بنام خدا

علی قاسم زاده ۴۰۱۱۶۴۳۳۹

۱- سوال ۵.۱۰ کتاب:

a) $$\min \quad \log\det\left(\sum_{i=1}^{p} x_i v_i v_i^T\right)^{-1} \quad \text{s.t.} \quad x \succeq 0,\ 1^T x = 1 \xrightarrow[X = \sum x_i v_i v_i^T]{\text{طبق صورت سوال}} \min \quad \log\det X^{-1} \quad \text{s.t.} \quad X = \sum_{i=1}^{p} x_i v_i v_i^T,\ x \succeq 0,\ 1^T x = 1$$

حالا از مسئله مان dual می گیریم:

$$L(X, x, Z, z, \lambda) = \log\det X^{-1} + tr\left(Z\left(X - \sum_{i=1}^{p} x_i v_i v_i^T\right)\right) - z^T x + \lambda(1^T x - 1) =$$

$$\log\det X^{-1} + tr(ZX) - tr\left(Z \sum_{i=1}^{p} x_i v_i v_i^T\right) - z^T x + \lambda(1^T x - 1) =$$

$$-\log\det X + tr(ZX) - tr\left(\sum_{i=1}^{p} x_i v_i^T Z v_i\right) - z^T x + \lambda(1^T x) - \lambda =$$

$$-\log\det X + tr(ZX) - \sum_{i=1}^{p} x_i v_i^T Z v_i - z^T x + \lambda(1^T x) - \lambda =$$

$$-\log\det X + tr(ZX) - \sum_{i=1}^{p} x_i (v_i^T Z v_i + z_i - \lambda) - \lambda =$$

حالا اگر نسبت به X مشتق بگیریم و برابر صفر قرار دهیم، خواهیم داشت که:

$$-X^{-1} + Z = 0 \longrightarrow Z = X^{-1}$$

پس داریم که:

$$\log\det Z + n - \sum_{i=1}^{p} x_i (v_i^T Z v_i + z_i - \lambda) - \lambda$$

حالا اگر از این عبارت نسبت به x مشتق بگیریم خواهیم داشت که:

$$\forall i: \quad -(v_i^T Z v_i + z_i - \lambda) = 0 \longrightarrow z_i = \lambda - v_i^T Z v_i$$

$$\Rightarrow \log\det(Z) + n - \lambda$$

Scanned with CamScanner

$$\Rightarrow g(Z,z) = \begin{cases} \log\det(Z) + n - \lambda & \text{if } \lambda - v_i^T Z v_i = z_i, \quad i = 1 \text{---} p \\ -\infty & \text{else} \end{cases}$$

$$\rightarrow \max \quad \log\det(Z) + n - \lambda$$

$$\text{s.t.} \quad \lambda - v_i^T Z v_i = z_i \quad i = 1 \text{---} p$$

$$z_i \geq 0$$

$\lambda = z_i + v_i^T Z v_i$

و داریم $z_i \geq 0$ پس این قید را می‌توانیم با $v_i^T Z v_i \leq \lambda$ جایگزین کنیم.

$$\Rightarrow \max \quad \log\det(Z) + n - \lambda$$

$$\text{s.t.} \quad v_i^T Z v_i \leq \lambda \quad i = 1 \text{---} p$$

همچنین می‌توانیم ماتریس Z/λ را w بنامیم در آن صورت خواهیم داشت که:

$$\max \quad \log\det(W\lambda) + n - \lambda \qquad\qquad \max \quad \log\det(W) + n\log\lambda + n - \lambda$$

$$\text{s.t.} \quad v_i^T W v_i \leq 1 \quad i = 1 \text{---} p \qquad \rightarrow \qquad \text{s.t.} \quad v_i^T W v_i \leq 1 \quad i = 1 \text{---} p$$

از طرفی با مشتق گرفتن نسبت به λ، بهترین مقدار برای آن برابر است با:

$$\lambda = n \leftarrow \frac{n}{\lambda} - 1 = 0$$

پس به جای λ، n قرار می‌دهیم، خواهیم داشت که:

$$\max \quad \log\det(W) + n\log n$$

$$\text{s.t.} \quad v_i^T W v_i \leq 1 \quad i = 1 \text{---} p$$

b) $\min \ tr\left(\sum_{i=1}^{p} x_i v_i v_i^T\right)^{-1}$

s.t. $x \succeq 0, \ 1^T x = 1$

$\xrightarrow{\text{طبق صورت سوال}}$

$\min \ tr(X^{-1})$

s.t. $X = \sum_{i=1}^{p} x_i v_i v_i^T$

$x \succeq 0, \ 1^T x = 1$

$$\xrightarrow{\text{dual}} L(X, x, Z, z, \lambda) = tr(X^{-1}) + tr\left(Z\left(X - \sum_{i=1}^{p} x_i v_i v_i^T\right)\right) - z^T x + \lambda(1^T x - 1)$$

$$\xrightarrow{\text{مشتق به } X} -X^{-2} + Z = 0 \longrightarrow Z = X^{-2} \longrightarrow X^{-1} = Z^{1/2}$$

$$\longrightarrow 2tr(Z^{1/2}) - \underbrace{tr\left(Z\left(\sum_{i=1}^{p} x_i v_i v_i^T\right)\right)}_{-\sum_{i=1}^{p} x_i v_i^T Z v_i} - z^T x + \lambda(1^T x - 1)$$

$$\longrightarrow 2tr(Z^{1/2}) - \sum_{i=1}^{p} x_i (v_i^T Z v_i + z_i - \lambda) - \lambda$$

