۸۱۰۱۰۰۴۴۱ HW6 محمدرضا کاشانی

1- حداقل یک تصویر برای هر مجموعه‌ی کانوکس بسته‌ای به ازای هر norm موجود است و این از دانسته‌های ماست.

حال ما یک $\bar{x} \in C$ در نظر می‌گیریم و تصویر را کمینه کردن تابع $\|x-x_0\|$ به روی تبدیل بسته که هم اشتراک C با مجموعه‌ی $\{x \mid \|x-x_0\| \leq \|\bar{x}-x_0\|\}$ که تصویر باید مقدار کمینه‌ای از این مقدار $\bar{x}$ کمینه‌تر باشد که جواب وجود دارد.

حال با نشان دادن ... اکید بودن نورم که همان به معنی یکتا بودن تصویر است در نظر می‌گیریم u و v که $u \neq v$ است و هر دو تصویر x_0 روی C باشند پس داریم:

$$\|u-x_0\| = \|v-x_0\| = D \longrightarrow$$

حال چون u و v در C هستند پس چون C کانوکس است $\frac{u+v}{2}$ هم در C وجود دارد ← حال

نامساوی مثلثی

$$\left\|\frac{u+v}{2}-x_0\right\| = \left\|\frac{1}{2}(u-x_0)+\frac{1}{2}(v-x_0)\right\| < \frac{1}{2}\|u-x_0\|+\frac{1}{2}\|v-x_0\| = D$$

پس نقطه‌ی $\frac{u+v}{2}$ فاصله‌ی کمتری پیدا می‌کند ← پس u و v تصویر نیستند ← پس فقط یک تصویر داریم

2- ما باید بیضی‌گونی را داشته باشیم که تمام بیضی‌گون‌های ε_i را شامل شود که بیضی‌گون ε باید شامل تمام تقاطع ε_i ها باشد یعنی برای تمام نقاط ε_i در فرمول بیضی‌گون در ε صدق کند که به این معنی است:

$$\sup_{x \in \varepsilon_i} \left[(x-x_0)^T A^{-1} (x-x_0)\right] \leq 1$$

حال این تعریف را بازنویسی می‌کنیم تا با تعریف ε_i ها یکی شود:

$$x^T A_i x + 2 b_i^T x + c_i \leq 0 \Rightarrow x^T A^{-1} x - 2 x_0^T A^{-1} x + x_0^T A^{-1} x_0 - 1 \leq 0$$

از S-procedure در Appendix-B، رابطه بالا درست است اگر و تنها اگر $\lambda_i \geq 0$ وجود داشته باشد که

$$\lambda_i \begin{bmatrix} A_i & b_i \\ b_i^T & c_i \end{bmatrix} \succeq \begin{bmatrix} A^{-1} & -A^{-1}x_0 \\ -(A^{-1}x_0)^T & x_0^T A^{-1} x_0 - 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \lambda_i A_i & \lambda_i b_i \\ \lambda_i b_i^T & \lambda_i c_i \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} A^{-1} & -A^{-1}x_0 \\ -(A^{-1}x_0)^T & x_0^T A^{-1} x_0 \end{bmatrix} \succeq 0$$

$$\Rightarrow \begin{bmatrix} \lambda_i A_i & \lambda_i b_i \\ \lambda_i b_i^T & \lambda_i c_i + 1 \end{bmatrix} - \begin{bmatrix} I \\ -x_0^T \end{bmatrix} A^{-1} \begin{bmatrix} I & -x_0 \end{bmatrix} \succeq 0 \xrightarrow{\text{sure complement}} \begin{bmatrix} A & I & -x_0^T \\ I & \lambda_i A_i & \lambda_i b_i \\ -x_0 & \lambda_i b_i^T & 1+\lambda_i c_i \end{bmatrix} \succeq 0$$

نامساوی LMI شد

حال می‌خواهیم حجم این بیضی‌گون را کمینه کنیم که حجم آن به $\log\det A^{-1}$ تعبیر می‌شود.

$$\text{minimize} \quad \log\det A^{-1}$$

$$\text{s.t.} \quad \begin{bmatrix} A & I & -x_0^T \\ I & \lambda_i A_i & \lambda_i b_i \\ -x_0 & \lambda_i b_i^T & 1+\lambda_i c_i \end{bmatrix} \succeq 0 \;;\; i=1,\dots,K \;,\; \lambda_i \geq 0 \;;\; i=1,\dots,K$$

