به نام خداوند بخشنده مهربان

جبر خطی

دانشکده مهندسی کامپیوتر

حمیدرضا ربیعی، مریم رمضانی
پاییز ۱۴۰۳

دستگاه مختصات، فضای برداری و زیرفضاها — **تمرین تئوری ششم** — تاریخ انتشار: ۱۳ دی ۱۴۰۳

---

۱. پرسش‌های خود درمورد این تمرین را در سامانه کوئرا مطرح کنید.

۲. **سیاست ارسال با تاخیر:** شما در مجموع در طول نیم‌سال می‌توانید از ۴ روز تاخیر استفاده کنید. این مقدار برای تمارین تئوری و عملی به‌صورت جداگانه حساب می‌شود. تاخیرها با مقیاس ساعت محاسبه شده و به بالا گرد می‌شوند.

۳. **سیاست مشارکت دانشجویان در حل کردن تمارین:** دانشجویان می‌توانند در حل تمارین برای رفع ابهام و یا به‌دست آوردن ایده‌ی کلی با یک‌دیگر مشورت و همفکری کنند. این کار مورد تایید و تشویق تیم ارائه‌ی درس می‌باشد؛ چرا که هم‌فکری و کار گروهی می‌تواند موجب تقویت یادگیری شود. اما به‌دست آوردن جزئیات راه‌حل و نگارش پاسخ باید تماما توسط خود دانشجو انجام شود. **حتما در انتهای پاسخ‌های ارسالی خود نام افرادی که با آن‌ها همفکری کردید را ذکر کنید.**

پرسش ۱ (۱۵ نمره) در این مسئله تمامی متغیرهایی که با حرف بزرگ لاتین مشخص شده‌اند، نشان‌دهنده ماتریس هستند. همچنین منظور از $\|\cdot\|_F$ نرم فروبنیوس می‌باشد.

(آ) مشتق تابع زیر را به‌دست بیاورید:

$$f(X) = \|AX - B\|_F^2$$
$$\frac{\partial f(X)}{\partial X} = ?$$

(ب) درستی عبارات زیر را نشان دهید:

i.

$$\|AB\|_F \leq \|A\|_F \|B\|_F$$

ii.

$$\|A\|_F \leq \sqrt{n}\|A\|_2$$

دقت کنید که داریم:

$$\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{\max}(A^T A)} = \sigma_{\max}(A)$$

پاسخ

(آ) مشتق تابع $f(X) = \|AX - B\|_F^2$ را محاسبه می‌کنیم:

ابتدا تابع را بازنویسی می‌کنیم:

$$f(X) = \mathrm{tr}((AX - B)^T (AX - B)).$$
$$f(X) = \mathrm{tr}(X^T A^T AX - 2B^T AX + B^T B).$$

اکنون مشتق $\frac{\partial f(X)}{\partial X}$ را محاسبه می‌کنیم:

$$\frac{\partial}{\partial X}\mathrm{tr}(X^T A^T AX) = 2A^T AX,$$
$$\frac{\partial}{\partial X}\mathrm{tr}(-2B^T AX) = -2A^T B,$$
$$\frac{\partial}{\partial X}\mathrm{tr}(B^T B) = 0.$$

بنابراین:

$$\frac{\partial f(X)}{\partial X} = 2A^T AX - 2A^T B.$$

(ب) i.

$$\|AB\|_F^2 \leq \|A\|_F^2 \|B\|_F^2$$
$$\|M\|_F^2 = \mathrm{tr}(M^T M).$$

برای حاصل‌ضرب AB داریم:

$$\|AB\|_F^2 = \mathrm{tr}((AB)^T (AB)) = \mathrm{tr}(B^T A^T AB).$$

$$\mathrm{tr}(XY) \leq \sqrt{\mathrm{tr}(X^2)} \cdot \sqrt{\mathrm{tr}(Y^2)}.$$

با قرار دادن $X = A^T A$ و $Y = BB^T$:

$$\|AB\|_F^2 \leq \sqrt{\mathrm{tr}((A^T A)^2)} \cdot \sqrt{\mathrm{tr}((BB^T)^2)}.$$

