بهینه سازی غیر خطی

شکل کلی یک مسئله بهینه سازی غیر خطی به صورت زیر است:

$$
\begin{aligned}
&\min(\max) f(x) \\
&s.t \\
&h_i(x) = 0, i \in I = \{1, \dots, m\} \\
&g_j(x) \le 0, j \in J = \{1, \dots, k\}
\end{aligned}
$$

که $x \in \mathbb{R}^n$. اگر مساله قبل هیچ قیدی نداشته باشد گوییم مساله نامقید است.

اگر f, g_j ها محدب و h_i ها خطی باشند گوییم مساله یک مساله بهینه سازی محدب است.

تعریف: تابع f محدب است هرگاه برای هر x_1, x_2 در دامنه آن که خود محدب است و هر $0 \le \lambda \le 1$ داشته باشیم

$$f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$

تابع f مقعراست هرگاه $-f$ محدب نباشد.

مثال1: توابع خطی همگی محدب هستند.

$f(x) = x^2$ محدب است. $f(x) = \log x$ مقعر است.

تعریف: فرض کنید S یک زیر مجموعه از $\mathbb{R}^n$ و $f:S\to R$.گوییم f در $\bar{x}$ مشتق پذیر است هرگاه بردار $\nabla f(\bar{x})$ و تابع $\alpha:\mathbb{R}^n\to\mathbb{R}$ وجود داشته باشد که به ازای هر $x\in S$

$$f(\mathrm{x})=\mathrm{f}(\bar{x})+\nabla \mathrm{f}(\bar{x})^t(\mathrm{x}-\bar{x})+\|\mathrm{x}-\bar{x}\|\,\alpha(\bar{x}:\mathrm{x}-\bar{x})$$

و $\lim_{x\to\bar{x}}\alpha(\bar{x};\mathrm{x}-\bar{x})=0$ که $\nabla f(\bar{x})=\left(\frac{\partial f(\bar{x})}{\partial x_1},\ldots,\frac{\partial f(\bar{x})}{\partial x_n}\right)^t$.

به همین صورت گوییم f دوبار مشتق پذیر در $\bar{x}$ است هرگاه بردار $\nabla f(\bar{x})$ و ماتریس هسیان $H(\bar{x})$ (ماتریس هسیان را با $\nabla^2 f$ نیز نشان می دهند) و $\alpha:\mathbb{R}^n\to\mathbb{R}$ وجود داشته باشند که

$$f(\mathrm{x})=\mathrm{f}(\bar{x})+\nabla \mathrm{f}(\bar{x})^t(\mathrm{x}-\bar{x})+\frac{1}{2}(\mathrm{x}-\bar{x})^t H(\bar{x})(\mathrm{x}-\bar{x})+\|\mathrm{x}-\bar{x}\|^2\,\alpha(\bar{x};\mathrm{x}-\bar{x})$$

که $\lim_{x\to\bar{x}}\alpha(\bar{x};\mathrm{x}-\bar{x})=0$ و
$$H(\bar{x})=\begin{bmatrix} f_{11}(\bar{x}) & f_{12}(\bar{x}) & \ldots & f_{1n}(\bar{x}) \\ f_{21}(\bar{x}) & f_{22}(\bar{x}) & \ldots & f_{2n}(\bar{x}) \\ \vdots & \vdots & \vdots & \vdots \\ f_{n1}(\bar{x}) & f_{n2}(\bar{x}) & \ldots & f_{nn}(\bar{x}) \end{bmatrix}$$

مثال2: اگر $f(\mathrm{x}_1,\mathrm{x}_2)=2\,\mathrm{x}_1+6x_2-2x_1^2-3x_2^2+4x_1x_2$ آنگاه:

$$\nabla f(\bar{x})=\begin{bmatrix} 2-4\bar{x}_1+4\bar{x}_2 \\ 6-6\bar{x}_2+4\bar{x}_1 \end{bmatrix} \text{و} \quad H(\bar{x})=\begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}$$

مثال3:اگر $f(\mathrm{x}_1,\mathrm{x}_2)=\mathrm{e}^{2x_1+3x_2}$ آنگاه:

$$\nabla f(\bar{x})=\begin{bmatrix} 2e^{2\bar{x}_1+3\bar{x}_2} \\ 3e^{2\bar{x}_1+3\bar{x}_2} \end{bmatrix} \text{و} \quad H(\bar{x})=\begin{bmatrix} 4e^{2\bar{x}_1+3\bar{x}_2} & 6e^{2\bar{x}_1+3\bar{x}_2} \\ 6e^{2\bar{x}_1+3\bar{x}_2} & 9e^{2\bar{x}_1+3\bar{x}_2} \end{bmatrix}$$

توجه داشته باشید ماتریس هسیان همواره متقارن است.

