

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

برخی کاربردهای بهینه‌سازی محدب قوی

نگارش

المیرا فتحی پسندیده

استاد راهنما

دکتر مجید سلیمانی دامنه

پایان‌نامه برای دریافت درجه کارشناسی
در رشته ریاضیات و کاربردها

تیر ۱۳۹۶

تقدیم بہ

پدرم کہ عالمانہ بہ من آموخت تا چگونہ در عرصہ زندگی، ایستادگی را تجربہ نمایم
و بہ مادرم، دریای بی کران فداکاری و عشق کہ وجودم برایش ہمہ رنج بود و وجودش برایم ہمہ مهر.

سپاسگزارى...

مشكر و سپاس يكران دارم خدمت استاد كرامت، جناب آقاى دكتر سليمانى.

الميرا فتحى پسندیده

تیر ۱۳۹۶

چکیده

در این پروژه به برخی کاربردهای بهینه‌سازی محدب خاص، در مسائل انتخاب سبد سهام، جواب‌های معادلات انتگرالی، جداسازی و کمترین مربعات می‌پردازیم. علاوه بر این، به دنبال مطالعه‌ی برخی روش‌های عددی روی مسائل بهینه‌سازی با تابع هدف محدب هستیم. مطالعه‌ی همگرایی، نرخ همگرایی و رفتار جواب، از اهداف دیگر این مطالعه می‌باشد. در این پروژه، از ابزارهایی مانند الگوریتم کاهش، الگوریتم حداقل نرم گرادیان و تابع تصویر بهره خواهیم برد.

واژگان کلیدی: بهینه‌سازی محدب، جواب حداقل نرم، انتخاب سهام

فهرست مطالب

۱	بهبودسازی محدب	۱
۱	۱.۱ تابع محدب قوی	۱
۲	۲.۱ بهبودسازی محدب	۲
۵	۳.۱ تابع تصویر متعامد	۵
۹	۴.۱ شرایط KKT	۹
۱۲	۲ کاربردهای بهبودسازی محدب	۱۲
۱۲	۱.۲ دسته‌بندی بردارها با استفاده از جداکننده‌ی خطی	۱۲
۲۳	۲.۲ بهبودسازی ترکیبیاتی	۲۳
۲۳	۳.۲ مسأله کمترین مربعات	۲۳
۲۵	۴.۲ مسئله کمترین مربعات خطی تنظیم شده (RLS)	۲۵
۲۶	۵.۲ برازش دایره	۲۶
۲۹	۳ چند ابزار پایه‌ای	۲۹
۲۹	۱.۳ ترازهای برشی	۲۹
۳۰	۲.۳ تصویر گرادیان	۳۰
۳۱	۱.۲.۳ ثابت لیپ‌شیتز L معلوم باشد	۳۱
۳۳	۲.۲.۳ ثابت لیپ‌شیتز L معلوم نباشد	۳۳

۳۵	۴	الگوریتم حداقل نرم گرادیان
۳۵	۱.۴	روش حداقل نرم گرادیان (با ثابت لیپ شیتز)
۳۶	۲.۴	روش حداقل نرم گرادیان (بدون ثابت لیپ شیتز)
۳۸	۳.۴	تحلیل همگرایی
۳۹	۵	مثال‌های کاربردی
۳۹	۱.۵	مسئله‌ی انتخاب سهام
۴۳	۲.۵	جواب‌های معادلات انتگرالی
۴۷		مراجع

فصل ۱

بهینه‌سازی محدب

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 $f(x) - \frac{\sigma}{4}\|x\|^2$ 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

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{\sigma}{4}\lambda(1 - \lambda)\|x - y\|^2 \quad (1.1)$$

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

□ برهان. به [۳] رجوع کنید.

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

□ برهان. به [۳] رجوع کنید.

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

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\sigma}{2} \|x - y\|^2, \quad \forall x, y \in C \quad (2.1)$$

□ برهان. به [۳] رجوع کنید.

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

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \sigma \|x - y\|^2.$$

□ برهان. به [۳] رجوع کنید.

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

$$\nabla^2 f(x) \geq \sigma I, \quad \forall x \in C \quad (3.1)$$

□ برهان. به [۳] رجوع کنید.

۲.۱ بهینه‌سازی محدب

ابتدا مسأله‌ی برنامه‌ریزی محدب را بصورت زیر تعریف می‌کنیم.

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

$$\min f(x) \quad (4.1)$$

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

قضیه ۲.۲.۱. [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 (۴.۱) باشد. پس $\varepsilon > 0$ وجود دارد، بطوریکه $f(\bar{x}) \leq f(x)$ 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

$$\lambda \in \left(0, \min \left\{ 1, \frac{\varepsilon}{\|x^* - \bar{x}\|} \right\} \right).$$

چون C محدب است، داریم

$$x^\circ = \lambda x^* + (1 - \lambda)\bar{x} \in C$$

به علاوه

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

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

$$\begin{aligned} f(\bar{x}) &\leq f(x^\circ) = f(\lambda x^* + (1 - \lambda)\bar{x}) \\ &\leq \lambda f(x^*) + (1 - \lambda)f(\bar{x}) \\ &< \lambda f(\bar{x}) + (1 - \lambda)f(\bar{x}) \\ &= f(\bar{x}) \end{aligned} \tag{۵.۱}$$

در نتیجه

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

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

قضیه ۳.۲.۱. در یک مسأله‌ی برنامه‌ریزی محدب، اگر f اکیداً محدب باشد، آنگاه هر جواب بهینه‌ی موضعی، یک جواب بهینه‌ی سراسری یکتا می‌باشد.

برهان. مشابه قضیه ۲.۲.۱ است. □

قضیه ۴.۲.۱. [۳] مجموعه جواب‌های بهینه‌ی یک مسأله‌ی برنامه‌ریزی محدب، مجموعه‌ای محدب است. برهان. فرض کنید O مجموعه‌ی جواب‌های بهینه‌ی مسأله‌ی برنامه‌ریزی محدب (۴.۱) باشد. حال فرض کنید

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

آنگاه داریم

$$\bar{x} \in O_p \implies f(\bar{x}) \leq f(x), \quad \forall x \in C$$

$$x^\circ \in O_p \implies f(x^\circ) \leq f(x), \quad \forall x \in C$$

با در نظر گرفتن $\lambda \in [0, 1]$ ، داریم

$$f(\lambda\bar{x} + (1-\lambda)x^\circ) \leq \lambda f(\bar{x}) + (1-\lambda)f(x^\circ) \leq f(x), \quad \forall x \in C$$

از طرفی چون C محدب است، داریم

$$\lambda\bar{x} + (1-\lambda)x^\circ \in C$$

پس

$$\lambda\bar{x} + (1-\lambda)x^\circ \in O_p$$

□

قضیه ۵.۲.۱. مجموعه جواب‌های بهینه‌ی مسأله‌ی بهینه‌سازی زیر، که در آن C مجموعه‌ای محدب و f تابعی اکیداً محدب است، تهی یا تک عضوی است.

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

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

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

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

$$f\left(\frac{1}{4}x + \frac{1}{4}y\right) < \frac{1}{4}f(x) + \frac{1}{4}f(y) = f(x)$$

در حالی که می‌دانیم، $k := \frac{1}{4}x + \frac{1}{4}y \in C$ ، پس نامساوی اکید فوق با مینیمم بودن $f(x)$ در تناقض است. \square

۳.۱ تابع تصویر متعامد

تعریف ۱.۳.۱. [۳] فرض کنیم $S \subseteq \mathbb{R}^n$ ناتهی، بسته و محدب باشد، آنگاه تابع فاصله‌ی متناظر با S را بصورت زیر تعریف می‌کنیم.

$$d_S : \mathbb{R}^n \longrightarrow \mathbb{R} \quad (۶.۱)$$

$$d_S(x) = \inf_{y \in S} \|x - y\|$$

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 ۲.۳.۱. [۳] برای $S \subseteq \mathbb{R}^n$ ناتهی، بسته و محدب، عملگر تصویر متعامد به صورت زیر تعریف می‌گردد.

$$P_S : \mathbb{R}^n \longrightarrow \mathbb{R}^n \quad (۷.۱)$$

$$P_S(x) = \arg \min_{y \in S} \|x - y\|^2$$

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

عملگر تصویر متعامد دارای خواص زیر است.