$$\xrightarrow{\text{مشتق به } x} \forall i \quad v_i^T Z v_i + z_i - \lambda = 0 \longrightarrow$$

$$g(Z, z, \lambda) = \begin{cases} 2tr(Z^{1/2}) - \lambda & \text{if} \quad \forall i \quad \underbrace{v_i^T Z v_i + z_i - \lambda = 0}_{v_i^T Z v_i + z_i = \lambda}, \ Z \succeq 0 \\ -\infty & \text{o.w.} \end{cases}$$

$$\longrightarrow v_i^T Z v_i \leq \lambda$$

$$\longrightarrow \text{maximize} \quad 2tr(Z^{1/2}) - \lambda$$

s.t. $v_i^T Z v_i \leq \lambda \quad i = 1 - p$

$Z \succeq 0$

از آنجایی X ماتریسی متقارن است که درنتیجه X^{-1} نیز ماتریسی متقارن است ،

$$(X^{-1})^T = (X^T)^{-1} = X^{-1}$$

حالا از طرفی X^{-2} هم ماتریسی متقارن است ؛

$$((X^{-1})^2)^T = (X^{-1} X^{-1})^T = X^{-T} X^{-T} = X^{-1} X^{-1} = X^{-2}$$

پس ماتریس Z را می‌توان به صورت WW نوشت چون هم PSD است و هم متقارن:

$W = Z^{1/2}$

$$\begin{aligned} \max \quad & 2\,tr(Z^{1/2}) - \lambda \\ s.t. \quad & Z \succeq 0, \\ & v_i^T Z v_i \le \lambda \quad i=1 \text{---} p \end{aligned} \longrightarrow \begin{aligned} \max \quad & 2\,tr(W) - \lambda \\ s.t. \quad & v_i^T W W v_i \le \lambda \quad i=1\text{---}p \end{aligned}$$

قید $WW \succeq 0$ هم که همواره برقرار است. حالا برای ساده‌تر شدن عبارات را بر λ تقسیم می‌کنیم، خواهیم داشت که: $M = W/\sqrt{\lambda}$:

$$\begin{aligned} \max \quad & 2\,tr(M\sqrt{\lambda}) - \lambda \\ s.t. \quad & v_i^T M M v_i \le 1 \quad i=1\text{---}p \end{aligned} \xrightarrow{\text{مشتق به } \lambda} \frac{tr(M)}{\sqrt{\lambda}} - 1 = 0 \longrightarrow tr(M) = \sqrt{\lambda}$$

$$\longrightarrow \begin{aligned} \max \quad & 2\,tr(M)\sqrt{\lambda} - \lambda \\ s.t. \quad & v_i^T M M v_i \le 1 \quad i=1\text{---}p \end{aligned} \longrightarrow \begin{aligned} \max \quad & tr(M)^2 \\ s.t. \quad & v_i^T M M v_i \le 1 \quad i=1\text{---}p \end{aligned}$$

سوال ۵.۴۰ کتاب:

$$\begin{aligned} \min \quad & \lambda_{max}\Big(\sum_{i=1}^{p} x_i v_i v_i^T\Big)^{-1} \\ s.t \quad & x \succeq 0, \quad 1^T x = 1 \end{aligned}$$

یک کران پایین روی ماتریس $\sum_{i=1}^{p} x_i v_i v_i^T$ در نظر می‌گیریم، به صورت tI پس داریم که:

$$tI \le \Big(\sum_{i=1}^{p} x_i v_i v_i^T\Big) \longrightarrow \Big(\sum_{i=1}^{p} x_i v_i v_i^T\Big) - tI \succeq 0 \longrightarrow$$

$$\lambda_{min}\Big(\sum_{i=1}^{M} x_i v_i v_i^T\Big) \ge t \longrightarrow$$

از تمرین اول می‌دانیم که $\lambda_{max}(A^{-1})$ برابر است با $\frac{1}{\lambda_{min}(A)}$ پس داریم که:

$$\lambda_{max}\Big(\sum_{i=1}^{p} x_i v_i v_i^T\Big)^{-1} \le \frac{1}{t}$$

درنتیجه حالا این $\frac{1}{t}$ ها هم قابل وصول است، پس می‌توانیم بجای کمینه کردن λ_{max}، کران بالایش یعنی $\frac{1}{t}$ را کمینه کنیم و یک قید به قیود مسئله اضافه کنیم. پس خواهیم داشت:

$$\begin{aligned} \min \quad & \frac{1}{t} \\ \text{s.t.} \quad & \sum_{i=1}^{p} x_i v_i v_i^T \succeq tI \\ & x \succeq 0, \quad 1^T x = 1 \end{aligned}$$

حالا مسئله‌ی جدید را باید برای فرم dual بنویسیم، خواهیم داشت که:

$$L(t, x, Z, z, \lambda) = \frac{1}{t} - tr\left(Z\left(\sum_{i=1}^{p} x_i v_i v_i^T - tI\right)\right) - z^T x + \lambda(1^T x - 1) =$$

$$\frac{1}{t} + t\, tr(Z) - \sum x_i v_i^T Z v_i - z^T x + \lambda(1^T x) - \lambda =$$

$$\frac{1}{t} + t\ tr(Z) - \sum x_i(v_i^T Z v_i + z_i - \lambda) - \lambda$$

حالا باز با مشتق گیری نسبت به x مانند مسائل قبلی قید $v_i^T Z v_i + z_i - \lambda = 0$ ایجاد می‌شود که اگر برقرار باشد، رابطه مینیمم دارد وگرنه $-\infty$. پس خواهیم داشت که:

$$\begin{cases} \frac{1}{t} + t\ tr(Z) - \lambda & \text{if} \quad v_i^T Z v_i + z_i - \lambda = 0 \\ -\infty & \text{o.w.} \end{cases}$$

