 1, 2, \cdots$ داریم $X^* \subseteq Q^k$ و اکنون با استقرا روی k نشان می‌دهیم که $X^* \subseteq \omega^k$.

برای $k = 1$ داریم $\omega^1 = \mathbb{R}^n$ و حکم بدیهی است. اکنون فرض می‌کنیم حکم برای $k = n$ برقرار باشد، یعنی $X^* \subseteq Q^n \cap \omega^n$.

فرض می‌کنیم $X^* \subseteq \omega^n$، برای اثبات $X^* \subseteq Q^{n+1} \cap \omega^{n+1}$ ابتدا $u \in X^*$ را در نظر می‌گیریم، بطوریکه $X^* \subseteq Q^n \cap \omega^n$، داریم $x_n = \Omega(Q^n \cap \omega^n)$.

بنابر رابطه‌ی (۲.۴) داریم

$$\langle \nabla\omega(x_n), x_n - u\rangle \leqslant 0$$
$$\Longrightarrow u \in \omega^{n+1}$$

پس $X^* \subseteq Q^k \cap \omega^k$ برای هر $k = 1, 2, \cdots$. $\square$

فصل ۵

تحلیل همگرایی

اولین هدف ما در این فصل آن است که نشان دهیم روش حداقل نرم گرادیان، دنباله‌ای مانند $\{x_k\}_{k\geqslant ۰}$ تولید کند بطوریکه به $x^*_{mn}=\Omega(X^*)$ همگرا باشد.

قضیه ۱.۰.۵ (همگرایی دنباله). [۲] *فرض کنید دنباله‌ی $\{x_k\}_{k\geqslant ۰}$ به وسیله‌ی روش حداقل نرم گرادیان با طول گام ثابت یا $backtracking$ تولید شده باشد، آنگاه گزاره‌های زیر برقرار است.*

۱. *دنباله‌ی $\{x_k\}_{k\geqslant ۰}$ کراندار است،*

۲. *نامساوی زیر برای هر $k=۱,۲,\cdots$ برقرار است.*

$$D_\omega(x_k,x_{k-۱})+D_\omega(x_{k-۱},a)\leqslant D_\omega(x_k,a) \qquad (۱.۵)$$

۳. *هرگاه $k\longrightarrow\infty$ آنگاه $x_k\longrightarrow x^*_{mn}$.*

برهان. قسمت (۱) : هرگاه $x_k=\Omega(Q^k\cap\omega^k)$، برای هر $u\in Q^k\cap\omega^k$ و بعلاوه $u\in X^*$ داریم

$$\omega(x_k)\leqslant\omega(u)$$

و با استفاده از رابطه‌ی (۶.۳) کرانداری $\{x_k\}_{k\geqslant ۰}$ نتیجه می‌شود.

قسمت (۲) : داریم

$$D_\omega(x_k,x_{k-۱})+D_\omega(x_{k-۱},a)-D_\omega(x_k,a)=\langle-\nabla\omega(x_{k-۱}),x_k-x_{k-۱}\rangle$$

با توجه به تعریف ω^k داریم

$$x_{k-1} = \Omega(\omega^k)$$

بعلاوه $x_k \in \omega^k$ و با توجه به رابطه‌ی (۲.۴) داریم

$$\langle \nabla\omega(x_{k-1}), x_k - x_{k-1} \rangle \geqslant 0$$

قسمت (۳) : برای هر $x \in \mathbb{R}^n$ داریم $D_\omega(x,a) = \omega(x)$. در نتیجه $\{\omega(x_k)\}_{k\geqslant 0} = \{D_\omega(x_t,a)\}_{k\geqslant 0}$ که نانزولی و کراندار می‌باشد. بنابراین $\lim_{k\longrightarrow\infty} \omega(x_k)$ وجود دارد.

بنابراین داریم

$$\lim_{x\longrightarrow\infty} D_\omega(x_k, x_{k-1}) = 0$$

، بنابراین عبارت زیر برقرار است.

$$D_\omega(x_k, x_{k-1}) \geqslant \frac{\delta}{2}\|x_k - x_{k-1}\|^2$$

بطوریکه

$$\lim_{k\longrightarrow\infty} \|x_k - x_{k-1}\| = 0 \tag{۲.۵}$$

هنگامیکه $x_k \in Q^k$ می‌باشد، داریم

$$\langle G_{L_k}(x_{k-1}), x_{k-1} - x_k \rangle \geqslant \frac{1}{\beta L_k}\|G_{L_k}(x_{k-1})\|^2$$

حال با توجه به کوشی شوارتز داریم

$$\frac{1}{\beta L_k}\|G_{L_k}(x_{k-1})\| \leqslant \|x_{k-1} - x_k\|$$

اکنون بنابر رابطه‌ی (۸.۴) و (۹.۴) داریم

$$\frac{1}{\eta L}\|G_{L_0}(x_{k-1})\| \leqslant \|x_{k-1} - x_k\|. \tag{۳.۵}$$

برای اینکه نشان دهیم $\{x_k\}_{k\geqslant 0}$ به x^*_{mn} همگراست، کافی است نشان دهیم هر زیر دنباله‌ی همگرا از آن به x^*_{mn} میل می‌کند. بنابراین $\{x_{k_n}\}_{n\geqslant 0}$ یک زیر دنباله‌ی همگرا می‌باشد که حد آن برابر با c است.بنابر روابط

(۲.۵) و (۳.۵) و پیوستگی $G_{L_\circ}$، داریم که $G_{L_\circ}(c) = \circ$ است. بنابراین، $\omega \in X^*$.