قضیه ۳.۳.۱. برای هر $x \in \mathbb{R}^n$ و $y \in S$ داریم

$$\langle x - P_S(x), y - P_S(x) \rangle \leq 0 \quad (۸.۱)$$

برهان. طبق تعریف ۲.۳.۱، 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

$$\nabla h(P_S(x))^T (y - P_S(x)) \geq 0, \quad \forall y \in S$$

$$\xrightarrow{\nabla h(y) = 2(y-x)} 2(P_S(x) - x)^T (y - P_S(x)) \geq 0, \quad \forall y \in S$$

$$\implies \langle (x - P_S(x)), (y - P_S(x)) \rangle \leq 0, \quad \forall y \in S$$

□

قضیه ۴.۳.۱. برای هر $x, y \in \mathbb{R}^n$ داریم

$$\langle P_S(x) - P_S(y), x - y \rangle \geq \|P_S(x) - P_S(y)\|^2 \quad (۹.۱)$$

برهان. بنابر قضیه قبل

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

با قرار دادن $Z = P_S(y)$ ، داریم

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

با قرار دادن $Z = P_S(x)$ ، داریم

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

با جمع دو رابطه‌ی اخیر داریم

$$\begin{aligned} & (-x + P_S(x) + y - P_S(y))^T (P_S(x) - P_S(y)) \leq 0 \\ \implies & (x - y)^T (P_S(x) - P_S(y)) \geq (P_S(x) - P_S(y))^T (P_S(x) - P_S(y)) \\ \implies & \langle P_S(x) - P_S(y), x - y \rangle \geq \|P_S(x) - P_S(y)\|^2. \end{aligned}$$

□

قضیه ۵.۳.۱. برای هر $x, y \in \mathbb{R}^n$ داریم

$$\|P_S(x) - P_S(y)\| \leq \|x - y\| \quad (۱۰.۱)$$

برهان. اگر $\|P_S(x) - P_S(y)\| = 0$ باشد، حکم بدیهی است. بنابراین فرض کنید

$$\|P_S(x) - P_S(y)\| \neq 0.$$

بنابر قضیه قبل، داریم

$$\begin{aligned} \|P_S(x) - P_S(y)\|^2 & \leq (P_S(x) - P_S(y))^T (x - y) \\ & \stackrel{\text{کوشی شوارتز}}{\implies} \|P_S(x) - P_S(y)\|^2 \leq \|P_S(x) - P_S(y)\| \|x - y\| \\ & \stackrel{\div \|P_S(x) - P_S(y)\|}{\implies} \|P_S(x) - P_S(y)\| \leq \|x - y\| \end{aligned}$$

□

مثال ۶.۳.۱. فرض کنید $C = B[\bar{x}, r]$ ، $x \in \mathbb{R}^n$ و $r > 0$ داده شده‌اند. می‌خواهیم یک رابطه برای

عملگر تصویر متعامد P_C پیدا کنیم. داریم:

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

و

$$(P) : \quad \min \|y - x\|^2$$

$$s.t. \quad \|y - \bar{x}\| \leq r$$

پس

$$\text{if } x \in C \implies P_C(x) = x$$

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

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

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

آنگاه

$$\nabla h(y^*) = 0$$

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

$$(q) : \quad \min \|y - x\|^2 \\ \text{s.t.} \quad \|y - \bar{x}\| = r$$

$$\begin{aligned} h(y) &= \|y - \bar{x} + \bar{x} - x\|^2 \\ &= \|y - \bar{x}\|^2 + \|\bar{x} - x\|^2 + 2(y - \bar{x})^T(\bar{x} - x) \\ &= r^2 + \|x - \bar{x}\|^2 + 2(y - \bar{x})^T(\bar{x} - x) \end{aligned}$$

که $r^2 + \|x - \bar{x}\|^2$ ثابت می‌باشد، پس به دنبال بهینه کردن $2(y - \bar{x})^T(\bar{x} - x)$ می‌باشیم. بنابراین داریم

$$(Z) : \quad \min 2(y - \bar{x})^T(\bar{x} - x) \\ \text{s.t.} \quad \|y - \bar{x}\| = r$$

بنابر نامساوی کوشی شواتز داریم

$$2(y - \bar{x})^T(\bar{x} - x) \geq -2\|y - \bar{x}\|\|\bar{x} - x\| = -2r\|\bar{x} - x\|$$

پس $\| \bar{x} - x \|$ 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

$$\nabla \left(\bar{x} + \frac{r(x-\bar{x})}{\|x-\bar{x}\|} - \bar{x} \right)^T (\bar{x} - x) = -\nabla r \frac{\|\bar{x} - x\|^2}{\|\bar{x} - x\|} = -\nabla r \|\bar{x} - x\|$$

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

۴.۱ شرایط KKT

تعریف ۱.۴.۱. [۳] نقاط KKT : در مسأله‌ی مینیمم‌سازی زیر

$$\min f(x)$$

$$s.t. \quad g_i(x) \leq 0 \quad i = 1, 2, \dots, m \quad (11.1)$$

$$h_j(x) = 0 \quad j = 1, 2, \dots, p$$

هرگاه توابع $f, g_1, g_2, \dots, g_m, h_1, \dots, h_p$ روی \mathbb{R}^n بطور پیوسته مشتق‌پذیر باشند، نقطه‌ی شدنی x^* را یک نقطه‌ی KKT گوئیم، هرگاه وجود داشته باشند $\lambda_1, \dots, \lambda_m \geq 0$ و $\mu_1, \mu_2, \dots, \mu_p \in \mathbb{R}$ بطوریکه

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^p \mu_j \nabla h_j(x^*) = 0$$

$$\lambda_i g_i(x^*) = 0 \quad i = 1, 2, \dots, m$$

قضیه ۲.۴.۱ (کافی بودن شرایط KKT برای حل مسائل برنامه‌ریزی محدب). [۳] فرض کنید x^* یک

جواب شدنی مسأله‌ی (۱۱.۱) باشد. همچنین f و g_1, \dots, g_m توابعی محدب و بطور پیوسته مشتق‌پذیر روی \mathbb{R}^n باشند و بعلاوه h_1, \dots, h_p توابعی آفینی باشند.

فرض کنید ضرایب $\lambda_1, \dots, \lambda_m \geq 0$ و $\mu_1, \dots, \mu_p \in \mathbb{R}$ وجود دارند، بطوریکه

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^p \mu_j \nabla h_j(x^*) = 0 \quad (12.1)$$

$$\lambda_i g_i(x^*) = 0 \quad i = 1, \dots, m$$

آنگاه x^* یک جواب بهینه‌ی مسأله‌ی (۱.۱) است.

برهان. نشان می‌دهیم به ازای $x \in \mathbb{R}^n$ ، $f(x) \geq f(x^*)$. تابع $S(\cdot)$ 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^* مینیمم‌کننده‌ی S روی \mathbb{R}^n می‌باشد. بنابراین

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

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

قضیه ۳.۴.۱ (لازم بودن شرایط KKT تحت شرایط اسلاتر). [۳] فرض کنید x^* یک جواب بهینه مسأله‌ی زیر باشد:

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

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

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

آنگاه \hat{x} یک نقطه KKT است.

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

$$\tilde{\lambda}_0 \nabla f(x^*) + \sum_{i=1}^m \tilde{\lambda}_i \nabla g_i(x^*) = 0 \quad (۳)$$

$$\tilde{\lambda}_i g_i(x^*) = 0 \quad i = 1, \dots, m$$

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

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

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 $i = 1, \dots, m$ داریم:

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

بنابراین داریم:

$$0 > \sum_{i=1}^m \tilde{\lambda}_i g_i(x^*) + \left(\sum_{i=1}^m \tilde{\lambda}_i \nabla g_i(x^*) \right)^T (\hat{x} - x^*) \quad (۱۳.۱)$$

که همگی $\tilde{\lambda}_i$ ها صفر می‌باشند که با توجه به رابطه‌ی (۱۳.۱) داریم $0 > 0$. بنابراین فرض خلف باطل و

□

حکم ثابت می‌شود.

فصل ۲

کاربردهای بهینه‌سازی محدب

۱.۲ دسته‌بندی بردارها با استفاده از جداکننده‌ی خطی

در علم داده‌کاوی با استفاده از داده‌های قبلی اطلاعات جدیدی به دست می‌آورند و یا با همان داده‌ها می‌توانند نتیجه‌ی آزمایشاتی را که هزینه زیادی دارند بدون انجام آزمایش پیش‌بینی کنند. یکی از روش‌های انجام روش‌های انجام این کار تکنیک کلاس‌بندی است.

یکی از این روش‌های دسته‌بندی، جداسازی با استفاده از یک جداساز خطی است.