حالا نسبت به t بهینه سازی می‌کنیم، خواهیم داشت که،

$$-\frac{1}{t^2} + tr(Z) = 0 \rightarrow tr(Z) = \frac{1}{t^2} \rightarrow$$

$$g(Z, z) = \begin{cases} 2\sqrt{tr(Z)} - \lambda & v_i^T Z v_i + z_i - \lambda = 0, \quad Z \succeq 0 \\ -\infty & \text{o.w.} \end{cases}$$

پس می‌توانیم مسئله‌ی مان را به این صورت بنویسیم که:

$$\max \quad 2\sqrt{tr(Z)} - \lambda$$

$$\text{s.t.} \quad \boxed{v_i^T Z v_i + z_i - \lambda = 0 \qquad i=1 - p}$$

$$Z \succeq 0$$

$$z \geq 0$$

این قید و قید $z_i \geq 0$ را هم می‌توانیم به صورت

$$v_i^T Z v_i \leq \lambda$$

بنویسیم. پس خواهیم داشت که،

$$\left.\begin{array}{ll} \max & 2\sqrt{tr(Z)} - \lambda \\ \text{s.t.} & v_i^T Z v_i \leq \lambda \quad i=1 - p \\ & Z \succeq 0 \end{array}\right\} \xrightarrow{Z/\lambda = W} \begin{array}{ll} \max & 2\sqrt{\lambda\, tr(W)} - \lambda \\ \text{s.t.} & v_i^T W v_i \leq 1 \qquad i=1 - p \\ \text{s.t.} & W \succeq 0 \end{array}$$

حال نسبت به λ مشتق می‌گیریم، خواهیم داشت که،

$$\frac{\sqrt{tr(W)}}{\sqrt{\lambda}} - 1 = 0 \longrightarrow tr(W) = \lambda$$

درنتیجه داریم که،

$$\begin{array}{ll} \max & tr(W) \\ \text{s.t.} & v_i^T W v_i \leq 1 \qquad i=1 - p \\ & W \succeq 0 \end{array}$$

$$\text{minimize} \; -\sum_{i=1}^{m} \log(b_i - a_i^T x) \xrightarrow{\text{طبق صورت سوال}} \min \; -\sum_{i=1}^{m} \log(y_i)$$

$$\text{s.t.} \quad y_i = b_i - a_i^T x \qquad i = 1 — m$$

a_iها را همگی را یک ماتریس A در نظر می‌گیریم و b_iها را b.

حالا باید dual این مسئله را بنویسیم:

$$L(y, x, \lambda) = -\sum_{i=1}^{m} \log(y_i) + \underbrace{\sum_{i=1}^{m} \lambda_i (y_i - b_i + a_i^T x)}_{\lambda^T (y - b + Ax)}$$

$$\Rightarrow L(y, x, \lambda) = -\sum_{i=1}^{m} \log(y_i) + \lambda^T (y - b + Ax)$$

حالا با مشتق گیری نسبت به x خواهیم داشت که: $\boxed{A^T \lambda = 0}$ به عنوان یک قید اضافه می‌شود.

$$\rightsquigarrow -\sum_{i=1}^{m} \log(y_i) + \lambda^T y - \lambda^T b$$

$$\sum_{i=1}^{m} \log(\lambda_i)$$

حالا با مشتق گیری نسبت به y خواهیم داشت که:

$$\forall i: -\frac{1}{y_i} + \lambda_i = 0 \longrightarrow \boxed{y_i = 1/\lambda_i}$$

پس خواهیم داشت که:

$$g(\lambda) = \sum_{i=1}^{m} \log(\lambda_i) + m - \lambda^T b \quad \text{if } A^T \lambda = 0 \text{, else } -\infty$$

$$\longrightarrow \max \quad m + \sum_{i=1}^{m} \log(\lambda_i) - \lambda^T b$$

$$\text{s.t.} \quad A^T \lambda = 0$$

این هم فرم مسئله دوگان ما است.

$$\min \; e^{-x} \quad \text{s.t.} \quad x^2/y \le 0 \quad , \quad D=\{(x,y)\mid y>0\}$$

a) $\longrightarrow$ تنها حالتی که داریم این است که $x=0$ هم باشد $\longrightarrow$ مقدار objective مان برابر است با 1.

از طرفی این یک مسئلهٔ convex است زیرا تابع objective که نمایی است و convex است،
همچنین قیدمان هم quad_over_linear است که سر کلاس اثبات شد که تابعی convex است و
$x^2/y \le 0$، ~~...~~ zero-sublevel برای تابع مان است که مجموعه‌ای convex است.
$\leftarrow$ مسئله مان convex است. تنها نقطهٔ feasible اش هم $x=0$ است که در نتیجه
نقطهٔ optimal هم همان است و مقدار optimal برابر است با 1.

b) dual این مسئله را می‌نویسیم، خواهیم داشت که:

$$L(x,y,\lambda) = e^{-x} + \lambda(x^2/y) \longrightarrow g(\lambda) = \inf_{x,y>0} L(x,y,\lambda) =$$

اگر $\lambda<0$ برقرار باشد، آنگاه با انتخاب یک مقدار بزرگ برای x و یک مقدار کوچک برای y، عبارت به $-\infty$ میل
می‌کند. همچنین برای $\lambda\ge 0$، کران پایین عبارت مان صفر است. (x را خیلی بزرگ برداریم و باز y را
نسبت به x^2 خیلی خیلی بزرگتر انتخاب کنیم، عبارت به صفر میل می‌کند.)

