810100441 HW7 حمیدرضا کاشانی

۱- تابع زیان ما به‌صورت زیر است:

(a

$$f(x) = \sum_{i=1}^{n} \psi(x_i - y_i) + \lambda \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

$$\nabla f(x) = \begin{bmatrix} \psi'(x_1 - y_1) - 2\lambda(x_2 - x_1) \\ \psi'(x_2 - y_2) + 2\lambda(x_2 - x_1) - 2\lambda(x_3 - x_2) \\ \vdots \\ \psi'(x_n - y_n) + 2\lambda(x_n - x_{n-1}) \end{bmatrix}$$

$$H(x) = \begin{bmatrix} \psi''(x_1 - y_1) - 2\lambda , & -2\lambda , & 0 , & 0 & \cdots & 0 & 0 \\ -2\lambda , & \psi''(x_2 - y_2) + 4\lambda , & -2\lambda , & 0 , & 0 , & 0 & \cdots \\ 0 , & 0 , & -2\lambda , & \psi''(x_3 - y_3) + 4 , & -2\lambda , & 0 , 0 , 0 & \\ & & \ddots & & & & \\ & 0 & & & & & \\ & & & & & -2\lambda , & \psi''(x_n - y_n) + 2\lambda \end{bmatrix}$$

$$\begin{bmatrix} \diagdown\diagdown\diagdown & 0 \\ 0 & \end{bmatrix}$$

(b برای این قسمت تابع به‌صورت زیر است است که دیتا دوبعدی است یعنی یک تصویر است:

$$f(x) = \sum_{i,j=1}^{n} \psi(x_{ij} - y_{ij}) + \lambda\left(\sum_{i=1}^{n-1}\sum_{j=1}^{n} (x_{i+1,j} - x_{ij})^2 + \sum_{i=1}^{n}\sum_{j=1}^{n-1} (x_{i,j+1} - x_{ij})^2 \right)$$

$$\frac{\partial f(x)}{\partial x_{ij}} = \psi'(x_{ij} - y_{ij}) + 2\lambda\big((x_{ij} - x_{i-1,j}) - (x_{i+1,j} - x_{ij}) + (x_{ij} - x_{i,j-1}) - (x_{i,j+1} - x_{ij}) \big)$$

$$= \psi'(x_{ij} - y_{ij}) + 2\lambda(4x_{ij} - x_{i-1,j} - x_{i+1,j} - x_{i,j-1} - x_{i,j+1})$$

و در ادامه داریم برای Hession به صورت زیر:

$$\frac{\partial^2 f}{\partial x_{ij}\,\partial x_{nm}} = \begin{cases} n=i,\ j=m & \Longrightarrow \phi''(x_{ij}-y_{ij})+4\lambda \\ n=i-1,\ j=m & \\ n=i+1,\ j=m & \Longrightarrow -2\lambda \\ n=i,\ j-1=m & \\ n=i,\ j+1=m & \\ \text{o.w} & \Longrightarrow 0 \end{cases}$$

با هم مثل قسمت یک مسئله است که ما داده ها را از صورت دو بعدی به صورت وکتور در آورده ایم و دقیقا همان کار قسمت اول را برای این داده ها می کنیم.

2- ابتدا باید Hess تابع $f(x)$ را محاسبه کنیم:

$$f(x) = \frac{1}{2}x^T x + \log \sum_{i=1}^{m} \exp(a_i^T x + b_i)$$

$$\longrightarrow \nabla f_x = x + \frac{\sum_{i=1}^{m} a_i \exp(a_i^T x + b_i)}{\sum_{i=1}^{m} \exp(a_i^T x + b)}$$

$$\longrightarrow \nabla^2 f_x = I + \frac{\sum_{i=1}^{m} a_i a_i^T \exp(a_i^T x + b_i) \sum \exp(a_i^T x + b) - \sum_{i=1}^{m} a_i^T \exp(a_i^T x + b) \sum_{i=1}^{m} a_i \exp(a_i^T x + b_i)}{\left(\sum_{i=1}^{m} \exp(a_i^T x + b)\right)^2}$$

حال با فرض $z = \dfrac{\exp(a_i^T x + b_i)}{\sum_{i=1}^{m} \exp(a_i^T x + b_i)}$ داریم و با ساده سازی داریم:

$$H(x) = \nabla^2 f_x = I + A^T(\underbrace{\text{diag}(z)}_{\text{diagonal}} - \underbrace{zz^T}_{\text{Low rank}})A$$

طبق اسلاید ۳۰-۱۰ می توان معادله نیوتن را برای آن به صورت زیر نوشت: $H\,\nabla x = -\nabla f$

و چون گفتیم H ساختار مشخص دارد می توان به فرم زیر نوشت:

$\text{diag}(z) - zz^T = L\,\text{diag}(z)^{-1} L^T$ و $L = \text{diag}(z) - zz^T$

با این تجزیه نیوتن راحت است داریم:

$$H\,\nabla x = -\nabla f \Rightarrow \left(I + A^T(\text{diag}(z) - zz^T)A\right)\Delta x = -\nabla f$$

که به صورت حل دستگاه زیر حل می‌شود:

$$\begin{bmatrix} I & A^T L \\ L^T A & -\mathrm{diag}(z)\end{bmatrix}\begin{bmatrix}\Delta x \\ \Delta x\end{bmatrix} = \begin{bmatrix}-\nabla f \\ 0\end{bmatrix} \Rightarrow \left(\mathrm{diag}(z)+L^T A A^T L\right)\Delta x = L^T A(-\nabla f)$$

که این رابطه دارای $m+1$ متغیر و از مرتبه $\frac{m^3}{3}+m^2 n$ می‌باشد.

---

3 a) مسئله یک مسئله max Likelihood است که قرار است احتمال رخداد را ماکسیمایز کنیم:

$$\text{maximize } L(x) = \log\left(\prod_{y_i=1} P_i(x) \prod_{y_i=-1}\left(1-P_i(x)\right)\right) = \sum_{y_i=1}\log P_i(x) + \sum_{y_i=-1}\log\left(1-P_i(x)\right)$$

$$P_i(x) = \Pr(a_i^T x + v_i \ge b_i) \rightarrow \text{نرمال} \Rightarrow \Phi(x \le X)$$

مسئله به این صورت می‌شود:

$$\text{minimize } -\sum_{y_i=1}^{m} \log\Phi(-b_i + a_i^T x) - \sum_{y_i=-1}^{m} \log\Phi(b_i - a_i^T x)$$

که Φ تابع توزیع تجمعی نرمال است $\Leftarrow$ $\Phi(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-\frac{t^2}{2}}dt$

Φ چون تابع توزیع تجمعی نرمال است و نرمال log-concave است یک تابع log-concave است که با یک تابع affine ترکیب شده پس $\Phi(a_i^T x - b)$ هم log-concave خواهد شد. پس مسئله یک بهینه‌سازی log-concave است.

b) برای تبدیل کردن این دو Σ به یک مجموع ابتکاری را استفاده می‌کنیم. جمع $y=-1$ را که a_i ها در آن‌ها قرینه شده و تمام داده‌ها را می‌توان به فرم $Ax-b$ نوشت به طوری که $A \leftarrow \mathrm{diag}(y)A$ و $b \leftarrow \mathrm{diag}(y)b$. در این صورت a_i ها و b هایی که برای $y=-1$ هستند قرینه می‌شوند و $y=1$ ها تغییری نمی‌کنند پس مسئله را این فرم می‌توان نوشت:

$$\text{minimize } h(Ax-b) \rightarrow h(w) = -\sum_{i=1}^{m}\log\Phi(w_i) \quad \Rightarrow \mathbb{R}^m \rightarrow \mathbb{R}$$

حال گرادیان و هسین تابع $f(x)$ را حساب می‌کنیم:

$$\nabla f(x) = A^T \nabla h(Ax-b) \quad \text{و} \quad \nabla^2 f(x) = A^T \nabla^2 h(Ax-b) A$$

گرادیان اول h به صورت روبرو است:

$$\frac{\partial h(w)}{\partial w_i} = -\frac{\phi'(w_i)}{\Phi(w_i)} = \frac{-1/\sqrt{2\pi}}{\exp(w_i^2/2)\,\Phi(w_i)}$$

حال هسین به صورت زیر است:

$$\frac{\partial^2 h(w)}{\partial w_i^2} = -\frac{\phi''(w_i)}{\Phi(w_i)} + \frac{\phi'(w_i)^2}{\Phi(w_i)^2} = \frac{w_i/\sqrt{2\pi}}{\exp(w_i^2/2)\,\Phi(w_i)} + \left(\frac{1/\sqrt{2\pi}}{\exp(w_i^2/2)\,\Phi(w_i)}\right)^2$$

و بقیه به صورت پایین صفر شده است.

۴ـ الف) گزینه غلط است، در الگوریتم های مبتنی بر decent کاهشی‌ترین لزوما جهت حرکت است و جهت صحیح حرکت است که همیشه به نقطه بهینه را سریع می‌کند.

ب) به صورت کلی می‌تواند صحیح نباشد در الگوریتم های مبتنی بر decent متناسب با شکل تابع هزینه حداقل انتخاب یک line-search درست می‌تواند باعث افزایش سرعت همگرایی شود.

ج) این گزینه غلط است زیرا همواره از متد decent در نقاط دور از نقاط min برای کارایی بهتر استفاده می‌شود و از متد نیوتن در نقاط نزدیک به بهینه زیرا گرادیان نزدیک نقاط بهینه بسیار کوچک می‌شود پس با کوچک شدن گام ها دیر به جواب بهینه می‌رسیم.