فرض کنید دو دسته نقاط x^1, \dots, x^m و x^{m+1}, \dots, x^k را داریم و می‌خواهیم آن‌ها را بوسیله‌ی

ابریصفحه‌ای مانند $H_{p,\alpha} = \{x \in \mathbb{R}^n \mid p^T x = \alpha\}$ جدا کنیم. پس به دنبال α و P هستیم، بطوریکه

$$\begin{cases} P^T x^i - \alpha < 0, & i = 1, \dots, m \\ P^T x^i - \alpha > 0, & i = m+1, \dots, k \end{cases} \quad (1.2)$$

فرض کنیم چنین ابریصفحه‌ای وجود دارد. معمولاً بی‌نهایت از این ابریصفحه‌ها وجود دارد. از بین این

ابریصفحه‌ها به دنبال ابریصفحه‌ای هستیم که بیشترین فاصله را تا نقاط داشته باشد. فاصله‌ی نقطه‌ی \bar{x} تا

$$d(H_{p,\alpha}, \bar{x}) = \frac{|p\bar{x} - \alpha|}{\|p\|}$$

ابریصفحه‌ی $H_{p,\alpha}$ برابر است با

برای اثبات این ادعا مسأله‌ی بهینه‌سازی محدب زیر را حل می‌کنیم:

فرض کنیم $r(A_{m \times n}) = m$ و $C = \{x \in \mathbb{R}^n \mid Ax = b\}$ و $\bar{y} \in \mathbb{R}^n$ ثابت می‌کنیم

(الف)

$$P_C(\bar{y}) = \bar{y} - A^T(AA^T)^{-1}(A\bar{y} - b)$$

(ب)

$$\text{dist}(\bar{y}, c) = \|A^T(AA^T)^{-1}(A\bar{y} - b)\|$$

برهان قسمت (الف) : $P_C(\bar{y})$ جواب بهینه‌ی مسأله‌ی مینیم‌سازی زیر است.

$$\min \| \bar{y} - x \|^2$$

$$s.t. \quad Ax = b$$

داریم

$$h(x) = \| \bar{y} - x \|^2 = \| \bar{y} \|^2 + \| x \|^2 - 2x^T \bar{y}$$

$$\implies \nabla h(x) = 2\mathbf{I}x - 2\bar{y} = 0$$

h یک تابع *Coercive* و پیوسته است و ناحیه شدنی هم‌بسته است. پس مسئله جواب بهینه دارد.

$$KKT \text{ شرایط} : 2(\bar{y} - x) - 2A^T\mu = 0 \quad (2.2)$$

$$Ax = b \quad (3.2)$$

$$\stackrel{(2.2)}{\implies} x = \bar{y} - A^T\mu$$

$$\stackrel{(2.3)}{\implies} b = A\bar{y} - AA^T\mu$$

$$\implies \mu = (AA^T)^{-1}b + (AA^T)^{-1}A\bar{y}$$

$$\implies x = \bar{y} - A^T((AA^T)^{-1}A\bar{y} - (AA^T)^{-1}b)$$

$$= \bar{y} - A^T(AA^T)^{-1}[A\bar{y} - b]$$

تنها x فوق نقطه‌ی KKT است و لذا بهینه است.

با توجه به رابطه‌ی (۳.۲)، $dist(\bar{y}, c)$ مقدار بهینه مسئله‌ی مینیم‌سازی زیر است.

$$\begin{aligned} \min \quad & \|\bar{y} - x\|^2 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

که برابر است با

$$\|\bar{y} - [\bar{y} - A^T(AA^T)^{-1}(A\bar{y} - b)]\| = \|A^T(AA^T)^{-1}(A\bar{y} - b)\|$$

تا اینجا (الف) و (ب) ثابت شدند. حال برای اثبات

$$d(H_{p,\alpha}, \bar{x}) = \frac{|p\bar{x} - \alpha|}{\|p\|}$$

از مسئله‌ی بالا استفاده کرده طبق رابطه‌ی (۳.۲)، قرار می‌دهیم $A = p^T$. پس

$$p_H(\bar{y}) = \bar{y} - p(p^T p)^{-1}(p^T \bar{y} - \alpha)$$

$$\begin{aligned} dist(\bar{y}, H) &= \|p(p^T p)^{-1}(p^T \bar{y} - \alpha)\| \\ &= |p^T \bar{y} - \alpha| \frac{1}{\|p\|^2} \|p\| = \frac{|p^T \bar{y} - \alpha|}{\|p\|} \end{aligned}$$

به مسأله‌ی اصلی برمی‌گردیم. پس به دنبال p و α هستیم که در (۱.۲) صدق کنند و $d(H_{p,\alpha}, x^i)$ برای

$i = 1, \dots, m, \dots, k$ ماکزیمم شود. بدین منظور $\min_{1 \leq i \leq k} d(H_{p,\alpha}, x^i)$ را \max می‌کنیم.

یعنی بدترین حالت را ماکزیمم می‌کنیم پس با مسئله‌ی زیر روبرو هستیم:

$$\beta_1 = \max_{P,\alpha} \min_{1 \leq i \leq k} \frac{|Px^i - \alpha|}{\|P\|}$$

$$\text{s.t.} \quad P^T x^i - \alpha < 0, \quad i = 1, \dots, m$$

$$P^T x^i - \alpha > 0, \quad i = m + 1, \dots, k$$

که این مسئله معادل است با

$$\beta_{\Psi} = \max_{P, \alpha} \frac{1}{\|P\|}$$

$$s.t. \quad \min_{1 \leq i \leq k} |P^T x^i - \alpha| = 1$$

$$P^T x^i - \alpha < 0, \quad i = 1, \dots, m$$

$$P^T x^i - \alpha > 0, \quad i = m + 1, \dots, k$$

و معادل است با مسئله

$$\beta_{\Psi} = \min \|p\|^{\Psi}$$

$$s.t. \quad \min_{1 \leq i \leq k} |p^T x^i - \alpha| = 1$$

$$p^T x^i - \alpha \leq -1, \quad i = 1, \dots, m$$

$$p^T x^i - \alpha \geq 1, \quad i = m + 1, \dots, k$$

و در نهایت معادل است با مسئله‌ی بهینه‌سازی محدب زیر:

$$\beta_{\Psi} = \min \|p\|^{\Psi}$$

$$s.t. \quad p^T x^i - \alpha \leq -1 \quad i = 1, \dots, m$$

$$p^T x^i - \alpha \geq 1 \quad i = m + 1, \dots, k$$

کد متلب آن به صورت زیر است.

```
%rng('shuffle');
data1=10*rand(2,200);
data2=10*rand(2,200)-15;
```

```
xmin=min([data1(1,:) data2(1,)]);
```

```

xmax=max([data1(1,:) data2(1,:)]);

x=xmin:0.1:xmax;

W=gAlgorithm(data1,data2,20,4000);

display(W);

y=-W(1)/W(2)*x-W(3)/W(2);

scatter(data1(1,:),data1(2,:),40,'fill');

hold;

scatter(data2(1,:),data2(2,:),40,'fill');

%hold;

scatter(x,y,15,'fill');

%display(xmin);

%display(xmax);

function r=fitFunc(W, beta, data1, data2)

    c=0;

    [m n]=size(data1);

    for i=1:n

        if (W*data1(:,i)+beta>-1)

            c=1;

```

```

        break;

    end

end

cp=0;

[m n]=size(data2);

for i=1:n

    if (W*data2(:,i)+beta<1)

        cp=1;

        break;

    end

end

if (cp==0 && c==0)

    r=0.5*power(norm(W,2),2);

else

    r=-1;

end

end

end

function W=gAlgorithm(data1 , data2 , iteration , beginingNumber)

    [n m]=size(data1);

    G=100*rand(n+2,beginingNumber)-50;

    %display(G);

    past=0;

    pastPos=0;

```



```

for ite=0:iteration

    rng( 'shuffle ' );

    %making new generation

    G=newGeneration(G);

    %some mutation happend

    G=mutation(G);

    %evaluation of generation

    [n m]=size(G);

    n=n-2;

    %display(size(G));

    %display(m);

    i=1;

    while i<=m

        %display(i);

        %display(size(G));

        %display(G(1:n,i));

        r=fitFunc(G(1:n,i),G(n+1,i),data1,data2);

        if (r>=0)

```

```

         $G(n+2, i) = r;$ 

    else

         $G(:, i) = [];$ 

         $i = i - 1;$ 

         $[nn \ m] = \mathbf{size}(G);$ 

    end

     $i = i + 1;$ 

end

 $[Y, I] = \mathbf{sort}(G(n+2, :));$ 

 $GG = G(:, I);$ 

 $G = GG;$ 

 $check = \mathbf{mean}(G(n+2, :));$ 

 $[nn \ m] = \mathbf{size}(G);$ 

 $i = 1;$ 

while  $i \leq m$ 

    if  $(G(n+2, i) > 5 * check)$ 

         $G(:, i) = [];$ 

         $i = i - 1;$ 

         $[nn \ m] = \mathbf{size}(G);$ 

         $\%display('warning');$ 

    end