$$\rightsquigarrow g(\lambda)=\begin{cases}0 & \lambda\ge 0\\ -\infty & \lambda<0\end{cases} \longrightarrow \max \; 0 \quad \text{s.t.} \; \lambda \ge 0 \longrightarrow d^*=0 \longrightarrow p^*-d^*=1$$

duality gap برابر است با 1.

c) شرایط slater برقرار نیست زیرا کلاً یک مقدار برای x داریم، و $y>0$ داریم که یک نیم‌خط تشکیل
می‌دهند $\leftarrow$ نقاط داخلی نداریم کلاً $\leftarrow$ شرایط slater برقرار نیست.

d) می‌دانیم که داشتیم:

حالا اگر $u=0$ باشد ←

$p^*(u)=1$

چرا که داشتیم:

$$p^*(u,v)=\inf\{f_0(x) \mid \exists x\in D,\ f_i(x)\le u_i,\ i=1 — m,\ h_i(x)=v_i,\ i=1 — p\}$$

وقتی $u=0$ باشد تنها عضو ممکن برای x، مقدار صفر است که $f_0(0)$ هم برابر 1 است ← $p^*(0)=1$

حالا اگر $u>0$ باشد، آنگاه، داریم که:

$$p^*(u)=\inf\{e^{-x} \mid x^2/y\le u\}$$

اگر $u>0$ باشد می‌توانیم x را خیلی بزرگ کنیم و y را طوری انتخاب کنیم که $x^2/y\le u$ باشد. پس داریم که ← e^{-x} برابر صفر می‌شود.

$u>0 \quad p^*(u)=0$

حالا اگر $u<0$ جوابی نداریم و $p^*(u)$ برابر می‌شود با $+\infty$

رابطه ←

$$p^*(u)\ge \underbrace{p^*(0)-\lambda^*u}_{1-\lambda^*u} \longrightarrow p^*(u)\ge 1-\lambda^*u$$

از آنجایی که λ^* براساس x و y انتخاب می‌شود. از L نسبت به x و y مشتق گرفته ایم سه رابطه‌ای به u ندارد و می‌توانیم داشته باشیم که ($u>0$ برای) $\lambda^*u<1$ در نتیجه باید همواره داشته باشیم که

به ازای این uها، که $u>0$ هم هست

$p^*(u)\ge 1-\lambda^*u>0 \longrightarrow$

← $p^*(u)=0$ ← نابرابری نمی‌تواند برقرار باشد. زیرا $0>0$ ✗

۴- a)

$$\text{minimize} \quad x_1^2 + x_2^2$$
$$\text{s.t.} \quad (x_1-1)^2 + (x_2-1)^2 \le 1$$
$$(x_1-1)^2 + (x_2+1)^2 \le 1$$

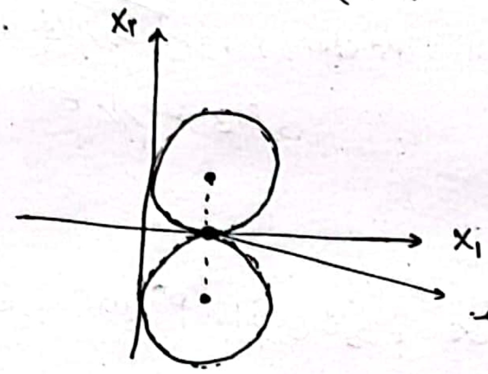
اشتراک دو مجموعه‌ی قیود عبارتند از:

$$\left.\begin{array}{l}(x_1-1)^2 + (x_2-1)^2 \le 1 \\ (x_1-1)^2 + (x_2+1)^2 \le 1\end{array}\right\} \xrightarrow{\text{جمع کردن}} (x_1-1)^2 + 2x_2^2 + 2 \le 2 \longrightarrow$$

$$(x_1-1)^2 + 2x_2^2 \le 0 \longrightarrow x_1 = 1, \ x_2 = 0$$

تنها جواب‌ها هستند.

به صورت شکل هم داریم که:



یک نقطه اشتراک دارند.

در نتیجه x^* همان $\begin{bmatrix}1\\0\end{bmatrix}$ است، و مقدار p^* برابر است با 1.

b) شرایط KKT را برای مسئله مان می‌نویسیم، خواهیم داشت که:

① قیود برقرار باشند:

$$(x_1-1)^2 + (x_2-1)^2 \le 1$$
$$(x_1-1)^2 + (x_2+1)^2 \le 1$$

② $\lambda_2 \ge 0, \ \lambda_1 \ge 0$

③ complementary slackness, $\longleftarrow \lambda_i f_i(x) = 0 \quad i = 1 \text{ — } m$

$$\lambda_1\left((x_1-1)^2 + (x_2-1)^2 - 1\right) = 0 \qquad \lambda_2\left((x_1-1)^2 + (x_2+1)^2 - 1\right) = 0$$

④ گرادیان لاگرانژین برابر صفر باشد:

$$2x_1 + 2\lambda_1(x_1-1) + 2\lambda_2(x_1-1) = 0$$
$$2x_2 + 2\lambda_1(x_2-1) + 2\lambda_2(x_2+1) = 0$$

حالا در نقطه‌ی $x=(1,0)$ داریم که: قیود زیر باید برقرار باشند،

① برقرار است. primal constraints

② $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ (باید برقرار باشد.) dual const

③ برقرار است. slackness

④ $-2\lambda_1+2\lambda_2=0$, $2=0$ (باید برقرار باشد.) ✗ gradient

← شرایط KKT برقرار نیست.