از تعریف نرم فروبینیوس:

$$\|A\|_F^2 = \mathrm{tr}(A^T A), \quad \|B\|_F^2 = \mathrm{tr}(B^T B).$$

بنابراین داریم:

$$\sqrt{\mathrm{tr}((A^T A)^2)} \leq \|A\|_F^2, \quad \sqrt{\mathrm{tr}((BB^T)^2)} \leq \|B\|_F^2.$$

در نتیجه:

$$\|AB\|_F^2 \leq \|A\|_F^2 \|B\|_F^2.$$

بنابراین:

$$\|AB\|_F \leq \|A\|_F \|B\|_F.$$

ii.

$$\|A\|_F^۲ = \mathrm{tr}(A^T A).$$

از اینجا داریم:

$$\|A\|_F^۲ = \sum_{i=۱}^{n} \sigma_i^۲ \le n \max_i \sigma_i^۲ = n\|A\|_۲^۲,$$

که در نتیجه:

$$\|A\|_F \le \sqrt{n}\|A\|_۲.$$

پرسش ۲ (۱۷ نمره) یکی از کاربرد های تجزیه مقدار تکین ، محاسبه تقریب رتبه پایین یک ماتریس است. ماتریس $A \in \mathbb{R}^{m\times n}$ با رتبه r را با تجزیه مقادیر تکین $A = U\Sigma V^T$ در نظر بگیرید. به ازای یک عدد $k \le r$ ، ماتریس A_k را به صورت $A_k = U_k \Sigma_k V_k^T$ تعریف می کنیم که در آن U_k و V_k به ترتیب k ستون اول U و V بوده و Σ_k بلوک $k \times k$ بالا و سمت چپ ماتریس Σ است که شامل k مقدار تکین بزرگ A می شود. به عبارتی می توان A_k را به صورت $A_k = \sum_{i=۱}^{k} \sigma_i u_i v_i^T$ بیان کرد که ماتریسی رتبه k است.

الف) نشان دهید A_k پاسخ مسئله بهینه سازی زیر است.

$$\min_{X\ ,\ rank(X)\le k} ||A - X||_F$$

ب) نشان دهید A_k پاسخ مسئله بهینه سازی زیر نیز می باشد.

$$\min_{X\ ,\ rank(X)\le k} ||A - X||_۲$$

راهنمایی : در هر دو مسئله باید نشان دهید

$$\forall X, rank(X) \le k : ||A - A_k|| \le ||A - X||$$

پاسخ

الف) قضیه:

$$\forall i, j : \sigma_i(X) + \sigma_j(Y) \ge \sigma_{i+j-۱}(X + Y)$$

ابتدا قضیه ذکر شده را اثبات می کنیم.

$$\forall X, Y : ||X||_۲ + ||Y||_۲ \ge ||X + Y||_۲ \Longrightarrow \sigma_۱(X) + \sigma_۱(Y) \ge \sigma_۱(X + Y)$$

$$\sigma_i(X) + \sigma_j(Y) = \sigma_۱(X - X_{i-۱}) + \sigma_۱(Y - Y_{j-۱}) \ge \sigma_۱(X + Y - X_{i-۱} - Y_{j-۱})$$

$$rank(X_{i-۱} + Y_{j-۱}) \le rank(X_{i-۱}) + rank(Y_{j-۱}) = i + j - ۲$$

$$\Longrightarrow \sigma_۱(X + Y - X_{i-۱} - Y_{j-۱}) \ge \sigma_۱(X + Y - (X + Y)_{i+j-۲}) = \sigma_{i+j-۱}(X + Y)$$

$$\Longrightarrow \sigma_i(X) + \sigma_j(Y) \ge \sigma_{i+j-۱}(X + Y)$$

حال با قرار دادن $X = A_k$ و $Y = Z$ که در آن $rank(Z) \le k$ است و $j = k + ۱$ می توان نتیجه گرفت که

$$\forall i : \sigma_i(A - Z) + \underbrace{\sigma_{k+۱}(Z)}_{۰} \ge \sigma_{i+k}(A)$$