```

```

        i=i+1;
    end

    display(ite);
    %display(m);
    [val index]=min(G(n+2,:));

    display(val);
    W=G(:,index);
    %display(index);
    if ite==0
        past=val;
    end

    %display(pastPos);
    if (ite-pastPos>5)
        if (val==past)
            ite=iteration+1;
            break;
        end
    else
        if val~=past
            pastPos=ite;

```

```

        end

        past=val;

    end

end

end

end

function newG=mutation(G)

    [mm n]=size(G);

    newC=floor(n/3);

    pos=randi(newC,1,n);

    %display(pos);

    s=rand(1);

    rng(s);

    for i=1:newC

        if(pos(i)~=1)

            r1=100*(2*rand(1)-1);

            r2=2*rand(1)-1;

            posp=randi(1,1,mm-1);

            G(posp,pos(i))=r2*G(posp,pos(i))+r1;

            posp1=randi(1,1,mm-1);

```

```

    posp2=randi(1,1,mm-1);

    t=G(posp1, pos(i));

    G(posp1, pos(i))=G(posp2, pos(i));

    G(posp2, pos(i))=t;

    end

end

newG=G;

end

function newG=newGeneration(G)

    [m n]=size(G);

    newG=G;

    newChild=floor(n/5);

    pos=randi(newChild,1,n);

    for i=1:newChild-1

        if (pos(i)~=1)

            r1=2*rand(1)-1;

            r2=2*rand(1)-1;

            nG=(r1 *G(:, pos(i))+r2 *G(:, pos(i+1)))/(r1+r2);

            newG=[newG, nG];

        end

    end

end

end

```

۲.۲ بهینه‌سازی ترکیبیاتی

در ریاضیات کاربردی، بهینه‌سازی ترکیبیاتی موضوعی است که شامل یافتن یک شیء بهینه از یک مجموعه‌ی محدود از اشیاء است. در بسیاری از این مسائل، جست و جوی جامع امکان‌پذیر نمی‌باشد و با استفاده از روش‌های بهینه‌سازی، می‌توان چنین مسائلی را حل کرد.

بهینه‌سازی ترکیبیاتی یک شاخه از بهینه‌سازی ریاضی است که با پژوهش عملیاتی، نظریه الگوریتم و نظریه‌ی پیچیدگی محاسباتی در ارتباط است. این شاخه کاربردی مهم در زمینه‌های مختلفی، مانند هوش مصنوعی، یادگیری ماشین و مهندسی نرم‌افزار کاربرد دارد. برخی کاربردهای بهینه‌سازی ترکیبیاتی به صورت زیر است:

۱. توسعه بهترین شبکه هواپیمایی،
 ۲. مسئله برش،
 ۳. مسئله برنامه‌ریزی پرستار،
 ۴. مسئله فروشنده دوره‌گرد،
 ۵. تصمیم‌گیری در مورد تاکسی‌هایی که در ناوگان برای رسیدن به مقصد هستند.
- تعداد زیادی الگوریتم برای حل مسائل ترکیبیاتی طراحی شده است که بخش قابل توجهی از آن‌ها بر پایه‌ی کلاس‌های خاصی از مسائل بهینه‌سازی خطی است. برخی نمونه‌های مسائل بهینه‌سازی ترکیبیاتی که در این چارچوب قرار می‌گیرند عبارتند از: کوتاهترین مسیر و مینیمم هزینه در شبکه جریان.

۳.۲ مسأله کمترین مربعات

دستگاه $Ax = b$ را در نظر بگیرید، که در آن A ماتریسی $m \times n$ است. اگر $r(A) = m$ باشد، این دستگاه جواب دارد و فرم عمومی جواب بصورت زیر است.

$$\begin{pmatrix} x_B \\ x_N \end{pmatrix} = \begin{pmatrix} B^{-1}b - B^{-1}Nx_N \\ x_N \end{pmatrix} \quad (۴.۲)$$

در اینجا فرض می‌کنیم $r(A) = n \leq m$ که به این حالت فرامعین گویند. در این حالت ممکن است دستگاه جواب نداشته باشد. در چنین حالتی به بردار x که به ازای آن فاصله بین Ax و b مینیمم می‌شود بسنده می‌کنیم. به عبارتی مسئله‌ی زیر را حل می‌کنیم.

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 \quad (5.2)$$

در واقع اگر داشته باشیم

$$\begin{bmatrix} a^1 \\ \vdots \\ a^m \end{bmatrix}$$

که a^i سطر i -ام A است، آنگاه

$$Ax = \begin{bmatrix} a^1 x \\ \vdots \\ a^m x \end{bmatrix} \quad (6.2)$$

$$\Rightarrow \|Ax - b\|^2 = \left\| \begin{bmatrix} a^1 x \\ \vdots \\ a^m x \end{bmatrix} - \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} a^1 x - b_1 \\ \vdots \\ a^m x - b_m \end{bmatrix} \right\|^2 \quad (7.2)$$

$$= \sum_{i=1}^m (a^i x - b_i)^2 \quad (8.2)$$

پس مسئله‌ی (5.2) معادل است با

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m (a^i x - b_i)^2 \quad (9.2)$$

به همین دلیل مسئله‌ی بدست آمده (۹.۲) را یک مسئله‌ی کمترین مربعات خطی می‌نامیم. در ادامه می‌خواهیم مسئله (۷.۲) را بررسی کنیم. داریم:

$$\begin{aligned}\|Ax - b\|^2 &= (Ax - b)^T(Ax - b) \\ &= (x^T A^T - b^T)(Ax - b) \\ &= x^T A^T Ax - x^T A^T b - b^T Ax + b^T b \\ &= x^T (A^T A)x + 2(-A^T b)^T x + \|b\|^2\end{aligned}$$

ملاحظه می‌شود که تابع هدف مسئله‌ی (۵.۲) یک تابع درجه دوم است. نقاط ایستای این تابع درجه‌ی دوم جواب‌های دستگاه $A^T Ax = A^T b$ هستند. این نقاط ایستا جواب سراسری مسئله (۵.۲) هستند زیرا $A^T A \succeq 0$. پس

$$x = (A^T A)^{-1} A^T b$$

جواب مسئله‌ی (۵.۲) است و آن را با x_{LS} نمایش می‌دهیم.

۴.۲ مسئله کمترین مربعات خطی تنظیم شده (RLS)

در برخی مسائل درحین اینکه $\|Ax - b\|^2$ را مینیمم می‌کنیم می‌خواهیم تابع دیگری مانند $R(x)$ نیز مینیمم شود. با مسئله‌ای به صورت زیر روبرو هستیم.

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 + \lambda R(x) \quad (10.2)$$

که $R: \mathbb{R}^n \rightarrow \mathbb{R}$ تابعی داده شده و $\lambda > 0$ ثابتی داده شده است. اسکالر λ ، اهمیت دو تابع $\|Ax - b\|^2$ و $R(x)$ را نسبت به هم تنظیم می‌کند. مسئله‌ی (۱۰.۲) را یک RLS می‌نامیم. نشان می‌دهیم که اگر $\|R(x)\|$ دارای فرم درجه ۲ باشد، $R(x) = Dx$ ، مسئله به سادگی قابل حل است. داریم:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 + \lambda \|Dx\|^2 \quad (11.2)$$

$$\text{تابع هدف} = \|Ax - b\|^2 + \lambda \|Dx\|^2 \quad (12.2)$$

$$= x^T (A^T A + \lambda D^T D)x + 2(-A^T b)^T x + \|b\|^2$$

نقاط ایستای تابع هدف، جواب‌های دستگاه

$$(A^T A + \lambda D^T D)x = A^T b \quad (13.2)$$

هستند. این جواب‌ها مینیمم سراسری هستند، زیرا

$$A^T A + \lambda D^T D \succeq 0$$

به علاوه اگر A تمام رتبه ستونی باشند، $A^T A + \lambda D^T D \succ 0$ و در نتیجه

$$x_{RLS} = (A^T A + \lambda D^T D)^{-1} A^T b$$

جواب یکتای مسئله‌ی (10.2) خواهد بود.

