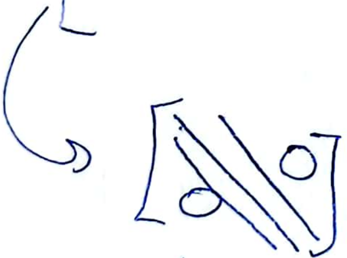
810100441 HW7 حمید رضا کاشانی

۱- (a تابع زیان ما به صورت زیر است:

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

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

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



(b برای این قسمت تابع به صورت زیر است که دیتا دو بعدی 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:

$$\frac{\partial^2 f}{\partial x_{ij}\,\partial x_{nm}} = \begin{cases} n=i,\ j=m & \Longrightarrow \phi''(x_{ij}-y_{ij}) + ۴\lambda \\ n=i-1 \text{ و } j=m & \\ n=i+1 \text{ و } j=m & \Longrightarrow -۲\lambda \\ n=i \text{ و } j-1=m & \\ n=i \text{ و } j+1=m & \\ o.w & \Longrightarrow 0 \end{cases}$$

با هم مثل قسمت یک سوال است که ما داده ها را از صورت دوبعدی به صورت وکتور در آوردیم و دقیقا همان کار قسمت اول را برای این داده ها می کنیم.

---

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

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

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

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

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

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

طبق اسلاید ۲۰-۱۰ 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:

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

با این تجزیه نیوتن را بسط داریم:

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

که به صورت حل دستگاه زیر می‌شود:

$$\begin{bmatrix} I & A^TL \\ L^TA & -diag(z) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta x \end{bmatrix} = \begin{bmatrix} -\nabla f \\ 0 \end{bmatrix} \Rightarrow (diag(z) + L^TAA^TL)\Delta x$$

$$= L^TA(-\nabla f)$$

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

---

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

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

$$P_i(x) = Pr(a_i^Tx + v_i \geq b_i) \rightarrow \text{نرمال} \Rightarrow \Phi(x \leq X)$$

مسئله به این صورت می‌شود:

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

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 $\Phi(a_i^Tx-b)$ هم log-concave خواهد بود. پس مسئله یک مسئله محدب log-concave است.

b) برای تبدیل کردن این دو $\sum$ به یک مجموع ابتکار زیر را استفاده می‌کنیم. جمع $y=-1$ را اگر همان u در آن‌ها قرینه شده و تمام داده‌ها را می‌توان به فرم $Ax-b$ نوشت به طوری که $A \leftarrow diag(y)A$ و $b \leftarrow diag(y)b$ 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:

$$\text{minimize } h(Ax-b) \rightarrow h(u) = -\sum_{i=1}^{m} \log\Phi(u_i): \Rightarrow R^m \rightarrow R$$

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:

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

ماتریس 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.

۴_ الف) گزینه غلط است، در الگوریتم‌های مبتنی بر decent کلیدی ترین پارامتر جهت حرکت است و جهت صحیح حرکت است که رسیدن به نقطه بهینه را تسریع می‌کند.

ب) ~~[illegible]~~ به صورت کلی می‌تواند صحیح نباشد در الگوریتم‌های 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 نقطه 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 گام‌ها دیر به جواب بهینه می‌رسیم.

د) این گزینه درست است زیرا روش نیوتن یک تقریب درجه دوم در هر نقطه می‌زند و نسبت به گرادیان‌ها در نزدیکی نقطه بهینه بسیار خوب عمل می‌کند. اما این برای زمانی است که ما ماتریس هسیان داریم و مشتق دوم پذیر باشد و صدق کند 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 نزدیک پیشین گرادیان عمل می‌کند و با انتخاب درست آن سرعت همگرایی افزایش می‌یابد.

c) غلط است زیرا روش نیوتن مستقل از تغییر مبدا مختصات است و انجام دادن این کار تاثیری در سرعت همگرایی ندارد و شرط affine بوده و نیوتن از تخمین درجه دو استفاده می‌کند.

۵-a) مسئله اصلی به صورت زیر است

$$\text{maximize} \quad \min_i \left(\alpha \log\left(1+\frac{G_{ii}P_i}{\sigma^2+\sum_{j\neq i} G_{ij}P_j}\right) - \lambda \sum_j P_j \right)$$

$$\text{s.t} \quad P_{min} \le P_j \le P_{max}$$

ضریب اهمیت برای محاسبه‌ی curve

این مسئله کانوکس نیست

مسئله را با آوردن قسمت غیر کانوکس در داخل قید و حذف log برای کانوکس کردن قید به مسئله کانوکس تبدیل می‌شود زیرا به دست آوردن سیگنال به نویز بهینه نزدیک به ما R_i را می‌دهد.

$$\frac{G_{ii}P_i}{\sigma^2+\sum_{j\neq i} G_{ij}P_j} \ge \lambda \Rightarrow$$

ضریب اهمیت را در داخل قید می‌آوریم

$$\boxed{\begin{aligned} &\text{minimize} \quad \sum_j P_j \\ &\text{s.t} \quad P_{min} \le P_j \le P_{max} \ \text{for } j=1,\dots,n, \quad -G_{ii}P_i + \lambda\left(\sigma^2 + \sum_{j\neq i} G_{ij}P_j\right) \le 0 \quad i=1,\dots,n \end{aligned}}$$

λ یک عدد است و تمام قیود کانوکس می‌باشند.

b) کد پایتون و نمودار ضمیمه شده است.