باید نشان دهیم

$$c = \Omega(X^*) = X^*_{mn}$$

بطوریکه

$$x_{k_n} = \Omega(Q^{k_n} \cap \omega^{k_n})$$

بنابر رابطه‌ی (۲.۴)، داریم

$$\langle \nabla\omega(x_{k_n}), z - x_{k_n} \rangle \geqslant \circ \quad \forall z \in Q^{k_n} \cap \omega^{k_n}$$

بنابراین داریم

$$\langle \nabla\omega(x_{k_n}), z - x_{k_n} \rangle \geqslant \circ \quad \forall z \in X^*$$

اگر $n \longrightarrow \infty$ و بنابر پیوستگی $\nabla\omega$ داریم

$$\langle \nabla\omega(c), z - c \rangle \geqslant \circ \quad \forall z \in X^*$$

به علاوه بنابر رابطه‌ی (۴.۲) داریم

$$c = \Omega(X^*) = X^*_{mn}.$$

□

قضیه ۲.۰.۵ (مرتبه همگرایی). *فرض کنید $\{x_k\}_{k\geqslant\circ}$ دنباله‌ی تولید شده توسط روش حداقل نرم گرادیان با یک طول گام ثابت یا backtracking باشد، آنگاه برای هر $k \geqslant ۱$ داریم*

$$\min_{۱\leqslant n\leqslant k} f(T_{L_n}(x_n)) - f^* \leqslant \frac{\beta\eta L\|a - x^*_{mn}\|^۲}{\sqrt{k}} \tag{۴.۵}$$

هنگامیکه β داده شده باشد، در قسمت (۷.۴).

اگر $X = \mathbb{R}^n$ آنگاه

$$\min_{۱\leqslant n\leqslant k} f(x_n) - f^* \leqslant \frac{\beta\eta L\|a - x^*_{mn}\|^۲}{\sqrt{k}} \tag{۵.۵}$$

برهان. به [۲] رجوع کنید. □

کتاب‌نامه

[1] A. Beck, *Convergence Rate Analysis of Gradient Based Algorithms*, PhD thesis, School of Mathematical Sciences, Technion–Israel Institute of Technology, (2003).

[2] A. Beck, *A first order method for finding minimal norm-like solutions of convex optimization problems*, Springer, (471): 25-46, (2013).

[3] A. Beck, *Introduction to Nonlinear Optimization Theory, Algorithms, And Applications With MATLAB*, SIAM, (2014).

[4] A. Ben-Israel, A. Ben-Tal, and S. Zlobec, *Optimality in convex programming: a feasible directions approach*, Optimality and stability in mathematical programming, (19): 16–38, (1982).

[5] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPS-SIAM Series on Optimization, (2001).

[6] D. P. Bertsekas, *Nonlinear Programming*, Belmont MA: Athena Scientific, Second edition, (1999).

[7] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM Journal on Optimization, 3(3): 538–543, (1993).

[8] B. Colson, P. Marcotte, G. Savard, *Bilevel programming: a survey*, 4OR, (3): 87–107, (2005).

[9] M. C. Ferris and O. L. Mangasarian, *Finite perturbation of Convex Programs*, Computer Science Tech. Rep. 802, (1991).

[10] M.P. Friedlander, P. Tseng, *Exact regularization of convex programs*, SIAM J. Optim., 18(4): 1326–1350, (2007).

[11] G.H. Golub, M. Heath, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21(2): 215–223, (1979).

[12] P.C. Hansen, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci.Stat. Comput., (14): 1487–1503, (1993).

[13] P.C. Hansen, *Regularization tools: a Matlab package for analysis of discrete regularization problems*, Numer. Algorithms, (6): 1–35, (1994).

[14] H. Qi C. Kanzow and L. Qi, *On the minimum norm solution of linear programs*, Journal of Optimization Theory and Applications, (312): 333-345, (2003).

[15] O. L. Mangasarian and R. R. Meyer, *Nonlinear perturbation of linear programs*, SIAM Journal on Control and Optimization, (17): 745-757, (1979).

[16] H. Markowitz, *Portfolio selection. J. Finance*, (7): 77–91, (1952).

[17] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity And Method Efficiency In Optimization*, A Wiley-Interscience Publication. John Wiley Sons Inc., New York, (1983).

[18] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer, Boston, (2004).

[19] D.L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind.*, J. Assoc. Comput. Mach., (9): 84–97, (1962).

[20] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, DC: V.H. Winston, (1977).

University of Tehran

School of Mathematics, Statistics and Computer Science

Minimal norm solutions in convex optimization

By:

Alireza Tirekar

Under Supervision of:

Dr. Majid Soleimani-damaneh

A Dissertation Submitted to the Graduate Studies Office

In Partial Fulfillment of the Requirements for

the Degree of B.Sc. in

Applied Mathematics

July 2017