د) این گزینه درست است زیرا روش نیوتن یک تقریب درجه دوم در هر نقطه می‌زند و نسبت به گرادیان در نزدیکی نقاط بهینه بسیار خوب عمل می‌کند ـ اما این برای زمانی است که ماتریس هم وارون‌پذیر و دیگر مشتق دوم برابر صفر نشود در غیر این صورت غلط است.

ه) روش نیوتن برای $h=1$ ممکن است جواب ندهد اما در حالت $h<1$ نرخ همگرایی افزایش پیدا می‌کند. غلط است

و) این گزینه صحیح است زیرا آدام نوعی پیشرفته گرادیان عمل می‌کند و با انتخاب درست آن سرعت همگرایی افزایش می‌یابد.

ز) علامت است زیرا روش نیوتن مستقل از تغییر مبدا مختصات است و انجام دادن آن کار تاثیری در سرعت همگرایی ندارد و شرط affine بوده و نیوتن از تخمین درجه ۲ می زند

---

6) در این مسئله هدف این است که در کمترین دور ممکن بدهی‌ها ۹۵ درصد تسویه شوند. برای رسیدن به این هدف باید در هر دور هر بانک بیشترین مقداری را که می‌تواند بپردازد، بپردازد (با فرض دادن پول و تمکین برای صاف کردن بدهی) پرداخت کند که برای این کار باید کمترین میزان پرداخت کل بانک‌ها در هر دور بیشینه گردد با این قید که پرداخت کل هر بانک از مقدار دارایی‌اش در هر زمان بیشتر نشود ($Px1 <= Ct$) و هیچ فردی داده نشود بیش در هر دوره بیشتر از بدهی آن دوره پرداخت نشود ($Lt - P >= 0$).

حال $L1$ و $C1$ به ما داده شده است. در هر دوره ابتدا مسئله convex گفته شده را حل می‌کنیم سپس مقادیر Lt و Ct را آپدیت کرده و در صورت ارضا نشدن stop condition با افزایش $t=t+1$ به دور بعدی حرکت می‌کنیم.

اثبات کانوکس بودن مسئله: obj → cp.maximize(cp.min($\underbrace{\text{P1 x np.ones(n)}}_{\text{affine}}$))

بیشینه کردن یک تابع مقعر ← concave

شرط‌ها هم به صورت $Px1 \leq Ct$ و $Lt-P>=0$ هم دو شرط affine هستند.

که به صورت پایتون آورده شده است.

با توجه به نمودار رسم شده برای I مقدار I از ۲۴ به بعد از ۳ کمتر می‌شود یعنی $T=24$ تمام بدهی‌ها صاف شده است ولی برای $T=47$ از یک صدم کمتر می‌شود که دیگر کلا بدهی‌ها صاف شده‌اند

a) مسئله حل مدار با استفاده از بهینه سازی کانوکس است ما در این مسئله KCL و KVL و قوانین فیزیکی موجود در عناصر را برای مدل‌ها به صورت قید داریم. تابع داده شده در Hint جمع توان های عناصر کل مدار است که برابر $\int_0^{i_j} \phi_j(u_j)\,du_j$ است تابع $\phi_j(u_j)$ ها غیرنزولی است و انتگرال توابع غیرنزولی تابع کانوکس است و ϕ جمع ϕ_j ها است که تابعی کانوکس است و برابر توان کل جمع عناصر است حال در صورت داشتن قید KVL یا KCL قانون پایستگی انرژی برقرار می‌شود پس نقطه‌ای می‌یابیم ما در $\phi = 0$ برقرار می‌شود که مسئله به این صورت نهایی می‌شود:

$$\text{minimum } \phi(i) \rightarrow \text{توان کل}$$

$$\text{s.t } Ai = 0 \rightarrow KCL$$

که این خود در مسئله بهینه سازی تمامی اطلاعات عناصر را در خود دارد که در کنار قیود KCL به ما تمامی V ها را می‌دهد

تابع دوال مشتق لاگرانژین برابر صفر $\longrightarrow \nabla\phi(i) + A^T\lambda = 0$

اگر ما $\lambda = \nabla\phi(i)$ بگیریم و $e = -\lambda$ درنظر داشته باشیم داریم:

$$\Rightarrow \nabla\phi(i) = -A^T\lambda = A^Te \Rightarrow \lambda = A^Te \underset{\text{V ولتاژها است}}{\xrightarrow{\lambda \text{ همان}}} \boxed{v = A^Te}$$