درنتیجه λ_1^* و λ_2^* ای که ثابت کند x^* بهینه است نداریم.

c) حالا ابتدا داریم که،

$$g(\lambda_1,\lambda_2)=\inf L(x,\lambda_1,\lambda_2)=\inf_x \left\{ x_1^2+x_2^2+\lambda_1\left((x_1-1)^2+(x_2-1)^2-1\right)+\lambda_2\left((x_1-1)^2+(x_2+1)^2-1\right) \right\}$$

حالا نسبت به x گرادیان می‌گیریم، یکبار به x_1 و یکبار به x_2 مشتق جزئی می‌گیریم، خواهیم داشت:

$$\frac{\partial L}{\partial x_1}=2x_1+2\lambda_1(x_1-1)+2\lambda_2(x_1-1)=0 \longrightarrow x_1(1+\lambda_1+\lambda_2)-\lambda_1-\lambda_2=0$$

$$\frac{\partial L}{\partial x_2}=2x_2+2\lambda_1(x_2-1)+2\lambda_2(x_2+1)=0 \longrightarrow x_2(1+\lambda_1+\lambda_2)-\lambda_1+\lambda_2=0$$

$$\longrightarrow \boxed{x_1=\frac{\lambda_1+\lambda_2}{1+\lambda_1+\lambda_2}\ ,\ x_2=\frac{\lambda_1-\lambda_2}{1+\lambda_1+\lambda_2}} \rightsquigarrow \lambda_1+\lambda_2-\frac{(\lambda_1+\lambda_2)^2+(\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2}$$

پس با جایگذاری این مقادیر خواهیم داشت که: (ابتدا مخرج باید مثبت باشد، $1+\lambda_1+\lambda_2>0$)

$$g(\lambda_1,\lambda_2)=\begin{cases}\lambda_1+\lambda_2-\dfrac{(\lambda_1+\lambda_2)^2+(\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2} & \text{if } 1+\lambda_1+\lambda_2>0\\ -\infty & \text{o.w.}\end{cases}$$

$$\lambda_1+\lambda_2-\frac{(\lambda_1+\lambda_2)^2+(\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2}=\left(\lambda_1+\lambda_2-(\lambda_1-\lambda_2)^2\right)/1+\lambda_1+\lambda_2$$

از طرفی می‌خواهیم در مسئله‌ی dual این تابع را ماکزیمم کنیم و می‌دانیم که اگر جای λ_1 و λ_2 را عوض کنیم باز هم همین معادله بدست می‌آید یعنی معادله نسبت به λ_1 و λ_2 متقارن است پس با مشتق گیری هم مقادیر یکسانی برای λ_1 و λ_2 بدست می‌آید.

← در نقطه‌ی بهینه $\lambda_1=\lambda_2$، حالا خواهیم داشت، $\frac{2\lambda_1}{1+2\lambda_1}$، برای ماکزیمم کردن این تابع هم λ_1 را به بینهایت میل می‌دهیم، به 1 میل می‌کند.

$$\max \frac{\lambda_1+\lambda_2-(\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2} \longrightarrow \max \frac{2\lambda_1}{1+2\lambda_1}$$

$$\text{s.t. } \lambda_1,\lambda_2\geq 0 \qquad \text{s.t. } \lambda_1\geq 0$$

← d^* برابر خواهد بود 1. و p^* هم برابر بود با 1 ولی در d^* ما با 1 میل می‌کنیم 1 جزو نقاطی که واقعاً attained هستند نیست و می‌دانیم که داشتیم شرایط KKT هم زمانی برقرار بود که، strong duality برقرار باشد و نقاط بهینه‌ی primal، بهینه‌ی dual، attained باشند ولی اینجا شرط سومی برقرار نیست و درنتیجه KKT هم برقرار نیست.

Scanned with CamScanner

۵-

minimize $\quad tr(X) - \log\det(X)$

subject to $\quad Xs = y$

شرایط KKT عبارتند از:

① قیود برقرار باشند: $\quad X \succ 0,\ Xs = y$

② $Z \succeq 0$

③ $z_i(Xs - y)_i = 0$

④ گرادیان صفر باشد.

$$\Rightarrow \nabla_X \{tr(X) - \log\det(X) + z^T(Xs - y)\} = 0 \longrightarrow$$

$$I - X^{-1} + \frac{1}{2}(zs^T + sz^T) = 0 \longrightarrow \boxed{X^{-1} = I + \frac{1}{2}(zs^T + sz^T)}$$

از آنجایی که $X \succ$ برقرار است ← می‌توانیم بنویسیم: $\quad s = X^{-1}y \leftarrow$

$$s = \left(I + \frac{1}{2}(zs^T + sz^T)\right)y = y + \frac{1}{2} z \underbrace{s^Ty}_{1} + \frac{1}{2} s(z^Ty) = y + \frac{1}{2} z + \frac{1}{2}(z^Ty)s$$

$$\longrightarrow \frac{1}{2} z = \left(1 - \frac{1}{2} z^Ty\right)s - y \longrightarrow z = (2 - z^Ty)s - 2y$$

حالا دو طرف را در y ضرب داخلی می‌کنیم، خواهیم داشت:

$$z^Ty = 2\underbrace{s^Ty}_{1} - (z^Ty)\underbrace{s^Ty}_{1} - 2y^Ty \longrightarrow 2z^Ty = 2 - 2y^Ty \longrightarrow \boxed{z^Ty = 1 - y^Ty}$$