یادآوری:

1)ماتریس متقارن A را معین مثبت گوییم هرگاه به ازای هر $x\neq 0$، $x^TAx>0$

2)اگر به ازای هر x، $x^TAx\geq 0$ آنگاه گوییم A نیمه معین مثبت است.

3) شرط لازم برای معین مثبت بودن یک ماتریس مثبت بودن عناصر قطری آن است.

قضیه 1: ماتریس A معین مثبت است اگر وتنها اگر همه مقادیر ویژه آن مثبت باشند.
همچنین ماتریس A نیمه معین مثبت است اگر وتنها اگر همه مقادیر ویژه آن بزرگتر یا مساوی صفر باشند.

اگرچه محاسبه مقادیر ویژه زمان بر است اما برای محاسبه چند مقدار ویژه بزرگ یا کوچک یک ماتریس روشهای کارایی وجود دارد. به عنوان مثال اگر کوچکترین مقدار ویژه یک ماتریس مثبت باشد آنگاه آن ماتریس معین مثبت است و اگر منفی باشد معین مثبت نیست.

قضیه 2: اگر $A\in ¡^{n\times n}$ ومتقارن باشد،گوییم A معین مثبت است اگر وفقط اگر مینورهای اصلی آن مثبت باشند.

مثال4) $A=\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}$ $\begin{array}{l}\Delta_1=\det(\mathrm{A}_1)=1\\ \Delta_2=\det(\mathrm{A}_2)=1-2\times 2=-3\end{array}$ پس ماتریس A معین مثبت نیست.

قضیه 3:اگر $f_1,\ldots,f_k$ توابع محدب باشند آنگاه توابع زیر نیز محدب هستند:

$$f(\mathrm{x})=\sum_{i=1}^{k} l_i f_i(\mathrm{x}),\ l_1,\ldots,l_k\geq 0 \quad (1$$

$$f(\mathrm{x})=\max_{1\leq i\leq k} f_i(\mathrm{x}) \quad (2$$

اثبات1: فرض کنید $k=2$ و $l_1,l_2\geq 0$ و طبق تعریف $f(\mathrm{x})=l_1 f_1(\mathrm{x})+l_2 f_2(\mathrm{x})$

$$\begin{aligned}
&f(l\,\mathrm{x}+(1-l)\,\mathrm{y})=l_1 f_1(l\,\mathrm{x}+(1-l)\,\mathrm{y})+l_2 f_2(l\,\mathrm{x}+(1-l)\,\mathrm{y})\leq\\
&l_1(l\,\mathrm{f}_1(\mathrm{x})+(1-l)\,\mathrm{f}_1(\mathrm{y}))+l_2(l\,\mathrm{f}(\mathrm{x})+(1-l)\,\mathrm{f}_2(\mathrm{y}))\leq\\
&l\,(l_1\,\mathrm{f}(\mathrm{x})+l_2\,\mathrm{f}_2(\mathrm{x}))+(1-l)(l\,\mathrm{f}(\mathrm{x})+l_2\,\mathrm{f}_2(\mathrm{y}))=l\,\mathrm{f}(\mathrm{x})+(1-l)\,\mathrm{f}(\mathrm{y})
\end{aligned}$$

پس f محدب است. حالت $k>2$ نیز بطور مشابه اثبات می شود.

اثبات 2: به عهده دانشجویان.

قضیه 4: تابع مشتق پذیر f محدب است اگر وفقط اگر $f(\mathrm{y})\geq \mathrm{f}(\mathrm{x})+\nabla\,\mathrm{f}(\mathrm{x})^T(\mathrm{y}-\mathrm{x})$ برای همه x,y ها.

قضیه 5:اگر f دوبار مشتق پذیر باشد آنگاه f محدب است اگر و فقط اگر ماتریس هسیان آن نیمه معین مثبت باشد.

اثبات: فرض کنید هسیان f نیمه معین مثبت باشد آنگاه برای هر $x, y \in D_f$ داریم:

$$f(\mathrm{y}) = \mathrm{f(x)} + \nabla \mathrm{f(x)}^T (\mathrm{y-x}) + \frac{1}{2}(\mathrm{y-x})^T \nabla^2 f(\mathrm{x}+q(\mathrm{y-x}))(\mathrm{y-x}) \geq$$
$$\mathrm{f(x)} + \nabla f(\mathrm{x})^T (\mathrm{y-x}) \qquad (*)$$

برای بعضی $0 \le q \le 1$. پس بنابر قضیه 4 f محدب است.

