به نام خدا

بهینه‌سازی محدّب ۱ (۱-۲۵۷۵۶)

تمرین سری ششم

ترم بهار ۰۲-۱۴۰۱

دانشکده‌ی مهندسی برق

دانشگاه صنعتی شریف

استاد: دکتر محمّدحسین یاسائی میبدی

مهلت تحویل: یک‌شنبه ۲۱ خرداد ۱۴۰۲، ساعت ۱۱:۵۹

$(*)$ مسائلی که با ستاره مشخّص شده‌اند امتیازی هستند و حل‌کردن آن‌ها نمره‌ی امتیازی خواهد داشت!

۱ کاهش گرادیان برای توابع مربعی!

در این مسئله به تحلیل الگوریتم کاهش گرادیان برای تابع ساده مربعی زیر می پردازیم:

$$f(x) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{b}^\top \mathbf{x} + c$$

فرض کنید که در قدم t ام از اندازه η_t استفاده می‌کنیم، یعنی:

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \eta_t \nabla f(\mathbf{x}_t).$$

۱. نشان دهید:

$$\mathbf{x}_{t+1} = (\mathbf{I} - \eta_t\mathbf{A})\mathbf{x}_t + \eta_t\mathbf{b}.$$

۲. فرض کنید $\mathbf{x}^*$ نقطه‌ی بهینه‌کننده تابع f باشد. علت درستی روابط زیر را بیان کنید:

$$\begin{aligned}\mathbf{x}_{t+1} - \mathbf{x}^* &= (\mathbf{x}_t - \eta_t(\mathbf{A}\mathbf{x}_t - \mathbf{b})) - (\mathbf{x}^* - \eta_t(\mathbf{A}\mathbf{x}^* - \mathbf{b})) \\ &= (\mathbf{I} - \eta_t\mathbf{A})(\mathbf{x}_t - \mathbf{x}^*) = \prod_{k=1}^{t}(\mathbf{I} - \eta_k\mathbf{A})(\mathbf{x}_1 - \mathbf{x}^*)\end{aligned}$$

۳. حال علت درستی روابط زیر را بیان کنید:

$$\begin{aligned}\|\mathbf{x}_{t+1} - \mathbf{x}^*\| &\le \|\prod_{k=1}^{t}(\mathbf{I} - \eta_k\mathbf{A})\|_2\|\mathbf{x}_1 - \mathbf{x}^*\|_2 \\ &\le (\prod_{k=1}^{t}\|\mathbf{I} - \eta_k\mathbf{A}\|_2\|\mathbf{x}_1 - \mathbf{x}^*\|_2.\end{aligned}$$

دقت کنید که نرم در نظر گرفته شده نرم عملگری ماتریس است که برابر با بزرگ‌ترین مقدار ویژه‌ی ماتریس است.

۴. حال فرض کنید که تابع f ، M–هموار و m–قویاً محدّب باشد. نشان دهید که دو شرط فوق معادل است با:

$$\frac{m}{M}\mathbf{I} \preceq \frac{1}{M}\mathbf{A} \preceq \mathbf{I}.$$

و در نتیجه:

$$\|\mathbf{I} - \frac{1}{\beta}\mathbf{A}\|_2 = 1 - \frac{\alpha}{\beta} = 1 - \frac{1}{\kappa}$$

که در آن α کوچک‌ترین مقدار ویژه $\mathbf{A}$ و β بزرگ‌ترین مقدار وِیژه‌ی آن می‌باشد.

۵. با استفاده از قسمت‌های قبل نشان دهید که برای قدم ثابت $\eta_t = \frac{1}{M}$ داریم:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2 \leq \exp\left(-\frac{t}{\kappa}\right)\|\mathbf{x}_1 - \mathbf{x}^*\|_2.$$

که در آن κ عدد حالت ماتریس A است.

۲ نیوتون!