بنابراین خواهیم داشت :

$$||A - Z||_F^۲ = \sum_{i=۱}^{\min(m,n)} \sigma_i^۲(A - Z) \ge \sum_{i=k+۱}^{\min(m,n)} \sigma_i^۲(A)$$

$$||A - A_k||_F^۲ = ||\sum_{i=k+۱}^{\min(m,n)} \sigma_i u_i v_i^T||_F^۲ = tr\left\{(\sum_{i=k+۱}^{\min(m,n)} \sigma_i u_i v_i^T)^T(\sum_{i=k+۱}^{\min(m,n)} \sigma_i u_i v_i^T)\right\}$$

$$= tr\left\{ \sum_{i=k+1}^{\min(m,n)} \sigma_i^2 v_i \underbrace{u_i^T u_i}_{1} v_i^T \right\} == tr\left\{ \sum_{i=k+1}^{\min(m,n)} \sigma_i^2 v_i v_i^T \right\} = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2 tr\{v_i v_i^T\} = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2 tr\{v_i^T v_i\} = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2$$

$$\Longrightarrow \forall Z, rank(Z) \leq k : ||A - Z||_F^2 \geq ||A - A_k||_F^2$$

ب)

$$rank(X) = k \Longrightarrow \dim \mathcal{N}(X) = n - k \ \ , \ \ \dim \mathcal{R}(V_{k+1}) = k + 1$$

$$\dim \mathcal{N}(X) + \dim \mathcal{R}(V_{k+1}) = n + 1 \Longrightarrow \exists x \in \mathcal{N}(X) \cap \mathcal{R}(V_{k+1}) : ||x||_2 = 1$$

$$x = \sum_{i=1}^{k+1} \alpha_i v_i \ \ , \ \ \sum_{i=1}^{k+1} \alpha_i^2 = 1$$

$$||A - X||_2 = \sup_{||x||_2 = 1} ||(A - X)x||_2 \Longrightarrow ||A - X||_2 \geq ||(A - X)x||_2 = ||Ax||_2$$

$$A = U\Sigma V^T \Longrightarrow Ax = U\Sigma V^T x = \sum_{i=1}^{k+1} \sigma_i \alpha_i u_i \Longrightarrow ||Ax||_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 \alpha_i^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \alpha_i^2 = \sigma_{k+1}^2$$

$$||A - X||_2 \geq \sigma_{k+1}$$

$$||A - A_k||_2 = || \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T - \sum_{i=1}^{k} \sigma_i u_i v_i^T ||_2 = || \sum_{i=k+1}^{\min(m,n)} \sigma_i u_i v_i^T ||_2 = \sigma_{k+1}$$

$$\Longrightarrow \forall X, rank(X) \leq k : ||A - X||_2 \geq ||A - A_k||_2$$

پرسش ۳ (۱۶ نمره) فرض کنید A یک ماتریس مربعی $n \times n$ و x برداری از $\mathbb{R}^n$ باشد. مقادیر زیر را بیابید:

(آ)

$$\frac{\partial(x^T Ax)}{\partial x}$$

(ب) اگر درایه‌های ماتریس A تابعی از یک اسکالر β باشند، مقدار زیر را محاسبه کنید:

$$\frac{\partial A^{-1}}{\partial \beta}.$$

پاسخ

فرض کنید A یک ماتریس مربعی $n \times n$ و x برداری از $\mathbb{R}^n$ باشد. مقادیر زیر را بیابید:

(آ)

$$\frac{\partial(x^T Ax)}{\partial x}$$

مقدار عبارت فوق برابر است با:

$$\frac{\partial(x^T Ax)}{\partial x} = (A + A^T)x.$$

اگر A متقارن باشد ($A = A^T$)، نتیجه به صورت زیر ساده می‌شود:

$$\frac{\partial(x^T Ax)}{\partial x} = 2Ax.$$

(ب) اگر درایه‌های ماتریس A تابعی از یک اسکالر β باشند، مقدار زیر را محاسبه کنید:

$$\frac{\partial A^{-۱}}{\partial \beta}.$$

ابتدا از رابطه اساسی ماتریسی شروع می‌کنیم:

$$AA^{-۱} = I,$$

که I ماتریس همانی است. دو طرف معادله را نسبت به β مشتق می‌گیریم:

$$\frac{\partial}{\partial \beta}\left(AA^{-۱}\right) = \frac{\partial I}{\partial \beta}.$$

از آنجا که I نسبت به β ثابت است، مشتق آن صفر می‌شود:

$$\frac{\partial}{\partial \beta}\left(AA^{-۱}\right) = ۰.$$

حال از قاعده زنجیره‌ای برای مشتق‌گیری ضرب ماتریسی استفاده می‌کنیم:

$$\frac{\partial A}{\partial \beta}A^{-۱} + A\frac{\partial A^{-۱}}{\partial \beta} = ۰.$$

برای پیدا کردن $\frac{\partial A^{-۱}}{\partial \beta}$، عبارت $\frac{\partial A}{\partial \beta}A^{-۱}$ را به طرف دیگر معادله منتقل می‌کنیم:

$$A\frac{\partial A^{-۱}}{\partial \beta} = -\frac{\partial A}{\partial \beta}A^{-۱}.$$

سپس دو طرف معادله را در $A^{-۱}$ از سمت چپ ضرب می‌کنیم:

$$\frac{\partial A^{-۱}}{\partial \beta} = -A^{-۱}\frac{\partial A}{\partial \beta}A^{-۱}.$$

بنابراین نتیجه نهایی به صورت زیر است:

$$\frac{\partial A^{-۱}}{\partial \beta} = -A^{-۱}\frac{\partial A}{\partial \beta}A^{-۱}.$$

پرسش ۴ (۱۶ نمره) در بسیاری از مسائل حداقل مربعات غیر خطی تابع باقیمانده به این صورت است :

$$f_i(x) = \Phi_i(a_i^T x - b_i), \quad i = ۱, \cdots, m$$

$$a_i \in \mathbb{R}^n \quad , \quad b_i \in \mathbb{R} \quad , \quad \Phi_i : \mathbb{R} \longrightarrow \mathbb{R}$$

در این حالت تابع هدف حداقل مربعات غیر خطی به این فرم است:

$$||f(x)||^۲ = \sum_{i=۱}^{m}(\Phi_i(a_i^T x - b_i))^۲$$

یک ماتریس $A \in \mathbb{R}^{m\times n}$ تعریف می کنیم که سطر های آن $a_۱^T, \cdots, a_m^T$ را داشته باشد. همچنین یک بردار $b \in \mathbb{R}^m$ تعریف می کنیم.
نشان دهید مشتق تابع به فرم زیر است :

$$Df(x) = diag(d)A$$
$$d_i = \Phi_i'(r_i) \quad , \quad r = Ax - b$$

پاسخ

$$Df(x) = \begin{bmatrix} \frac{\partial f_۱}{\partial x} \\ \vdots \\ \frac{\partial f_m}{\partial x} \end{bmatrix}$$

$$r = Ax - b \Longrightarrow r_i = a_i^T x - b_i \Longrightarrow \frac{\partial r_i}{\partial x} = a_i^T$$

$$\frac{\partial f_i}{\partial x} = \frac{\partial f_i}{\partial r_i}\frac{\partial r_i}{\partial x} = \Phi_i'(r_i)a_i^T = d_i a_i^T$$

$$Df(x) = \begin{bmatrix} d_۱ a_۱^T \\ \vdots \\ \vdots \\ d_m a_m^T \end{bmatrix} = \begin{bmatrix} d_۱ & ۰ & \cdots & ۰ \\ ۰ & d_۲ & \ddots & \vdots \\ \vdots & \ddots & \ddots & ۰ \\ ۰ & \cdots & ۰ & d_m \end{bmatrix} \begin{bmatrix} a_۱^T \\ \vdots \\ \vdots \\ a_m^T \end{bmatrix} = diag(d)A$$