پس خواهیم داشت که:

$$\boxed{z = -2y + (1 + y^Ty)s}$$

پس داریم که:

$$X^{-1} = I + \frac{1}{2}\left(\left(-2y + (1+y^Ty)s\right)s^T + s\left(-2y + (1+y^Ty)s\right)^T\right)$$

$$\rightarrow X^{-1} = I + \frac{1}{2}\left(-2ys^T + (1+y^Ty)ss^T - 2sy^T + (1+y^Ty)\,ss^T\right)$$

$$= I + \frac{1}{2}\left(-2ys^T - 2sy^T + 2(1+y^Ty)ss^T\right)$$

$$= I + (1+y^Ty)ss^T - sy^T - ys^T$$

پس X ای که بهینه باشد به صورت یکتا بدست آمد. حالا فقط باید ببینیم X^* ای که سوال داده وارون این عبارت هست یا خیر:

$$X^* = I + yy^T - \frac{1}{s^Ts}ss^T \longrightarrow$$

$$X^*\left(I + (1+y^Ty)ss^T - sy^T - ys^T\right) =$$

$$I + yy^T - \cancel{\frac{1}{s^Ts}}ss^T + (1+y^Ty)\left(ss^T + \cancel{(y^Ts)}^{1}\, ys^T - ss^T\right) - sy^T - \underbrace{yy^Ts\,y^T}_{-yy^T} + sy^T$$

$$- ys^T - (y^Ty)\,ys^T + \frac{\cancel{ss^T}}{\cancel{s^Ts}} =$$

$$I + yy^T + ys^T + (y^Ty)ys^T - yy^T - ys^T - (y^Ty)ys^T = I$$

حالا X^* را از طرف دیگر در ماتریس حاصل ضرب می‌کنیم:

$$\left(I + (1+y^Ty)ss^T - sy^T - ys^T\right)X^* = I + yy^T - \frac{1}{s^Ts}ss^T + (1+y^Ty)\left(\cancel{ss^T} + ss^Tyy^T - \cancel{ss^T}\right)$$

$$- sy^T - (y^Ty)\,sy^T + \frac{1}{s^Ts}ss^T - \cancel{ys^T} - yy^T + \cancel{ys^T} =$$

$$I + \cancel{sy^T} + (y^Ty)sy^T - \cancel{sy^T} - (y^Ty)sy^T = I$$

← X^* وارون X^{-1} است یعنی X^* همان X بهینه (optimal) است.

$$\begin{aligned} \min \quad & x^T W x \\ \text{s.t.} \quad & x_i^2 = 1, \; i = 1 - n \end{aligned} \xrightarrow{\text{dual}} \begin{aligned} \max \quad & -1^T v \\ \text{s.t.} \quad & W + \mathrm{diag}(v) \succeq 0 \end{aligned}$$

a) ابتدا از راهنمایی سوال استفاده می‌کنیم و ابتدا آنرا اثبات می‌کنیم. $X \in S^n$ است، پس می‌توانیم آنرا به صورت $Q \Lambda Q^T$ بنویسیم، از طرفی چون رنک X برابر 1 است پس تنها یک مقدار ویژه‌ی ناصفر دارد پس خواهیم داشت که $X = Q\begin{bmatrix}\lambda & & \\ & 0 & \\ & & 0\end{bmatrix}Q^T \leftarrow X = \lambda q q^T$. حالا از آنجایی که $X \succeq 0 \leftarrow \lambda \geq 0$ درنتیجه می‌توانیم $\sqrt{\lambda} q$ را بردار x بگیریم، خواهیم داشت که $X = xx^T$.

حالا از آنجایی که عناصر روی قطر اصلی یک هستند پس x_i^2 ها برابر 1 هستند $\leftarrow$ x_i ها عضو $\{-1, +1\}$ هستند. حالا خواهیم داشت که:

$$tr(WX) = tr(Wxx^T) = tr(x^T W x) = x^T W x$$

همچنین از طرفی داریم که $X_{ii} = x_i^2 \leftarrow$ شرط $x_i^2 = 1$ معادل است با اینکه $X_{ii} = 1$.

از طرفی $tr(WX)$ هم برابر شد با $x^T W x$

پس مسئله‌ی اولیه‌مان را می‌توانیم به این فرم بنویسیم:

$$\begin{aligned} \min \quad & x^T W x \\ \text{s.t.} \quad & x_i^2 = 1 \quad i = 1 - n \end{aligned} \longleftrightarrow \begin{aligned} \min \quad & tr(WX) \\ \text{s.t.} \quad & X \succeq 0, \; \mathrm{rank}(X) = 1 \\ & X_{ii} = 1 \quad i = 1 - n \end{aligned}$$

b) از قسمت قبلی می‌دانیم که اگر یک قید رنک 1 بودن به این مسئله اضافه کنیم، معادل با مسئله‌ی اولیه می‌شود، در واقع ما با محدودتر کردن دامنه‌ی X، به جواب مسئله‌ی اولیه می‌رسیم، از طرفی داریم همان objective اولیه را مینیمایز می‌کنیم روی مجموعه‌ای بزرگتری از X ها $\leftarrow$ جواب بهینه کمتر مساوی است از جواب بهینه مسئله‌ی two-way partitioning

$$\begin{aligned} \min \quad & tr(WX) \\ \text{s.t.} \quad & X \succeq 0 \\ & X_{ii} = 1 \quad i = 1 - n \end{aligned}$$

اگر رنک ماتریس X، در این مسئله 1 شود، آنگاه جواب‌مان همان جواب مسئله‌ی two-way-partitioning است.