که متغیرهای ما $A\in S^n$ و $x_0\in R^n$ و λ_i برای تمام iها است

3-a) وقتی پنالتی فانکشن L2_norm باشد برای مقادیر زیاد مجازات‌های زیادتری را در نظر می‌گیرد و حدود را در یک اندازه کمینه می‌کند یعنی نمی‌تواند مقداری بیشتر از بقیه باشد و تمامی فواصل باید مساوی شود پس حاصل می‌شود:

$$p^{(i)} = \left(1-\frac{i-1}{N-1}\right)q + \left(\frac{i-1}{N-1}\right)r \rightarrow i=1,\dots,N$$

به طور مساوی دیتاها طوری انتخاب می‌شوند.

b) Hellinger $\Leftarrow$ $d^{hel}(u,v) = \sum_{i=1}^{n} (\sqrt{u_i}-\sqrt{v_i})^2$

مسئله را باز کنیم $\Rightarrow d^{hel}(u,v) = \sum_{i=1}^{n} u_i + v_i - 2\sqrt{u_i v_i} = (u+v)^T 1 - 2\sum_{i=1}^{n}\sqrt{u_i v_i}$

حاصل جمع دو affine منهای دو برابر یک تابع concave که convex است

c) Kolmogorov $\Leftarrow$ $d^{Kol}(u,v) = \max_{i=1,\dots,n} \left|\sum_{j=1}^{i} u_j - \sum_{j=1}^{i} v_j\right|$

قدر مطلق تابعی convex است و داخل آن یک تابع affine وجود دارد پس حاصل یک تابع convex می‌شود و max توابع convex برابر convex است

d) که ضمیمه شده است مسئله به صورت زیر است.

بایستن

$$\text{minimize} \quad \sum_{i=1}^{N-1} d(p^{(i)}, p^{(i+1)})$$

$$\text{s.t.} \quad p^{(i)} \succeq 0 \,,\; 1^T p^{(i)} = 1 \,,\; i=1,2,\dots,N$$

$$p^{(1)} = q \,,\; p^{(N)} = r$$

۴_ مسئله برابر این صورت است که $h(x)$ را باید min کرده و x مورد نظر را پیدا کنیم:

(a

$$\min h(x) = \frac{\lambda}{2}\sum_{i=1}^{n}\|x-\mathrm{prox}^{\lambda}_{f_i}(x)\|_2^2+\sum_{i=1}^{m} f_i(\mathrm{prox}^{\lambda}_{f_i})$$

$$\mathrm{prox}^{\lambda}_{f}(x)=\arg\min_y\left\{f(y)+\frac{\lambda}{2}\|y-x\|_2^2\right\}$$

یک متغیر جدید z_i اضافه می‌کنیم که $z_i=\mathrm{prox}^{\lambda}_{f_i}(x)$ و داریم:

$$\min \frac{\lambda}{2}\sum_{i=1}^{n}\|x-z_i\|_2^2+\sum_{i=1}^{n} f_i(z_i)$$

$$\text{s.t.}\quad z_i=\arg\min_{z_i}\left\{f(z_i)+\frac{\lambda}{2}\|z_i-x\|_2^2\right\}$$

هر دو convex بوده پس مسئله کانوکس است

(b تابع prox را ابتدا حل می‌کنیم: $f_i(x)=\frac{1}{2}(a_i^Tx-b_i)^2$

$$\mathrm{prox}^{\lambda}_{f}(x)=\arg\min_y\left\{\frac{1}{2}(a_i^Ty-b_i)^2+\frac{\lambda}{2}\|y-x\|_2^2\right\}$$

$$\frac{\partial\,\mathrm{prox}}{\partial y}=0\rightarrow (a_i^Ty-b_i)a_i+\lambda(y-x)=0\Rightarrow y(a_i^Ta+\lambda)-b_ia_i-\lambda x=0$$

$$\Rightarrow y_i=\frac{b_ia_i-\lambda x}{a_i^Ta+\lambda}=\frac{-\lambda}{a_i^Ta+\lambda}x+\frac{b_ia_i}{a_i^Ta+\lambda}=a'_ix+b'_i$$

$$\Rightarrow \mathrm{prox}^{\lambda}_{f_i}(x)=y_i$$

در h جایگذاری

$$\rightarrow h(x)=\frac{\lambda}{2}\sum_{i=1}^{m}\|x-y_i\|_2^2+\sum_{i=1}^{m}\frac{1}{2}(a_i^Ty_i-b_i)^2$$

$$=\frac{\lambda}{2}\sum_{i=1}^{m}\|x-a'_ix-b'_i\|_2^2+\sum_{i=1}^{m}\frac{1}{2}(a(a'_ix+b'_i)-b_i)^2$$

$$=\frac{\lambda}{2}\sum_{i=1}^{m}\|(1-a'_i)x-b'_i\|_2^2+\sum_{i=1}^{m}\frac{1}{2}(aa'_ix+ab'_i-b_i)^2=*$$

نسبت به x quadratic است

$$\rightarrow \min *$$

۴- تابع D را باید نسبت به قیود max کنیم

max $D(x)$ $\longrightarrow$ کانوکس نیست

s.t $1^T x = 1$ و $0 \le x \le M$.