پرسش ۵ (۱۶ نمره) کمینه‌سازی زیر را در نظر بگیرید:

$$\begin{aligned} &\min \|\mathbf{x}\|_2 \\ &\text{subject to } A\mathbf{x} = \mathbf{y}. \end{aligned}$$

به وضوح عبارت زیر در شرط مسئله صدق میکند:

$$x_۰ = A^T(AA^T)^{-۱}\mathbf{y}.$$

(آ) اثبات کنید که هر جوابی برای $A\mathbf{x} = \mathbf{y}$ به صورت زیر قابل نمایش است:

$$x = x_۰ + \mathbf{w},$$

که $\mathbf{w}$ در $N(A)$ قرار دارد.

(ب) به کمک بخش قبل نشان دهید $\mathbf{w} \perp x_۰$.

(ج) به کمک بخش قبل نشان دهید $x_۰$ کوچک ترین جواب و به عبارتی همان خواسته مسئله است.

(د) حال سعی کنید با ضرایب لاگرانژ $x_۰$ را استخراج کنید.

(ه) حال مسئله را به شکل زیر تغییر دهید:

$$J(\mathbf{x}) = \|A\mathbf{x} - \mathbf{y}\|_۲^۲ + \lambda\|\mathbf{x}\|_۲^۲.$$

این مسئله به Regularized Least Squares (RLS) معروف است. به طور کلی Regularization روشی است که برای جلوگیری از overfitting در مسائل بهینه‌سازی به کار می‌رود. این تکنیک با اضافه کردن یک جمله‌ی پنالتی به تابع هدف، پیچیدگی مدل را کاهش می‌دهد و تعادل بین برازش داده‌ها و ساده‌سازی مدل را حفظ می‌کند. در اینجا، مسئله‌ی RLS با اضافه کردن نرم ۲ متغیر هدف به تابع هزینه تعریف می‌شود. جواب آن را با استفاده از مشتق گیری به دست آورید و رابطه‌ی آن را با مسئله اولیه بنویسید.

پاسخ

(آ) از $A\mathbf{x} = \mathbf{y}$ نتیجه می‌گیریم:

$$A(\mathbf{x} - \mathbf{x}_۰) = y - y = ۰$$

بنابراین:

$$\mathbf{x} - \mathbf{x}_۰ \in N(A)$$

(ب)

$$(x - x_۰)^T x_۰ = (x - x_۰)^T A^T(AA^T)^{-۱}y$$

$$= (A(x - x_۰))^T A^T(AA^T)^{-۱}y = ۰$$

(ج)

$$\|\mathbf{x}\|^۲ = \|\mathbf{x}_۰ + \mathbf{w}\|^۲ = \|\mathbf{x}_۰\|^۲ + \|\mathbf{w}\|^۲ \geq \|\mathbf{x}_۰\|^۲$$

(د) ضرایب لاگرانژ را تعریف می‌کنیم:

$$L(\mathbf{x}, \lambda) = \|\mathbf{x}\|^۲ + \lambda^T(A\mathbf{x} - \mathbf{y})$$

با مشتق‌گیری:

$$\begin{aligned} \nabla_{\mathbf{x}} L &= ۲\mathbf{x} + A^T\lambda = ۰ \\ \nabla_\lambda L &= A\mathbf{x} - \mathbf{y} = ۰ \\ \implies \mathbf{x}_۰ &= A^T(AA^T)^{-۱}\mathbf{y} \end{aligned}$$