بالعکس: فرض کنید f محدب است ولی هسیان آن نیمه معین مثبت نیست. پس وجود دارد $y \in ¡^n$ بطوریکه

$$(\mathrm{y-x})^T \nabla^2 f(\mathrm{x})(\mathrm{y-x}) < 0$$

اگر $\nabla^2 f$ پیوسته باشد با انتخاب $y \in ¡^n$ مناسب، پس برای همه $0 \le q \le 1$ داریم:

$$(\mathrm{y-x})^T \nabla^2 \mathrm{f}(\mathrm{x}+q(\mathrm{y-x}))(\mathrm{y-x}) < 0$$

حال باتوجه به رابطه (*)داریم: $f(\mathrm{y}) < \mathrm{f(x)} + \nabla \mathrm{f(x)}^T (\mathrm{y-x})$ که این با محدب بودن f در تناقض است.

تعریف: گوییم x^* مینیمم موضعی(نسبی یا محلی) تابع f است اگر $f(\mathrm{x}^*) \le \mathrm{f(x)}, \forall \mathrm{x} \in \mathrm{N}$ که در ان N یک همسایگی از x^* است.

اگر برای هر $x \in ¡^n$ داشته باشیم $f(\mathrm{x}^*) \le \mathrm{f(x)}$ گوییم x^* یک مینیمم سراسری f است.

قضیه 6: اگر f یک تابع محدب باشد آنگاه هر مینیمم موضعی آن یک مینیمم سراسری است.

اثبات: فرض کنیم x^* مینیمم موضعی تابع f باشد. اگر x^* مینیمم سراسری f نباشد پس $\exists y \ s.t. \ f(\mathrm{y}) < \mathrm{f}(\mathrm{x}^*)$

از طرفی داریم

$$for\ l \in (0,1); \mathrm{f}(l\ \mathrm{x}^* + (1-l)\,\mathrm{y}) \le l\ \mathrm{f}(\mathrm{x}^*) + (1-l)\,\mathrm{f}(\mathrm{y}) <$$
$$l\ \mathrm{f}(\mathrm{x}^*) + (1-l)\,\mathrm{f}(\mathrm{x}^*) = \mathrm{f}(\mathrm{x}^*)$$

حال اگر $l \to 1$ آنگاه $l\,x^* + (1-l)\,\mathrm{y}$ به x^* نزدیک می شود که این با مینیمم موضعی بودن x^* در تناقض است. پس هر مینیمم موضعی f ، مینیمم سراسری آن نیز است.

نتیجه: اگر تابع اکیدا محدب f یک مینیمم سراسری داشته باشد، آنگاه یکتاست.

اثبات : فرض کنید x^*, y^* دو مینیمم موضعی f باشند، چون f اکیدا محدب است داریم:

$$f(l\,x^* + (1-l)\,\mathrm{y}^*) < l\ \mathrm{f}(\mathrm{x}^*) + (1-l)\,\mathrm{f}(\mathrm{y}^*) = \mathrm{f}(\mathrm{y}^*)$$

که این تناقض است چون در $(l\,x^* + (1-l)\,\mathrm{y}^*)$ مقدارتابع کمتر است پس با مینیمم سراسری بودن x^* در تناقض است. پس مینیمم سراسری f یکتاست.

قضیه 7: فرض کنید $f : S \to ¡$ محدب ومشتق پذیر روی S باشد آنگاه $\bar{x} \in X$ نقطه مینیمم سراسری f است اگر وفقط اگر $\nabla f(\bar{x}) = 0$.

اثبات: فرض کنید $\bar{x}$ مینیمم f است پس

$$\forall s \in S, \forall l \in ¡ ; f(\bar{x}) \le \mathrm{f}(\bar{x} + l\ \mathrm{s})$$

حال اگر $t > 0$ آنگاه $\frac{f(\bar{x} + l\ \mathrm{s}) - \mathrm{f}(\bar{x})}{l} \ge 0$ حال از سمت چپ وقتی $l \to 0^+$ حد می گیریم پس داریم: $\nabla f(\bar{x})^T\ s \ge 0$ چون S دلخواست پس $\nabla f(\bar{x})^T(-\mathrm{s}) \ge 0$.

بنابراین داریم $\nabla f(\bar{x})^T\ s = 0$. حال اگر $s = e_i$ پس $\nabla f(\bar{x}) = 0$.

بالعکس: اگر $\nabla f(\bar{x}) = 0$ و f محدب باشد آنگاه $\bar{x}$ مینیمم سراسری f است. چون f محدب است پس برای هر $y \in S$ داریم: $f(\mathrm{y}) \ge \mathrm{f}(\bar{x}) + \nabla\,\mathrm{f}(\bar{x})^T(\mathrm{y} - \bar{x})$.