5.2 برآزش دایره

• داده‌ها: $a_1 \dots a_m \in \mathbb{R}^n$

• هدف: یافتن مرکزیک گوی ($x \in \mathbb{R}^n$) و شعاع آن (r) است به طوری که تا حد امکان a_i ها روی

مرز $B(x; r)$ باشند. پس، مجهولات $x \in \mathbb{R}^n$ و $r \geq 0$ هستند به طوری که

$$\|x - a_i\| \approx r \quad i = 1, \dots, m$$

یا به طور معادل $\|x - a_i\|^2 \approx r^2$. با تکنیک کمترین مربعات به دنبال حل مسئله زیر هستیم.

$$x \in \mathbb{R}^n \quad \sum_{i=1}^m (\|x - a_i\|^2 - r^2)^2 \quad (14.2)$$

داریم

$$\begin{aligned} \|x - a_i\|^2 - r^2 &= (x - a_i)^T (x - a_i) - r^2 \\ &= \|x\|^2 - r^2 - 2a_i^T x + \|a_i\|^2 \end{aligned}$$

در ادامه می‌خواهیم مسئله (۱۴.۲) را به یک LS خطی تبدیل کنیم و جواب آن را (تحت فرضیاتی) بدست

آوریم. از تغییر متغیر $R = \|x\|^2 - r^2$ استفاده می‌کنیم. داریم:

$$\begin{aligned} R = \|x\|^2 - r^2 &\implies \|x\|^2 - R = r^2 \geq 0 \\ &\implies \|x\|^2 \geq R \end{aligned}$$

با این تغییر متغیر مسئله (۱۴.۲) به مسئله زیر تبدیل می‌شود.

$$\min_{\substack{x \in \mathbb{R}^n \\ R \in \mathbb{R}}} \sum_{i=1}^m (-2a_i^T x + R + \|a_i\|^2)^2 \quad (15.2)$$

$$s.t. \quad \|x\|^2 \geq R$$

در هر جواب بهینه مسئله

$$\min_{\substack{x \in \mathbb{R}^n \\ R \in \mathbb{R}}} \sum_{i=1}^m (-2a_i^T x + R + \|a_i\|^2)^2 \quad (16.2)$$

داریم $\|x\|^2 \geq R$. پس می‌توان قید $\|x\|^2 \geq R$ را از مسئله (۱۵.۲) حذف کرد. به عبارتی به مسئله

(۱۶.۲) می‌رسیم. با قرار دادن

$$\tilde{A} = \begin{bmatrix} -2a_1^T & 1 \\ \vdots & \vdots \\ -2a_m^T & 1 \end{bmatrix},$$

$$\tilde{b} = \begin{bmatrix} -\|a_1\|^2 \\ \vdots \\ -\|a_m\|^2 \end{bmatrix},$$

$$\tilde{y} = \begin{pmatrix} x \\ R \end{pmatrix} \in \mathbb{R}^{n+1}$$

مسئله‌ی (۱۶.۲) معادل است با

$$\min \|\tilde{A}\tilde{y} - \tilde{b}\|^2$$

$$\tilde{y} \in \mathbb{R}^{n+1}$$

اگر \tilde{A} تمام رتبه ستونی باشد جواب بهینه مسئله به صورت زیر است.

$$\tilde{y}^* = \begin{pmatrix} x^* \\ R^* \end{pmatrix} = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T \tilde{b}$$

و

$$\text{شعاع} : r = \sqrt{\|x^*\|^2 - R^*}$$

```
function h = circle(x,y,r)
```

```
hold on
```

```
th = 0:pi/50:2*pi;
```

```
xunit = r * cos(th) + x;
```

```
yunit = r * sin(th) + y;
```

```
h = plot(xunit, yunit);
```

```
hold off
```

```
function [X,R] = circle_fit(x,y)
```

```
x=x(:); y=y(:);
```

```
a=[x y ones(size(x))] \ [-(x.^2+y.^2)];
```

```
xc = -.5*a(1);
```

```
yc = -.5*a(2);
```

```
R = sqrt((a(1)^2+a(2)^2)/4-a(3));
```

```
X=[xc, yc];
```

فصل ۳

چند ابزار پایه‌ای

۱.۳ ترازهای برشی

مسئله‌ی زیر را به گونه‌ای در نظر بگیرید، 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.

$$\begin{aligned} \min f(x) \\ \text{s.t. } x \in X \end{aligned} \tag{۱.۳}$$

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

$$X^* \subseteq \{z \in \mathbb{R}^n : \nabla f(x)^T(x - z) \geq 0\}.$$

اهمیت نتیجه بالا، استخراج ابر صفحه‌ی زیر از آن می‌باشد

$$\{z \in \mathbb{R}^n : \langle \nabla f(x), x - z \rangle < 0\}.$$

توجه داریم که x متعلق به ابر صفحه‌ی زیر است

$$H = \{z \in \mathbb{R}^n : \langle \nabla f(x), x - z \rangle = 0\}. \quad (2.3)$$

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

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

۲.۳ تصویر گرادیان

تعریف ۱.۲.۳. [۲] فرض کنید $M > 0$.

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

$$T_M(x) \equiv P_X(x - \frac{1}{M} \nabla f(x)) \quad (3.3)$$

(ب) نگاشت گرادیان مقید بصورت زیر تعریف می‌گردد.

$$G_M(x) \equiv M(x - T_M(x)) = M[x - P_X(x - \frac{1}{M} \nabla f(x))] \quad (4.3)$$

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 $I - \frac{1}{M} \nabla f$

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

توجه داریم که

$$x \in X^* \iff G_M(x) = 0$$

برهان. فرض کنیم $x^* \in X^*$. آنگاه $x^* = P_X(x^* - \frac{1}{M}\nabla f(x^*))$. پس

$$(x^* - \frac{1}{M}\nabla f(x^*) - x^*)^T(x - x^*) \leq 0, \quad \forall x \in X$$

$$\iff -\frac{1}{M}\nabla f(x^*)^T(x - x^*) \leq 0, \quad \forall x \in X$$

$$\stackrel{M>0}{\iff} \nabla f(x^*)^T(x - x^*) \geq 0, \quad \forall x \in X$$

$$\iff x^* \in X^* \iff x^* \text{ ایستای مقید برای } P \text{ است}$$

□

نیم فضای متناظر با برش‌های عمیق بصورت زیر تعریف می‌گردد.

$$Q_{M,\alpha,x} \equiv \{z \in \mathbb{R}^n : \langle G_M(x), x - z \rangle \geq \frac{1}{\alpha M} \|G_M(x)\|^2\} \quad (5.3)$$

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 (5.3) بصورت زیر تعریف می‌گردد.

$$Q_{M,\alpha,x} \equiv \{z \in \mathbb{R}^n : \langle \nabla f(x), x - z \rangle \geq \frac{1}{\alpha M} \|\nabla f(x)\|^2\} \quad (6.3)$$

حال شرایط مسأله را به دو حالت تقسیم می‌کنیم.

۱. ثابت لیپ‌شیتز L معلوم باشد.

۲. ثابت لیپ‌شیتز L معلوم نباشد.

۱.۲.۳ ثابت لیپ‌شیتز L معلوم باشد

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

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2, \quad \forall x, y \in \mathbb{R}^n \quad (7.3)$$

با قراردادن $y = x^*$ بطوریکه $x^* \in X^*$ در رابطه‌ی (۷.۳)، با توجه به $\nabla f(x^*) = 0$ ، به رابطه‌ی زیر

می‌رسیم

$$\langle \nabla f(x), x - x^* \rangle \geq \frac{1}{L} \|\nabla f(x)\|^2, \quad \forall x \in \mathbb{R}^n, x^* \in X^* \quad (8.3)$$

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

زمانیکه مسأله مقید است و $X \neq \mathbb{R}^n$ ، رابطه‌ی (۸.۳) کمی پیچیده‌تر خواهد شد و داریم $X^* \subseteq Q_{L,\frac{1}{L},x}$.

این نتیجه بر پایه‌ی خواص نگاشت گرادیان G_L بدست می‌آید.

لم ۲.۲.۳. [۲] برای هر $x, y \in \mathbb{R}^n$ داریم

$$\langle G_L(x) - G_L(y), x - y \rangle \geq \frac{1}{4L} \|G_L(x) - G_L(y)\|^2. \quad (۹.۳)$$

برهان. بنابر (۹.۱)، داریم

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

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

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

بنابراین

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

در نتیجه

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

حال بنابر (۷.۳) داریم

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

$$\xrightarrow{\text{کوشی شوارتز}} L \langle G_L(x) - G_L(y), x - y \rangle \geq \|G_L(x) - G_L(y)\|^2$$

$$+ \|\nabla f(x) - \nabla f(y)\|^2 - \|\nabla G_L(x) - \nabla G_L(y)\| \cdot \|\nabla f(x) - \nabla f(y)\|$$

(۱۰.۳)

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

$$\alpha^2 + \beta^2 - \alpha\beta = \frac{3}{4}\alpha^2 + \left(\frac{\alpha}{2} - \beta\right)^2 \geq \frac{3}{4}\alpha^2$$

با توجه به رابطه‌ی (۱۰.۳)، داریم

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

با قرار دادن $y = x^*$ که $x^* \in X^*$ در رابطه‌ی (۹.۳)، نتیجه می‌گیریم

$$X^* \subseteq Q_{L, \frac{3}{4}, x}$$

□

لم زیر خلاصه بحث فوق است.