(ه) تابع هزینه به صورت زیر است:

$$J(\mathbf{x}) = \|A\mathbf{x} - \mathbf{y}\|_۲^۲ + \lambda\|\mathbf{x}\|_۲^۲$$

$$J(\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} - ۲\mathbf{y}^T A\mathbf{x} + \mathbf{y}^T\mathbf{y} + \lambda\mathbf{x}^T\mathbf{x}$$

مشتق آن نسبت به $\mathbf{x}$:

$$\nabla J(\mathbf{x}) = ۲A^T A\mathbf{x} - ۲A^T\mathbf{y} + ۲\lambda\mathbf{x}$$

برای کمینه کردن، مشتق برابر صفر قرار می‌گیرد:

$$(A^T A + \lambda I)\mathbf{x} = A^T\mathbf{y}$$

و در نهایت:

$$\mathbf{x}_\lambda = (A^T A + \lambda I)^{-۱} A^T\mathbf{y}$$

وقتی که $\mathbf{y}$ در $R(A)$ قرار دارد و لاندا برابر با ۱ است معادل با مسئله قبل است

پرسش ۶ (۲۰ نمره)

فرض کنید مسئله کمترین مربعات به فرم زیر داده شده است:

$$\min_x \left(\|Ax - b\|_۲^۲ + c^T x + d\right),$$

که در آن:

- A یک ماتریس $m \times n$،
- b برداری در $\mathbb{R}^m$،
- c برداری در $\mathbb{R}^n$،
- و d یک اسکالر است.

(آ) این مسئله را به فرم استاندارد کمترین مربعات ($\min_x \|Tx - v\|$) تبدیل کنید.

(ب) مسئله تبدیل‌شده را حل کرده و مقدار بهینه x را به دست آورید.

پاسخ

(آ) **تبدیل به فرم استاندارد کمترین مربعات:**

ابتدا عبارت $\|Ax - b\|_۲^۲$ را بازنویسی می‌کنیم:

$$\|Ax - b\|_۲^۲ = (Ax - b)^T(Ax - b) = x^T A^T Ax - ۲b^T Ax + b^T b.$$

حال کل تابع هدف را بازنویسی می‌کنیم:

$$\|Ax - b\|_۲^۲ + c^T x + d = x^T A^T Ax - ۲b^T Ax + b^T b + c^T x + d.$$

ترم‌های خطی و ثابت را دسته‌بندی می‌کنیم:

$$= x^T A^T Ax + (-۲A^T b + c)^T x + (b^T b + d).$$

بنابراین، مسئله به فرم استاندارد کمترین مربعات تبدیل می‌شود:

$$\min_x \frac{۱}{۲} x^T Qx + q^T x + r,$$

که در آن:

- $Q = ۲A^T A$ (ماتریس ضرایب درجه دوم)،
- $q = -۲A^T b + c$ (بردار ضرایب خطی)،
- $r = b^T b + d$ (مقدار ثابت).

(ب) **حل مسئله و یافتن مقدار بهینه x:**

برای پیدا کردن مقدار بهینه x، گرادیان تابع هدف را نسبت به x محاسبه کرده و برابر صفر قرار می‌دهیم:

$$f(x) = x^T A^T Ax - ۲b^T Ax + c^T x + b^T b + d.$$

گرادیان $f(x)$ نسبت به x برابر است با:

$$\nabla f(x) = ۲A^T Ax - ۲A^T b + c.$$

شرط ایستایی ($\nabla f(x) = ۰$) به صورت زیر است:

$$۲A^T Ax - ۲A^T b + c = ۰.$$

این معادله را حل می‌کنیم:

$$۲A^T Ax = ۲A^T b - c.$$

دو طرف را بر ۲ تقسیم می‌کنیم:

$$A^T Ax = A^T b - \frac{۱}{۲}c.$$

اگر $A^T A$ معکوس‌پذیر باشد، مقدار x بهینه برابر است با:

$$x^* = (A^T A)^{-۱} \left(A^T b - \frac{۱}{۲}c \right).$$