$v_j = \phi_j(i_j)$ و $Ai = 0$ تمام قیود با این مسئله برقرار می‌شود

---

b) در مدار داده شده دو مقاومت دو دیود و یک منبع ولتاژ داریم که برای تمامی آنها تابع ϕ_j را محاسبه کرده و از آن $\phi(i)$ را بدست می‌آوریم:

برای مقاومت‌ها $\leftarrow \int_0^{i} ru\,du = \frac{1}{2} r i^2$

برای منبع ولتاژ $\leftarrow \int_0^{i} v + ru\,du = vi + \frac{1}{2} r i^2$

برای دیود $\leftarrow \int_0^{i} V_T \log(1 + u/I_s)\,du = V_T I_s\left((1 + \frac{i}{I_s})\log(1 + \frac{i}{I_s}) - i/I_s\right)$

مسئله:

$$\text{minimize } v_1 i_1 + \frac{1}{2}\sum_{j=1,2,5} R_j i_j^2 + \sum_{j\in 3,4} V_T I_s\left((1 + \frac{i_j}{I_s})\log(1 + \frac{i_j}{I_s}) - \frac{i_j}{I_s}\right)$$

$$\text{s.t } Ai = 0$$

در فایل پیوست کل مسئله حل و e و v و i و A^Te بدست آمده و دقت مسئله گفته شده است.

8- ابتدا شروط لازم یعنی $Y \geq 0$ و $X[0] = Y[0] = 0$ و اینکه Y در بازه‌ی غیرمجاز نباید تغییر کند را در قید مسئله می‌نویسیم، سپس شرط صعودی بودن X یعنی حرکت به سمت جلو در جاده برای همه t ها و کم بودن سرعت در کل بازه از S_max ~~[illegible]~~ را اعمال می‌کنیم حال شرط اینکه سرعت از یک مقداری کم نشود یک شرط غیرکانوکس است ولی چون این شرط فقط در زمانی صدق می‌کند که ما حرکت در راستای x داریم پس به جای استفاده از ΔP از Δx برای نمایش سرعت استفاده می‌کنیم که $\Delta x \geq S_{min}$ دیگر غیرکانوکس نیست، حال objective را کمینه کردن مقدار کل شتاب ماشین در نظر می‌گیریم:

$$\text{minimize} \sum_{t=1}^{T-1} \|(P_{t+1} - P_t) - (P_t - P_{t-1})\|_2$$

$$\text{s.t} \quad 0 \leq Y \text{ و } Y \leq L \text{ و } X[0] = 0 \text{ و } Y[0] = 0 \text{ , } Y[0:T\_start] = 0$$

$$\text{ و } Y[T\_end:] = 1 \text{ و } \|P_{t+1} - P_t\|_2 \leq S\_max \text{ و } \|X_{t+1} - X_t\|_1 \geq S_{min}$$

t های غیر از زمان سبقت

$$\text{ و } X_{t+1} - X_t \geq 0$$

با رسم P_t برای همه t ها می‌بینیم در ترین حالت سبقت برای زمان T-end تا T-start به صورت خطی است و با کمترین میزان تغییرات شتاب لاین عوض می‌شود

8- ابتدا شروط لازم یعنی $L \geq Y \geq 0$ و $X[0]=Y[0]=0$ و اینکه Y در بازه‌ی عبور عابر باید تغییر نکند را در قیود مسئله می‌نویسیم، سپس شرط صعودی بودن X یعنی حرکت به سمت جلو در جاده برای تمام t ها و کم بودن سرعت در کل بازه از s_max ~~...~~ را اعمال می‌کنیم حال شرط اینکه سرعت از یک مقداری کم نشود یک شرط غیر کانوکس است ولی چون این شرط فقط در زمانی صدق می‌کند که ما حرکت در راستای x داریم پس به جای استفاده از Δp از Δx برای نمایش سرعت استفاده می‌کنیم که $\Delta x \geq s_{min}$ دیگر غیر کانوکس نیست، حال objective را کمینه کردن مقدار کل شتاب ماشین در نظر می‌گیریم:

$$\text{minimize} \sum_{t=1}^{T-1} \|(p_{t+1}-p_t)-(p_t-p_{t-1})\|_2$$

$$\text{s.t.} \quad 0 \leq y \text{ و } y \leq L \text{ و } x[0]=0 \text{ و } y[0]=0 \text{ ، } y[0:T\text{-}start]=0$$

$$\text{و } y[T\text{-}end:]=1 \text{ و } \|p_{t+1}-p_t\|_2 \leq s\_max \text{ و } \|x_{t+1}-x_t\|_1 \geq s_{min}$$

t های غیر از زمان سبقت

$$\text{و } x_{t+1}-x_t \geq 0$$

با رسم p_t برای تمام t ها می‌بینیم نرم‌ترین حالت سبقت برای زمان T-start تا T-end به صورت خطی است و با کمترین میزان تغییرات شتاب لاین عوض می‌شود