لم ۳.۲.۳. [۲] برای هر $x \in \mathbb{R}^n$ و $x^* \in X^*$ داریم

$$\langle G_L(x), (x - x^*) \rangle \geq \frac{3}{4L}\|G_L(x)\|^2 \quad (۱۱.۳)$$

که این یعنی

$$X^* \subseteq Q_{L, \frac{3}{4}, x}$$

اگر $X = \mathbb{R}^n$ باشد، آنگاه

$$\langle \nabla f(x), x - x^* \rangle \geq \frac{1}{L}\|\nabla f(x)\|^2 \quad (۱۲.۳)$$

که این یعنی

$$X^* \subseteq Q_{L, 1, x}$$

۲.۲.۳ ثابت لیپشیتز L معلوم نباشد

لم ۴.۲.۳. [۲] فرض کنید $x \in \mathbb{R}^n$ در رابطه‌ی زیر صدق کند

$$f(T_M(x)) \leq f(x) + \langle \nabla f(x), T_M(x) - x \rangle + \frac{M}{4}\|T_M(x) - x\|^2 \quad (۱۳.۳)$$

آنگاه برای هر $x^* \in X^*$ داریم

$$\langle G_M(x), x - x^* \rangle \geq \frac{1}{M} \|G_M(x)\|^2 \quad (۱۴.۳)$$

که این یعنی

$$X^* \subseteq Q_{M, \nu, x}$$

برهان. با در نظر گرفتن $x^* \in X^*$ و با توجه به رابطه (۱۳.۳)، داریم

$$\circ \leq f(T_M(x)) - f^* \leq f(x) - f^* + \langle \nabla f(x), T_M(x) - x \rangle + \frac{M}{\nu} \|T_M(x) - x\|^2. \quad (۱۵.۳)$$

با توجه به محدب بودن تابع f ، داریم

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

که با توجه به رابطه (۱۵.۳)، داریم

$$\circ \leq \langle \nabla f(x), T_M(x) - x^* \rangle + \frac{M}{\nu} \|T_M(x) - x\|^2 \quad (۱۶.۳)$$

حال با توجه به تعریف T_M و رابطه (۱۰.۳) داریم

$$\langle x - \frac{1}{M} \nabla f(x) - T_M(x), T_M(x) - x^* \rangle \geq \circ$$

با ضرب رابطه اخیر در M و جمع آن با رابطه (۱۶.۳)، داریم

$$M \langle x - T_M(x), T_M(x) - x^* \rangle + \frac{M}{\nu} \|T_M(x) - x\|^2 \geq \circ$$

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

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

□

لم ۵.۲.۳ (کاهشی). فرض کنید f تابعی بطور پیوسته مشتق پذیر باشد و بعلاوه $f \in C_L^1$. آنگاه

$$\forall x, y \in \mathbb{R}^n : f(y) \leq f(x) + \langle \nabla f(y), x - y \rangle + \frac{L}{\nu} \|x - y\|^2 \quad (۱۷.۳)$$

□

برهان. به [۲] رجوع کنید.

فصل ۴

الگوریتم حداقل نرم گرادیان

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.

$$\Omega(S) \equiv \arg \min_{x \in S} \omega(x) \quad (1.4)$$

با توجه به شرایط بهینگی در مسأله‌ی بالا، داریم

$$\tilde{x} = \Omega(S) \iff \langle \nabla \omega(\tilde{x}), x - \tilde{x} \rangle \geq 0, \quad \forall x \in S \quad (2.4)$$

هرگاه $\omega(x) = \frac{1}{2} \|x - a\|^2$ 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 لیپ‌شیتز)

• ورودی: L - ثابت لیپ‌شیتز برای ∇f

• مقدار آغازین: $x_0 = a$

• مرحله‌ی k -ام $k = 1, 2, \dots$

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

بطوریکه

$$Q^k = Q_{L, \beta, x_{k-1}}$$

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

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: $x_0 = a$

• مرحله k -ام $k = 1, 2, \dots$ 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

$$\begin{aligned} f(T_{\bar{L}}(x_{k-1})) &\leq f(x_{k-1}) + \langle \nabla f(x_{k-1}), T_{\bar{L}}(x_{k-1}) - x_{k-1} \rangle \\ &\quad + \frac{\bar{L}}{4} \|T_{\bar{L}}(x_{k-1}) - x_{k-1}\|^2 \end{aligned}$$

قرار دهید $L_k = \bar{L}$

قرار دهید:

$$x_k = \Omega(Q^k \cap W^k), \quad (3.4)$$

که در آن

$$Q^k = Q_{L,\beta,x_{k-1}}, \quad (4.4)$$

$$W^k = \{z \in \mathbb{R}^n : \langle \nabla \omega(x_{k-1}), z - x_{k-1} \rangle \geq 0\},$$

اکنون برای طول گام ثابت قرار می‌دهیم:

$$L_k = L, \quad \forall k \quad \eta = 1 \quad (5.4)$$

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

$$Q^k = Q_{L_k,\beta,x_{k-1}} \quad (6.4)$$

که داریم

$$\beta = \begin{cases} \frac{4}{3} & X \neq \mathbb{R}^n \text{ ثابت لیپشیتز مشخص} \\ 1 & X = \mathbb{R}^n \text{ ثابت لیپشیتز مشخص} \\ 2 & X = \mathbb{R}^n \text{ ثابت لیپشیتز نامشخص} \end{cases} \quad (7.4)$$

از طرفی می‌توانیم نشان دهیم، هنگامیکه داریم

$$L_0 \leq L_k \leq \eta L \quad k = 0, 1, 2, \dots \quad (8.4)$$

می‌توانیم نشان دهیم

$$\|G_{L_0}(x)\| \leq \|G_{L_k}(x)\| \leq \|G_{\eta L}(x)\| \quad (9.4)$$

لم ۱۰.۲.۴. [۲] فرض کنید دنباله‌ی $\{x_k\}_{k \geq 0}$ به وسیله‌ی روش حداقل نرم گرادیان توسط یک طول گام ثابت یا *backtracking* تولید شده باشد. آنگاه

$$X^* \subseteq Q^k \cap \omega^k \quad \forall k = 1, 2, \dots \quad (10.4)$$

۳.۴ تحلیل همگرایی

اولین هدف ما در این بخش آن است، که نشان دهیم روش حداقل نرم گرادیان، دنباله‌ای مانند $\{x_k\}_{k \geq 0}$ تولید کند بطوریکه به $x_{mn}^* = \Omega(X^*)$ همگرا باشد.

قضیه ۱.۳.۴ (همگرایی دنباله). [۲] فرض کنید دنباله‌ی $\{x_k\}_{k \geq 0}$ به وسیله‌ی روش حداقل نرم گرادیان با طول گام ثابت یا *backtracking* تولید شده باشد، آنگاه گزاره‌های زیر برقرار است.

۱. دنباله‌ی $\{x_k\}_{k \geq 0}$ کراندار است،

۲. نامساوی زیر برای هر $k = 1, 2, \dots$ برقرار است.

$$D_\omega(x_k, x_{k-1}) + D_\omega(x_{k-1}, a) \leq D_\omega(x_k, a) \quad (11.4)$$

۳. هرگاه $k \rightarrow \infty$ آنگاه $x_k \rightarrow x_{mn}^*$.

قضیه ۲.۳.۴ (مرتبه همگرایی). [۲] فرض کنید $\{x_k\}_{k \geq 0}$ دنباله‌ی تولید شده توسط روش حداقل نرم

گرادیان با یک طول گام ثابت یا *backtracking* باشد، آنگاه برای هر $k \geq 1$ داریم

$$\min_{1 \leq n \leq k} f(T_{L_n}(x_n)) - f^* \leq \frac{\beta \eta L \|a - x_{mn}^*\|^2}{\sqrt{k}} \quad (12.4)$$

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

اگر $X = \mathbb{R}^n$ آنگاه

$$\min_{1 \leq n \leq k} f(x_n) - f^* \leq \frac{\beta \eta L \|a - x_{mn}^*\|^2}{\sqrt{k}} \quad (13.4)$$

فصل ۵

مثال‌های کاربردی

در این فصل، کاربرد الگوریتم گفته شده در فصل قبل را با دو مثال کاربردی بررسی می‌کنیم.