۶-a) در این مسئله برای هر T یک مسئله کانوکس حل می‌کنیم و کوچکترین T که مسئله فیزیبل باشد جواب مسئله ما می‌شود

تمام شروط را در قالب شروط کانوکس در قید مسئله می‌آوریم و تابع هدف را صفر می‌کنیم تا یک مسئله فیزیبیلیتی حل کنیم

مسئله به صورت روبه‌رو می‌شود: T ثابت است برای هر مسئله:

$$\text{minimize} \quad 0$$

$$\begin{aligned} \text{s.t} \quad & L_{t+1} = L_t - P_t, \quad t=1,2,\dots,T-1 \\ & C_{t+1} = C_t - P_t\mathbf{1} + P_t^T\mathbf{1}, \quad t=1,2,\dots,T-1 \\ & P_t\mathbf{1} \preceq C_t, \quad t=1,2,\dots,T \\ & L_T = 0 \\ & L_t \succeq 0,\ C_t \succeq 0,\ t=1,2,\dots,T \quad \text{و} \quad P_t \succeq 0,\ t=1,2,\dots,T \end{aligned}$$

b) کد پایتون به فایل ضمیمه شده است

a) مسئله حل مدار با استفاده از بهینه سازی کانوکس است. ما در این مسئله KCL و KVL و قوانین عنصری موجود در عناصر را برای ولتاژ 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 غیر نزولی است و انتگرال توابع غیر نزولی تابع کانوکس است و φ جمع ϕ_j هاست که تابعی کانوکس است و برابر توان جمع عناصر است. حال در صورت داشتن قید KVL یا KCL قانون پایستگی انرژی برقرار شود. پس نقطه مینیمم ما در $\varphi=0$ برقرار می شود که مسئله به این صورت نهایی می شود:

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

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

که این نوع در مسئله بهینه سازی تمام اطلاعات عناصر را در خود دارد که در کنار قیود KCL برای ما تمام ولتاژ ها را می دهد

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

اگر ما $\lambda=\nabla\varphi(i)$ بگیریم و $e=-\lambda$ در نظر داشته باشیم داریم:

$$\Rightarrow \nabla\varphi(i) = -A^T\lambda = A^T e \Rightarrow \lambda = A^T e \xRightarrow[\text{ها است } v]{\text{همان } \lambda} \boxed{v = A^T e}$$

تمام قیود با به این مسئله برقرار می شود

$$v_j = \phi_j(i_j) \quad \text{و} \quad Ai=0$$

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 $\varphi(i)$ را بدست می آوریم:

برای مقاومت ها $\leftarrow \int_0^{i} r u\,du = \frac{1}{2} r i^2$

برای منبع ولتاژ $\rightarrow \int_0^{i} v + r u\,du = v i + \frac{1}{2} r i^2$

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

مسئله:

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

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

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

8- ابتدا شروط لازم یعنی $0 \le Y$ و $X[0]=Y[0]=0$ و اینکه Y در بازه‌ی عبور هیچ‌جا نباید تغییر کند را در قید مسئله می‌نویسیم، سپس شرط صعودی بودن X نیز چون که به سمت جلو در جاده برای تمام t ها و کم بودن سرعت در کل بازه از S_max را اضافه می‌کنیم حال شرط این که 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 $\Delta x < S_{min}$ دیگر غیرکانوکس نیست، حال objective را کمینه کردن مقدار کل شتاب ماشین در نظر می‌گیریم:

$$\text{minimize} \quad \sum_{t=1}^{T-1} \|(p_{t+1}-p_t)-(p_t-p_{t-1})\|_2^2$$

$$\text{s.t.} \quad 0 \le Y \ \text{و}\ Y \le L \ \text{و}\ X[0]=0 \ \text{و}\ Y[0]=0 \ \text{و}\ Y[0:T\text{-start}]=0$$

$$\text{و}\ Y[T\text{-end}:]=1 \ \text{و}\ \|p_{t+1}-p_t\|_2 \le S\_max \ \text{و}\ \|X_{t+1}-X_t\|_1 \ge S_{min}$$

← t های غیر از زمان سبقت

$$\text{و}\ X_{t+1}-X_t \ge 0$$

با رسم p_t برای تمام t ها می‌بینیم در بهترین حالت سبقت برای زمان T-start تا T-end به صورت تدریجی با [illegible] است و با کمترین میزان تغییرات شتاب لاین عوض می‌شود

8- ابتدا شروط لازم یعنی $Y \geq 0$ و $X[0] = Y[0] = 0$ و اینکه Y در بازه‌ی عبور مجاز نباید تغییر کند را در قید 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 ~~[illegible]~~ را اعمال می‌کنیم حال شرط داریم که سرعت از یک مقداری کم نشود یک شرط غیر کانوکس است ولی چون این شرط فقط در زمان عبور 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:

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

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

$$Y[\cancel{T\_start}\ T\_end :] = 1, \quad \|P_{t+1} - P_t\|_2 \leq S\_max, \quad \|X_{t+1} - X_t\|_1 \geq S\_min$$

برای t های غیر از زمان سبقت

$$X_{t+1} - X_t \geq 0$$

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