فرض کنید $f : \mathbb{R}^n \to \mathbb{R}$ یک تابع محدب و دوبار مشتق‌پذیر با مشتقات پیوسته است. فرض کنید ∇f تابع L–لیدشیتز است و f یک تابع قویاً محدب با پارامتر $m > 0$ است و فرض کنید $\nabla^2 f$ یک تابع M ـلیپشیتز است یعنی:

$$\|\nabla^2 f(\mathbf{x}) - \nabla^2 f(\mathbf{y})\|_{\text{op}} \leq M\|\mathbf{x} - \mathbf{y}\|_2$$

فرض کنید $\mathbf{x}^*$ نقطه‌ی کمینه‌کننده‌ی سراسری f باشد.

۱. نشان دهید که برای هر $\mathbf{x}$ داریم :

$$f(\mathbf{x}) - f\left(\mathbf{x}^*\right) \leq \frac{1}{2m}\|\nabla f(\mathbf{x})\|_2^2.$$

۲. نشان دهید برای تکرار مرحله‌ی t–ام الگوریتم نیوتن داریم:

$$\frac{M}{2m^2}\left\|\nabla f\left(\mathbf{x}_{t+1}\right)\right\|_2 \leq \left(\frac{M}{2m^2}\left\|\nabla f\left(\mathbf{x}_t\right)\right\|_2\right)^2.$$

۳. فرض کنید که بعد از k_0 مرحله، شرط زیر را داریم:

$$\left\|\nabla f\left(\mathbf{x}_{k_0}\right)\right\|_2 < \eta \leq \frac{m^2}{M}$$

نشان دهید برای $k > k_0$:

$$f\left(\mathbf{x}_k\right) - f\left(\mathbf{x}^*\right) \leq \frac{2m^3}{M^2}\left(\frac{1}{2}\right)^{2^{k-k_0+1}}$$

۳ چند پرسش!

به پرسش‌های زیر پاسخ دهید:

۱. فرض کنید f یک تابع محدب است و روی یک مجموعه‌ی محدب و بسته تعریف شده است. ثابت کنید f یک تابع قویا محدب با پارامتر m است اگر و تنها اگر ∇f^* یک تابع $\frac{1}{m}$–لیپشیتز باشد. (f^* تابع مزدوج f است.)

۲. مساله‌ی بهینه‌سازی زیر را در نظر بگیرید:

$$\begin{aligned} \text{Minimize}_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

الگوریتم زیر با انتخاب یک نقطه‌ی اولیه‌ی $\mathbf{u}^{(0)}$ برای حل این مساله داده شده است.

$$\begin{aligned} &\mathbf{x}^{(k)} \in \arg\min_{\mathbf{z}} f(\mathbf{z}) + \left(u^{(k-1)}\right)^{\top}\mathbf{A}\mathbf{z} \\ &\mathbf{u}^{(k)} = \mathbf{u}^{(k-1)} + t_k\left(\mathbf{A}\mathbf{x}^{(k)} - \mathbf{b}\right) \end{aligned}$$

که در آن t_k اندازه‌ی گام (نرخ یادگیری) برای به‌روزرسانی $\mathbf{u}^{(k)}$ است. آیا می‌توانید بگویید این الگوریتم از چه مساله‌ای نشات گرفته است؟
راهنمایی: مساله‌ی دوگان را در نظر گرفته و در صدد برقراری شرط صفر بودن گردیان تابع هدف برآیید.

۳. الگوریتم ارائه شده در قسمت قبل را یک بار دیگر برای تابع f با شرط قویاً محدب با پارامتر m و همچنین یک‌بار برای تابع f با شرط قویا محدب با پارامتر m و گرادیان L لیپشیتز آنالیز همگرایی کنید و تعداد مراحل همگرایی با خطای ϵ را به دست آورید. **راهنمایی:** می‌توانید از آنالیز همگرایی کاهش گرادیان استفاده کنید.

۴ روش گرادیان و شروط تساوی!