۱.۵ مسأله‌ی انتخاب سهام

فرض کنید یک سرمایه‌گذار مایل به سرمایه‌گذاری در n موقعیت به شماره‌های $1, 2, \dots, n$ باشد. متغیر

تصادفی y_j ، $(j = 1, \dots, n)$ بیانگر بازده سرمایه در موقعیت j ام می‌باشد.

فرض کنیم، بازده مورد انتظار بصورت زیر باشد:

$$\mu_j = E(y_j), \quad j = 1, 2, \dots, n \quad (1.5)$$

همچنین کواریانس جفت متغیرها بصورت زیر باشد:

$$\sigma_{i,j} = Cov(y_i, y_j), \quad i, j = 1, 2, \dots, n \quad (2.5)$$

متغیرهای x_1, \dots, x_n متغیرهای تصمیم هستند، بطوریکه x_j نشان‌دهنده‌ی درصد سرمایه‌گذاری در موقعیت

j ام است. داریم

$$\sum_{j=1}^n x_j = 1, \quad x_1, \dots, x_n \geq 0 \quad (3.5)$$

بازده سبد سرمایه حاصل، یک متغیر تصادفی بصورت زیر تعریف می‌گردد

$$R = \sum_{j=1}^n x_j y_j \quad (4.5)$$

که دارای واریانس و امید ریاضی زیر است:

$$V(R) = x^T C x \quad (5.5)$$

$$E(R) = \mu^T x$$

که در آن $\mu = (\mu_1, \dots, \mu_n)^T$ و C ماتریس کواریانس می‌باشد، بطوریکه درایه‌های آن بصورت زیر هستند:

$$C_{i,j} = \sigma_{i,j}, \quad 1 \leq i, j \leq n$$

فرض می‌کنیم ماتریس کواریانس همواره یک ماتریس نیمه معین مثبت است. بعلاوه $x^T C x$ که همان واریانس انتخاب سهام می‌باشد، میزان ریسک را نمایش می‌دهد.

مدل‌سازی‌های متنوعی برای مسأله‌ی انتخاب سهام ایجاد شده است که تمام آن‌ها به مدل مارکوویتز که کاری از هری مارکوویتز^۱ است، باز می‌گردد. وی اولین بار در سال ۱۹۵۲ این مدل را ارائه کرد.

یکی از این مدل‌سازی‌ها، بدین صورت است که میزان ریسک را تحت شرایط حداقل سطح بازده، مینیمم می‌کند:

$$\begin{aligned} \min x^T C x \\ \text{s.t. } \mu^T x \geq \alpha \\ e^T x = 1 \\ x \geq 0 \end{aligned} \quad (6.5)$$

که در آن α حداقل بازده مورد انتظار است.

یکی دیگر از مدل‌سازی‌ها برای حل مسأله‌ی سهام بدین صورت است که، میزان بازده مورد انتظار را

^۱Harry Markowitz

تحت شرایطی روی میزان ریسک، ماکسیم کنیم:

$$\begin{aligned} \max \mu^T x \\ \text{s.t. } x^T C x \leq \beta \\ e^T x = 1 \\ x \geq 0 \end{aligned} \quad (7.5)$$

که در آن، β حداکثر میزان ریسک مورد تحمل است. در انتها مدل سازی سوم را بیان می کنیم، که تابع هدف آن ترکیبی از بازده مورد انتظار و ریسک است:

$$\begin{aligned} \min -\mu^T x + \lambda(x^T C x) \\ \text{s.t. } e^T x = 1 \\ x \geq 0 \end{aligned} \quad (8.5)$$

مسئله‌ی بهینه سازی انتخاب سهام^۲ زیر را در نظر بگیرید [۱۶].

فرض کنید فردی دارای N موقعیت سرمایه گذاری است که به شکل $1, 2, \dots, N$ شماره گذاری شده اند و می خواهد سرمایه های خود را در موقعیت های متفاوتی سرمایه گذاری کند بطوریکه انتظار او از سود حاصل از این سرمایه گذاری ها برابر $\mu \in \mathbb{R}^n$ می باشد. بعلاوه ماتریس کواریانس^۳ بصورت $\Sigma \in \mathbb{R}^{N \times N}$ است. همانطور که اشاره شد، در مسئله‌ی سهام بدنبال مینیم کردن ریسک هستیم، بطوریکه که بازده مورد انتظار از مقدار معین از پیش تعیین شده r_0 بیشتر یا مساوی باشد:

$$\begin{aligned} \min \omega^T \sum \omega \\ \text{s.t. } \sum_{i=1}^N \omega_i = 1 \\ \omega^T \mu \geq r_0 \\ \omega \geq 0 \end{aligned} \quad (9.5)$$

^۲Markowitz Portfolio

^۳Covariance Matrix

متغیر تصمیم را که برای بیان میزان سرمایه‌گذاری در هر یک از موقعیت‌ها تعریف می‌شود با w نمایش می‌دهیم. هنگامی که ماتریس کوواریانس نیمه معین مثبت باشد اما معین مثبت نباشد، جواب بهینه‌ی مسأله یکتا نیست و مشکل طبیعی ایجاد شده آن است که چگونه از بین جواب‌های بهینه، بهترین را انتخاب کنیم. این همان مسأله‌ی مینیمم نرم^۴ است.

توجه داریم که اغلب اوقات ماتریس کوواریانس نیمه معین مثبت است. زیرا این ماتریس را با توجه به داده‌های قبلی بدست می‌آورند و هنگامی که تعداد نمونه‌ها از تعداد موقعیت‌های سرمایه‌گذاری کمتر باشد، این مشکل ایجاد می‌شود.

بر اساس یک مثال خاص، فرض کنید در مسأله‌ی (۹.۵)، $N = 8$ باشد و ماتریس کوواریانس آن بصورت زیر باشد. این مسأله از میزان اوراق خزانه‌داری آمریکا در طول ۳ ماه استخراج شده است و بازده سالانه از سال ۱۹۷۳ تا ۱۹۹۴ در نظر گرفته شده است. داده‌ها را می‌توانید از سایت^۵، اشاره شده در پاورقی ببینید. اما در اینجا از داده‌های مربوط به سال‌های ۱۹۹۴ تا ۱۹۹۷، برای بدست آوردن Σ و μ استفاده کرده‌ایم.

$$\Sigma = \begin{pmatrix} 0/0002 & 0/0005 & 0/0028 & 0/0032 & 0/0039 & 0/0007 & 0/0024 & 0/0048 \\ 0/0005 & 0/0061 & 0/0132 & 0/0136 & 0/0126 & 0/0049 & 0/0003 & 0/0154 \\ 0/0028 & 0/0132 & 0/0837 & 0/0866 & 0/0810 & 0/0196 & 0/0544 & 0/1159 \\ 0/0032 & 0/0136 & 0/0866 & 0/0904 & 0/0868 & 0/0203 & 0/0587 & 0/1227 \\ 0/0039 & 0/0126 & 0/0810 & 0/0868 & 0/0904 & 0/0192 & 0/0620 & 0/1232 \\ 0/0007 & 0/0049 & 0/0196 & 0/0203 & 0/0192 & 0/0054 & 0/0090 & 0/0261 \\ 0/0024 & 0/0003 & 0/0544 & 0/0587 & 0/0620 & 0/0090 & 0/0619 & 0/0900 \\ 0/0048 & 0/0154 & 0/1159 & 0/1227 & 0/1232 & 0/0261 & 0/0900 & 0/1725 \end{pmatrix}$$

ماتریس کوواریانس نمونه با استفاده از فرمول زیر محاسبه شده است.

$$\Sigma := \frac{1}{T-1} R(I_T - \frac{1}{T} \mathbf{1}\mathbf{1}^T) R^T \quad (10.5)$$

^۴Minimum Norm Problem (MNP)

^۵ <http://www.princeton.edu/~rvdb/ampl/nlmodels/markowitz/>

در اینجا $T = 4$ (تعداد دوره‌ها) و R یک ماتریس 4×8 می‌باشد، که شامل سرمایه‌های بازگشتی در 4 سال است.

رتبه‌ی ماتریس Σ حداکثر می‌تواند مقدار 4 را داشته باشد، بنابراین مسأله دارای جواب بهینه‌ی یکتا نیست پس با یک مسأله‌ی مینیمم نرم مواجه هستیم.