اگر مسئله dual مربوط به two-way-partitioning را در نظر بگیریم، خواهیم داشت که:

$$\begin{array}{lll} \max & -1^T\nu \\ \text{s.t.} & W+\mathrm{diag}(\nu)\succeq 0 \end{array} \quad \xleftrightarrow{\text{معادل اند}} \quad \begin{array}{lll} \min & 1^T\nu \\ \text{s.t.} & W+\mathrm{diag}(\nu)\succeq 0 \end{array}$$

حالا برای مسئله دوم dual می‌نویسیم، خواهیم داشت که:

$$L(\nu, X) = 1^T\nu - \mathrm{tr}(X(W+\mathrm{diag}(\nu))) = -\mathrm{tr}(XW) + \sum_{i=1}^{n}(1-x_{ii})\nu_i$$

← مشتق نسبت به ν بگیریم، قید اینکه $X_{ii}=1$ به ازای $i=1 \ldots n$ اضافه می‌شود.

$$\rightarrow g(X) = \begin{cases} -\mathrm{tr}(WX) & X_{ii}=1, X\succeq 0 \\ -\infty & \text{o.w.} \end{cases} \rightarrow \begin{array}{ll} \max & -\mathrm{tr}(WX) \\ \text{s.t.} & X\succeq 0 \\ & X_{ii}=1 \quad i=1 \ldots n \end{array}$$

که این مسئله هم معادل است با:

$$\begin{array}{ll} \min & \mathrm{tr}(WX) \\ \text{s.t.} & X\succeq 0 \\ & X_{ii}=1 \quad i=1 \ldots n \end{array}$$

* پس مسئله بخش a در واقع معادل dual مسئله two-way-partitioning است.

چون هر دو مسئله ها convex ای اند و شرایط KKT در آنها برقرار است پس

• جواب این دو یکسان است و کران پایینی برای جواب مسئله two-way-partitioning است.

-۷

P) $\min\ c^T x$ s.t. $F(x) \preceq 0$

D) $\max\ tr(F_0 Z)$ s.t. $tr(F_i Z) + c_i = 0 \quad i=1 - m$, $Z \succeq 0$

شرایط KKT برقرار است (مسئله‌مان SDP, convex است و شرایط $Z \succeq 0$, $tr(F_i Z) + c_i = 0$, $F(x) \preceq 0$ نیز در حال داریم.)

حالا از strong duality داریم که، $c^T x^* = tr(F_0 Z^*)$

مسئله‌ی بعد قیمان عبارت است از:

$$\text{minimize} \quad c^T x + M \max\{0, \lambda_{max}(F(x))\} \quad , M > tr(Z^*)$$

این مسئله‌ای convex است، آنرا به صورت زیر بازنویسی می‌کنیم:

$$\min\ c^T x + t \quad \text{s.t.} \quad \max\{0, \lambda_{max}(F(x))\} \le t \longrightarrow \min\ c^T x + t \quad \text{s.t.} \quad \underbrace{M\lambda_{max}(F(x)) \le t}_{\boxed{MF(x) \preceq tI}}, \quad t \ge 0$$

از آنجایی که $Z \succeq 0$, $M > tr(Z)$ ← $M > 0$ ← $F(x) \preceq \frac{t}{M} I$

حالا مسئله‌مان convex است و شرایط KKT نیز برقرار است: (ماتریس Z را در نظر می‌گیریم $Z \succeq 0$)

$$\boxed{\begin{array}{l} M\lambda_{max}(F(x)) - t \le 0 \\ Z \succeq 0 \\ tr(Z(MF(x) - tI)) = 0 \\ \text{مکمل صفر} \end{array}} \longrightarrow L(x, t, \nu, Z) = c^T x + t + tr(Z(F(x) - \tfrac{t}{M} I)) - \nu t$$

حالا مشتق می‌گیریم و برابر صفر قرار می‌دهیم:

$$\boxed{\frac{\partial L}{\partial x_i} = c_i + tr(F_i Z) = 0 \quad i = 1 - m}$$

$$\frac{\partial L}{\partial t} = 1 - \frac{1}{M} tr(Z) - \nu = 0 \rightarrow \boxed{\nu = 1 - \frac{1}{M} tr(Z)}$$

$$1 - \frac{1}{M} tr(Z) \ge 0 \leftarrow \nu \ge 0$$

درنتیجه خواهیم داشت:

$$t + \underbrace{c^T x + tr(Z(F(x) - F_0))}_{0} + tr(F_0 Z) \underbrace{- t\ tr(Z)} - t + t\ tr(Z) =$$

$\longrightarrow \; tr(F_0 Z)$

$$\Rightarrow \; g(Z,\nu) = \begin{cases} tr(F_0 Z) & tr(F_i Z) + c_i = 0 \quad i = 1 - m \;, \quad Z \succeq 0 \\ & \underbrace{1 - \frac{1}{M} tr(Z) \geq 0}_{tr(Z) \leq M} \\ -\infty & \text{o.w.} \end{cases}$$

$$\rightarrow \quad \max \quad tr(F_0 Z)$$

$$\text{s.t.} \quad Z \succeq 0$$

$$tr(F_i Z) + c_i = 0 \quad i = 1 - m$$

$$tr(Z) \leq M$$

حالا هر جواب Z^* برای مسئله D ، یک جواب برای این مسئله dual نیز هست.

چرا که در تمامی قیدها یکسان اند و از طرفی داشتیم که $M > tr(Z^*)$ که در این رابطه هم صدق می‌کند.