بزرگتر از صفر را باید چک کرد

$$D(x) = \frac{a^T x}{(x^T \Sigma x)^{1/2}} \xrightarrow[\text{set}]{\alpha\text{-superlevel}} \frac{a^T x}{(x^T \Sigma x)^{1/2}} \ge \alpha \to \underbrace{0 \ge \alpha (x^T \Sigma x)^{1/2} - a^T x}_{\text{zero-sublevel set}}$$

$(x^T \Sigma x)^{1/2}$ کانوکس است تابع Least square است در منفی یک affine شده پس 0-sub... کانوکس

شد برای تمام مسئله ها [illegible] همه convex هر [illegible] مسئله quasiconcave است.

می توانیم با تغییر متغیر $x = \frac{y}{k}$ داشته باشیم حالت زیر را به شرطی که $a^T y = 1$ و $k > 0$ باشد:

$$D(y,k) = \frac{a^T y / k}{((y/k)^T \Sigma (y/k))^{1/2}} = \frac{a^T y}{(y^T \Sigma y)^{1/2}} \xrightarrow{a^T y = 1} \frac{1}{(y^T \Sigma y)^{1/2}}$$

این را باید max کنیم که یعنی $(y^T \Sigma y)^{1/2}$ را min کنیم.

شروط زیر را داریم:

$0 \le x \le M \to 0 \le y/k \le M \to 0 \le y \le kM$

به علاوه $k \ge 0$

$1^T x = 1 \to 1^T \frac{y}{k} = 1 \to 1^T y = k$

مسئله به شکل زیر در آمد

min $y^T \Sigma y$

s.t $a^T y = 1$, $0 \le y \le kM$, $1^T y = k$

- کد پایتون ضمیمه شده است.

۵- برای سوال پنجم کد پایتون ضمیمه شد که کد پایتون مسئله بهینه سازی حل شده تمام outlier را به درستی جدا می کند که یعنی مسئله درست حل شده است اما به علت نبود تابع plot بیضی ها در مسئله علاقمند [illegible] تلاش برای کشیدن بیضی ها و نوشتن تابع ellipse_draw که قرار بود داده مسئله باشد موفق به کشیدن بیضی ها درست نشدم.

۷- a) مسئله کمینه کردن بیشترین فاصله از y_kها است که $\|y_k - x\|_2$ لیست اسکور است و تابع کانوکس همینطور max جند تابع کانوکس کانوکس است که کمینه کل نقاط ما کانوکس شده است.

b) مسئله socp باید یک نقطه را نشان داشته باشد که در شرط یک least squar affine کوچکتر مساوی یک affine دیگر شود که به این صورت میتوان نوشت:

$$\min_x \max_{k=1,\dots,n} \|y_k - x\|_2 \equiv \min_{\theta, x} \theta$$

$$\text{s.t.} \quad \|y_k - x\|_2 \le \theta \quad \text{for} \quad k=1,\dots,n$$

در اینجا θ یک آپرباند برای تمامی فاصله‌ها است و اگر ما θ را کمینه کنیم در حقیقت θ تا حد امکان آپرباند کوچکتری برای تمام فاصله‌ها میشود که معادل این است که بیشترین فاصله را با تغییر x کمینه کنیم.