در این سوال به دنبال یافتن تعمیم روش گرادیان برای مسائل دارای شروط تساوی خواهیم بود. فرض کنید که f یک تابع محدب و مشتق پذیر باشد، و $\mathbf{x} \in \mathrm{dom}\, f$ در شرط $\mathbf{Ax} = \mathbf{b}$ صدق کند، که در آن $\mathbf{A} \in \mathbb{R}^{p\times n}$ با $\mathrm{rank}(\mathbf{A}) = p < n$ است.
تصویر اقلیدسی منفی گرادیان $-\nabla f(\mathbf{x})$ بر روی $\mathrm{null}(\mathbf{A})$ توسط رابطه‌ی زیر داده می‌شود:

$$\Delta \mathbf{x}_{\mathrm{pg}} = \arg\min_{\mathbf{Au}=۰} \| -\nabla f(\mathbf{x}) - \mathbf{u}\|_۲$$

۱. فرض کنید که $(\mathbf{v}, \mathbf{w})$ جواب یکتای مسئله‌ی زیر باشند.

$$\begin{bmatrix} \mathbf{I} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ \mathbf{0} \end{bmatrix}$$

نشان دهید که :

$$\mathbf{v} = \Delta \mathbf{x}_{\mathrm{pg}}$$

و

$$\mathbf{w} = \arg\min_{\mathbf{y}} \|\nabla f(\mathbf{x}) + \mathbf{A}^\top \mathbf{y}\|_۲.$$

۲. رابطه‌ی بین منفی گرادیان تصویر شده یعنی $\Delta \mathbf{x}_{\mathrm{pg}}$ و منفی گرادیان مسئله‌ی کاهش یافته که به فرمت زیر است:

$$\text{Minimize} \quad \tilde{f}(\mathbf{z}) = f(\mathbf{Fz} + \hat{\mathbf{x}})$$

و با فرض آنکه داشته باشیم $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$، $\mathbf{Fz} \in \mathrm{null}(\mathbf{A})$ و $\mathbf{F}^\top\mathbf{F} = \mathbf{I}$ چیست؟

۳. روش گرادیان تصویر شده برای حل یک مسئله‌ی بهینه‌سازی با شروط تساوی از طول گام $\Delta \mathbf{x}_{\mathrm{pg}}$ و روش جستجوی خط backtracking بر روی f استفاده می‌کند. با استفاده از بخش قبل، شروطی را مطرح کنید که تحت آن روش گرادیان تصویر شده به جواب بهینه همگرا شود. فرض کنید که نقطه‌ی شروع الگوریتم نقطه‌ی $x_۰ \in \mathrm{dom}\, f$ است که در شرط $\mathbf{Ax}_۰ = \mathbf{b}$ صدق می‌کند.

۵ $(*)$ طرّاحی آزمایش!

مسائل زیر در طراحی آزمایشی مطرح می‌شوند:
الف) مسئله‌ی طراحی D-optimal:

$$\begin{aligned} &\text{Minimize} \quad \log\det\left\{\left(\sum_{i=۱}^{p} x_i \mathbf{v}_i \mathbf{v}_i^\top\right)^{-۱}\right\} \\ &\text{subject to:} \quad \mathbf{x} \succeq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x} = ۱ \end{aligned}$$

ب) مسئله‌ی طراحی A-optimal:

$$\begin{aligned} &\text{Minimize} \quad \mathrm{tr}\left\{\left(\sum_{i=۱}^{p} x_i \mathbf{v}_i \mathbf{v}_i^\top\right)^{-۱}\right\} \\ &\text{subject to:} \quad \mathbf{x} \succeq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x} = ۱ \end{aligned}$$

دامنه‌ی هر دوی این مسائل $\{\mathbf{x} \mid \sum_{i=1}^{p} x_i \mathbf{v}_i \mathbf{v}_i^T \succ \mathbf{0}\}$ است، که در آن متغیر $\mathbf{x} \in \mathbb{R}^p$ است و بردارهای $\mathbf{v}_1, \dots, \mathbf{v}_p \in \mathbb{R}^n$ داده شده‌اند.

دوگان مسائل فوق را با تعریف یک متغیر جدید $\mathbf{X} \in \mathbf{S}^n$ و هم چنین اضافه کردن شرط تساوی $\mathbf{X} = \sum_{i=1}^{p} x_i \mathbf{v}_i \mathbf{v}_i^\top$ سپس به کار بردن روش لاگرانژین بیابید. مسائل دوگان را تا جایی که می‌شود ساده کنید.