حداقل بازده انتظاری را میزان $r_0 = 1/05$ در نظر می‌گیریم. در این حالت مسأله‌ی (9.5) دارای جواب بهینه‌ی دگرین است. برای حل مسأله‌ی مینیمم نرم تابع هدف مسأله‌ی بیرونی را بصورت زیر در نظر می‌گیریم.

$$\omega(x) = \frac{1}{4} \|x - z\|^2 \quad (11.5)$$

فرض کنید تا جایی که ممکن است، در موقعیت سرمایه‌گذاری هشتم سرمایه‌گذاری کنیم.

بنابراین $a = (0, 0, 0, 0, 0, 0, 0, 1)^T$ و جواب مسأله بصورت زیر است:

$$(0/0000, 0/0000, 0/0995, 0/1421, 0/2323, 0/0000, 0/1261, 0/3999)^T$$

همانطور که می‌بینیم، میزان سرمایه‌گذاری پراکنده شده، پس می‌توانیم a را بصورت زیر انتخاب کنیم.

$$a = \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8} \right)$$

جواب حاصل از الگوریتم بصورت زیر خواهد بود، که با جواب بهینه‌ی قبلی کاملاً متفاوت می‌باشد.

$$(0/1531, 0/1214, 0/0457, 0/0545, 0/1004, 0/1227, 0/1558, 0/2466)^T$$

توجه داریم که مبلغ سرمایه‌گذاری شده در موقعیت سرمایه‌گذاری هشتم، در این جواب کمتر است، ولی میزان پراکندگی سرمایه‌گذاری کاهش یافته است. پس می‌توانیم با انتخاب a های متفاوت نسبت به اولویت‌های هر فرد، بهترین میزان سرمایه‌گذاری را در هر موقعیت بدست آوریم.

2.5 جواب‌های معادلات انتگرالی

جواب مسأله‌ی حداقل نرم می‌تواند نوعی از جواب تنظیم شده برای مسأله‌ی درونی باشد. در ادامه این میزان تنظیم شده را با حل مسأله‌ی معادلات انتگرالی در حالت گفته شده، یعنی مسائل انتگرالی نوع اول به وسیله‌ی روش حداقل نرم گرادیان نشان می‌دهیم.

مسأله‌ی مورد بحث از [۱۳] اقتباس شده است. در ابتدا مسأله‌ی مشهور فیلیپس^۶ [۱۹] را می‌بینیم، که با حدس تابع $f(t)$ سعی در حل معادله‌ی انتگرالی زیر دارد.

$$\int_{-6}^6 k(s-t)f(t) = g(s)$$

بطوریکه

$$k(t) = \begin{cases} 1 + \cos\left(\frac{\pi t}{3}\right), & |t| < 6 \\ 0, & \text{otherwise} \end{cases}$$

همچنین

$$g(s) = (6 - |s|) \left(1 + \frac{1}{3} \cos\left(\frac{\pi s}{3}\right) \right) + \frac{9}{2\pi} \sin\left(\frac{\pi |s|}{3}\right) \quad (12.5)$$

که در آن $\sigma = 0.52$. گالرکین^۷ مسأله را به حالت خطی کاهش داده و به فرم $Ax_T = b_T$ تبدیل کرده است. این سیستم و جواب‌هایش که در تابع فیلیپس پیاده‌سازی شده‌اند در سایت^۸ اشاره شده در پاورقی، قابل مشاهده می‌باشند. تابع در برنامه متلب با $[A, b, x] = \text{phillips}(1, 000)$ اجرا شده است. در این مثال $n = 1000$ در نظر گرفته شده است، بنابراین تعداد متغیرها برابر ۱۰۰۰ می‌باشد. با استفاده از روش، بردارهای مقادیر سمت راست بصورت زیر نمایش داده می‌شود.

$$b = b_T + \sigma.w \quad (13.5)$$

w ها از یک توزیع نرمال و w میان صفر ساخته شده‌اند.

در اینجا مسأله‌ی بیرونی را مسأله‌ی کمترین مربعات در نظر می‌گیریم.

$$\min_x \|Ax - b\|^2 \quad (14.5)$$

ماتریس A دارای مقدار ویژه‌ی صفر است. بنابراین، مسأله‌ی بیرونی دارای جواب بهینه‌ی دگرین می‌باشد. تابع هدف مسأله‌ی درونی را به شکل زیر در نظر می‌گیریم.

$$\omega(x) = x^T Q x \quad (15.5)$$

^۶Phillips

^۷Galerkin

^۸<http://www2.imm.dtu.dk/pcha/Regutools/>

که در آن، $Q = L^T L + I$ و L تقریباً مشتق اول اپراتور قرار داده شده در $get_l(1, 000, 1)$ از قوانین تنظیم شده است [۲].

برای حل مسأله با الگوریتم حداقل نرم گرادیان، نقطه‌ی آغازین را $x = 0$ در نظر می‌گیریم. متناظر با مسأله‌ی تیخونوف، جواب‌های متفاوتی در نظر گرفته می‌شود.

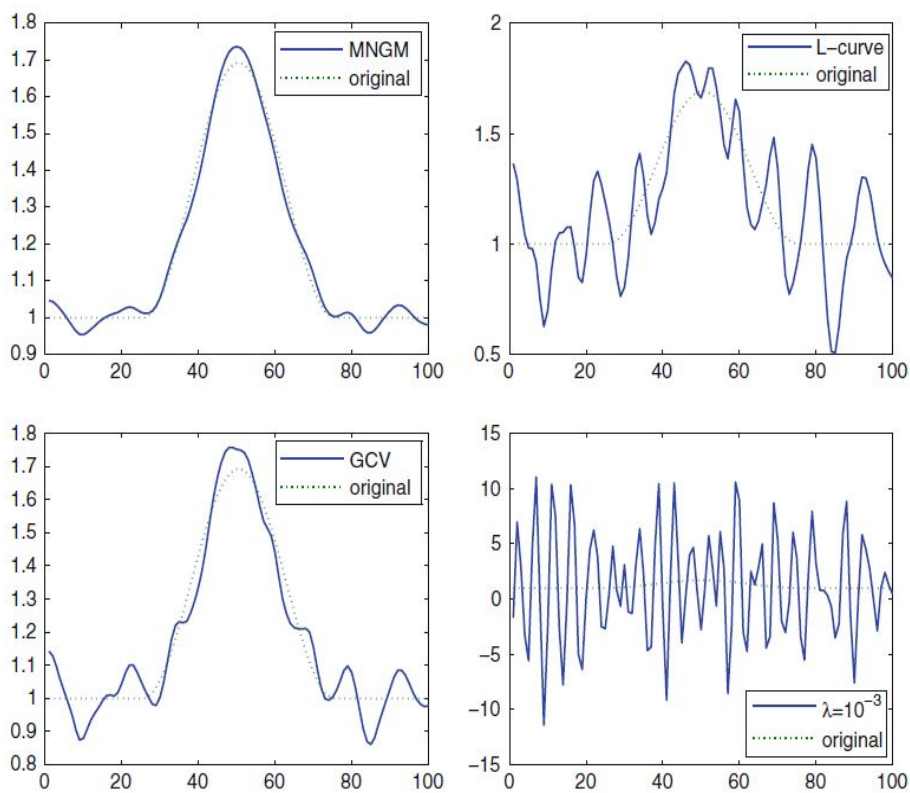
$$x_\lambda = \arg \min_x \{ \|Ax - b\|^2 + \lambda x^T Q x \}$$

که در آن λ با استفاده از استراتژی $L - Cure$ [۱۲] یا GCV ^۹ [۱۱] انتخاب شده است.

بعلاوه، معمولاً فرض می‌کنیم $\lambda = 10^{-3}$ مقدار کوچک λ را نشان می‌دهد. نتایج در شکل ۱.۵ نشان داده شده است.

واضح است که حل مسأله با استفاده از روش حداقل نرم گرادیان نتیجه‌ی بهتری را نسبت به جواب‌های بهینه‌ی دیگر در اختیار ما قرار می‌دهد [۲].

^۹Generalized Cross Validation



شکل ۱.۵: نتیجه‌ی روش‌های انجام شده برای مسأله‌ی فیلیپس. عکس بالا سمت چپ، نتیجه‌ی روش شبه نرم‌گرادیان می‌باشد. عکس بالا سمت راست، نتیجه‌ی روش $L - Curve$ با پارامترهای تیخونوف است. عکس پایین سمت راست و چپ، با استفاده از روش GCV با پارامترهای تیخونوف می‌باشد [۲].

کتاب نامه

- [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

Applications of strongly convex optimization

By:

Elmira Fathi Pasandideh

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