بنابراین $f(\mathrm{y}) \ge \mathrm{f}(\bar{x})$ پس $\bar{x}$ مینیمم سراسری f است.

مثال: تابع زیر را درنصر بگیرید:

$$f(\mathrm{x}) = -\mathrm{Ln}(1-\mathrm{x}_1-\mathrm{x}_2) - \mathrm{Lnx}_1 - Lnx_2$$

گرادیان وهسیان آن بصورت زیر است:

$$\nabla f(\mathrm{x}) = \begin{pmatrix} \frac{1}{1-x_1-x_2} - \frac{1}{x_1} \\ \frac{1}{1-x_1-x_2} - \frac{1}{x_2} \end{pmatrix},$$

$$H(\mathrm{x}) = \begin{pmatrix} (\frac{1}{1-x_1-x_2})^2 + (\frac{1}{x_1})^2 & (\frac{1}{1-x_1-x_2})^2 \\ (\frac{1}{1-x_1-x_2})^2 & (\frac{1}{1-x_1-x_2})^2 + (\frac{1}{x_1})^2 \end{pmatrix}$$

به آسانی می توان نشان داد که هسیان تابع f در دامنه آن یعنی

$$X = \{(\mathrm{x}_1, \mathrm{x}_2) : \mathrm{x}_1 > 0, \mathrm{x}_2 > 0, \mathrm{x}_1 + \mathrm{x}_2 < 1\}$$

معین مثبت است وگرادیان ان در $\bar{x} = (\frac{1}{3}, \frac{1}{3})$ برابر صفر است. لذا این نقطه ،نقطه مینیمم سراسری منحصربفرد آن است.

$$\frac{1}{1-x_1-x_2} = \frac{1}{x_1}, \frac{1}{1-x_1-x_2} = \frac{1}{x_2} \Rightarrow x_1 = x_2 \Rightarrow \frac{1}{1-2x_1} = \frac{1}{x_1} \Rightarrow x_1 = x_2 = \frac{1}{3}$$

مثال: گرادیان و هسیان توابع زیر را بدست آورید.

$$f(\mathrm{x}) = \mathrm{x}^T px + 2q^T x + r.$$

$$f(\mathrm{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} x_i x_j + 2\sum_{i=1}^{n} q_i x_i + r,$$

$$\frac{\partial f(\mathrm{x})}{\partial x_k} = (\sum_{j=1}^{n} p_{kj} x_j)\, \mathrm{x}_k + (\sum_{i=1}^{n} p_{ik} x_i)\, \mathrm{x}_k + 2q_k x_k =$$

$$p_{k.} x + p_{.k}^T + 2q_k$$

(در اینجا $p_{k.}$ یعنی فقط سطر k ام در هر ستونی و $p_{.k}^T$ یعنی فقط ستون k ام در هر سطری)

هسیان آن بصورت زیر است:

$$\nabla f(\mathrm{x}) = \begin{bmatrix} \dfrac{\partial f(\mathrm{x})}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(\mathrm{x})}{\partial x_n} \end{bmatrix} = (\mathrm{p} + \mathrm{p}^T)\,\mathrm{x} + 2\,\mathrm{q}$$

$$\nabla^2 f(\mathrm{x}) = \begin{bmatrix} \dfrac{\partial^2 f(\mathrm{x})}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f(\mathrm{x})}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(\mathrm{x})}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(\mathrm{x})}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\mathrm{x})}{\partial x_2 \partial x_2} & \cdots & \dfrac{\partial^2 f(\mathrm{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(\mathrm{x})}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(\mathrm{x})}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(\mathrm{x})}{\partial x_n \partial x_n} \end{bmatrix}, \frac{\partial^2 f(\mathrm{x})}{\partial x_k \partial x_l} = p_{kl} + p_{lk}, \nabla^2 f(\mathrm{x}) = \mathrm{p} + \mathrm{p}^T$$

توضیحاتی بیشتردرباره محاسبات مثال بالا:

شکل کلی توابع درجه دوم در مسائل غیر خطی بصورت زیر است:

$$x^T A x + b^T x + c$$
$$A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^{n\times 1}, x \in \mathbb{R}^{n\times 1}, c \in \mathbb{R}$$

حال طبق صورت سوال داریم: $f(\mathrm{x}) = \mathrm{x}^T\ px + 2q^T x + r$

برای محاسبه $\mathrm{x}^T\ p x$:

$$\mathrm{x}^T\ px = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \left(\sum_{i=1}^{n} p_{i1} x_i, \dots, \sum_{i=1}^{n} p_{in} x_i\right) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} (\mathrm{p}_{ij} \mathrm{x}_i)\, \mathrm{x}_j$$