شرط $\|y_k - x\|_2^2 \le \theta^2$

c) به دست آوردن تابع لاگرانژین دوال:

$$L(x, t, \lambda) = \theta + \sum_{k=1}^{n} \lambda_k \left(\|y_k - x\|_2^2 - \theta^2\right)$$

$$\frac{\partial L}{\partial x} = -2 \sum \lambda_k (y_k - x) = 0 \longrightarrow x = \frac{1}{\sum \lambda_k} \left(\sum \lambda_k y_k\right)$$

$$\frac{\partial L}{\partial \theta} = 1 - 2 \sum \lambda_k \theta = 0 \longrightarrow \theta = \frac{1}{2 \sum \lambda_k}$$

$$g(\lambda) = \frac{1}{2\sum \lambda_k} + \sum_{k=1}^{n} \lambda_k \left(\left\| y_k - \frac{1}{\sum \lambda_k} \left(\sum \lambda_k y_k\right) \right\|_2 - \left(\frac{1}{2\sum \lambda_k}\right)^2 \right)$$

$$\Longrightarrow g(\lambda) = \frac{1}{4 \sum \lambda_k} + \sum_{k=1}^{n} \lambda_k \left\| y_k - \frac{1}{\sum \lambda_k} \left(\sum \lambda_k y_k\right) \right\|_2$$

حال مسئله اصلی کانوکس است و میتوان در شرط یک نقطه پیدا کرد که strictly شدنی شود پس شرایط slater را داریم و strong duality برای ما برقرار است.

d) با حل مسئله دوال ما یک λ^* داریم و با حل مسئله primal ما یک x^* و از تابع لاگرانژین نقطه زینی چون strong duality برقرار است λ^* و θ^* و x^* را به ما میدهد یعنی با پیدا کردن λ^* و جایگذاری در L و بهینه‌سازی کردن x و t مقادیر x^* و θ^* به دست می‌آید پس داریم:

$$x^* = \frac{1}{\sum \lambda_k^*} \left(\sum \lambda_k^* y_k\right) \quad \text{و} \quad \theta^* = \frac{1}{2 \sum \lambda_k^*}$$

۸- ما باید یک مجموعه شدنی پیدا کنیم که P، D و A را برای دیزاین ما ارضا کنند. با استفاده از n مثال $w^{(k)}$ که داده شده‌اند

ما یک مجموعه w داریم که به صورت زیر تعریف می‌کنیم:

$$w = \theta_1 w^1 + \theta_2 w^2 + \dots + \theta_k w^k$$

$$1^T\theta = 1 \quad , \quad \theta \succeq 0$$

حال بنابر نامساوی Jensen داریم:

$$f(w) \leq \theta_1 f(w^1) + \dots + \theta_k f(w^k)$$

ولی برای تابع f داریم که باید کانوکس باشند که توابع پوزی نومیال P و D و A کانوکس نیستند که با تغییر متغیر باید این توابع را کانوکس کنیم ← تابع $\log(f(e^x))$ کانوکس است

پس جوری تغییر متغیر می‌دهیم که این تابع درست شود که کانوکس را داریم که این با $x = \log w$ ای می‌شود و توابع A و D و P با این تغییر متغیر کانوکس می‌شوند. می‌توانیم نامساوی را برای آن بنویسیم

$$\log P(\exp x) \leq \theta_1 \log P(\exp x^{(1)}) + \dots + \theta_k \log P(\exp x^{(k)})$$

$$\theta \succeq 0 \quad , \quad 1^T\theta = 1 \quad , \quad x^{(i)} = \log w^{(i)}$$

$$\Rightarrow \log P(w) \leq \theta_1 \log P(w^1) + \dots + \theta_k \log P(w^k)$$
$$= \theta_1 \log P^1 + \dots + \theta_k \log P^k \Rightarrow *$$

$$\log w = \theta_1 \log w^{(1)} + \dots + \theta_k \log w^k \quad ,$$

حال چون جمله‌ی $*$ یک آپر باند برای $\log P(w)$ است برای w هایی که بتوان با w ها و θ نوشت)

$$\Rightarrow w_i = (w_i^{(1)})^{\theta_1} \times \dots \times (w_i^{(k)})^{\theta_k}$$

حال اگر این آپر باند از مقادیر مشخص شده‌ی spec کوچکتر شود برای P و A، D در این صورت w مورد نظر هم این شروط را ارضا می‌کند. مسئله ما به این صورت می‌شود: feasible set

$$\text{minimize } \theta \quad (\text{find } \theta)$$

$$\text{s.t.} \quad \sum_{j=1}^{k} \theta_j \log P^{(j)} \leq \log P_{spec}$$

$$\sum_{j=1}^{k} \theta_j \log D^{(j)} \leq \log D_{spec}$$

$$\sum_{j=1}^{k} \theta_j \log A^{(j)} \leq \log A_{spec}$$

$$1^T\theta = 1, \quad \theta \succeq 0$$

برای θ هایی که به دست می‌آید یک w می‌شود نوشت که در شرایط feasible مسئله باشد.

کد قسمت دوم پیوست شد