← Z^* در این هم feasible است، همچنین چون که $tr(F_0 Z^*) = c^T x^*$

است و به ازای هر x ای غیر از x^* (و کلا عضو این مجموعه بهینه)، خواهیم داشت $c^T x \geq c^T x^*$

پس ترم $M \max(0, \lambda_{max}(F(x)))$ در مسئله unconstraint مان باید ۰ شود

چرا که جواب‌های بهینه dual مسئله P ، جواب‌های بهینه dual مسئله بدون قیدمان نیز هستند.

همچنین d^* مسئله بدون قیدمان برابر است با p^* آن ← ترم دوم یعنی $M \max(0, \lambda_{max}(F(x)))$ باید صفر شود ← $\lambda_{max}(F(x)) \leq 0$ ← $F(x) \preceq 0$

باید برقرار باشد ← هر جواب بهینه x^* در مسئله بدون قید، یک جواب بهینه برای مسئله P مان است.

$$\min \quad \alpha^T y - 2|\gamma|^T \sqrt{y}$$
$$\text{s.t.} \quad \beta^T y = 0$$
$$y \geq 0$$

$$\xrightarrow{\text{dual}} L(y, \lambda, \nu) = \alpha^T y - 2|\gamma|^T \sqrt{y} - \lambda(\beta^T y) - \nu^T y$$

④ حالا باید نسبت به y مشتق بگیریم و برابر صفر قرار دهیم. ← grad

③ به ازای تمام iها باید $\nu_i y_i = 0$ ← slack

② $\nu \geq 0$ ← dual const

① $y \geq 0$, $\beta^T y = 0$ ← Primal const

اگر مشتق بگیریم، خواهیم داشت که،

$$\frac{\partial L}{\partial y_i} = \alpha_i - \frac{|\gamma_i|}{\sqrt{y_i}} - \lambda \beta_i - \nu_i = 0$$

اگر این رابطه بخواهد برقرار باشد باید به ازای هر i ای داشته باشیم $0 < y_i$ ← از شرط ③ $\nu_i y_i = 0$

داریم که به ازای هر iای، ν_i برابر است با صفر.

درنتیجه خواهیم داشت که:

$$\alpha_i - \frac{|\gamma_i|}{\sqrt{y_i}} - \lambda \beta_i = 0 \rightarrow \alpha_i - \lambda \beta_i = \frac{|\gamma_i|}{\sqrt{y_i}}$$

$$\rightarrow \sqrt{y_i} = \frac{|\gamma_i|}{\alpha_i - \lambda \beta_i} > 0 \rightarrow \underbrace{\alpha_i - \lambda \beta_i > 0}$$

حالا داریم که،

$$\inf L = \sum_{i=1}^{m} y_i \Big(\underbrace{\alpha_i - \frac{|\gamma_i|}{\sqrt{y_i}} - \lambda \beta_i}_{0} - \cancel{\nu_i} \Big) - \sum_{i=1}^{m} \sqrt{y_i}\, |\gamma_i| = - \sum_{i=1}^{m} \underbrace{\sqrt{y_i}}_{|\gamma_i|/\alpha_i - \lambda\beta_i}\, |\gamma_i|$$

$$\rightarrow g(\lambda) = \begin{cases} -\sum_{i=1}^{m} \frac{\gamma_i^2}{\alpha_i - \lambda \beta_i} & \text{if constraints hold} \\ -\infty & \text{o.w.} \end{cases}$$

حالا برای محاسبه λ^* و p و g داریم که :

$$\frac{\partial g}{\partial \lambda} = 0 \longrightarrow -\sum_{i=1}^{m} \frac{-\beta_i \gamma_i^2}{(\alpha_i - \lambda \beta_i)^2} = 0 \longrightarrow \sum_{i=1}^{m} \beta_i \gamma_i^2 \left(\frac{1}{\alpha_i - \lambda \beta_i}\right)^2 = 0$$

حالا کل عبارت را اگر در $\prod_{i=1}^{m} (\alpha_i - \lambda \beta_i)^2$ ضرب کنیم، بازهم ریشه ها ثابت می مانند، چراکه می دانیم تمامی $(\alpha_i - \lambda \beta_i)^2$ ها مثبت اند و درنتیجه انگاری داریم مخرج مشترک می گیریم.

ممکن است ریشه‌های جدیدی اضافه شوند ولی قبلی‌ها باقی می‌مانند

$$\xrightarrow{\prod_{i=1}^{m} (\alpha_i - \lambda \beta_i)^2} \sum_{i=1}^{m} \beta_i \gamma_i^2 \left(\frac{1}{\alpha_i - \lambda \beta_i}\right)^2 \prod_{j=1}^{m} (\alpha_j - \lambda \beta_j)^2 = \underbrace{\sum_{i=1}^{m} \beta_i \gamma_i^2 \prod_{i \neq j}^{m} (\alpha_j - \lambda \beta_j)^2}$$

حالا این چند جمله‌ای ممکن است علاوه بر ریشه‌های قبلی، ریشه‌های دیگری هم داشته باشد ←

λ^* ای که در $g(\lambda)$ استفاده شود، از طریق محاسبه ریشه‌های تعدادی چند جمله‌ای به فرم $P(\lambda)$ حاصل می شود.

$$\longrightarrow \sum_{i \in C} \beta_i \gamma_i^2 \prod_{\substack{i \neq j \\ j \in C}}^{m} (\alpha_j - \lambda \beta_j)^2$$

C هر یک مجموعه‌ای مشخص است ($\subseteq \{1, 2, \dots, m\}$).