# ادامه مبحث بهینه سازی غیر خطی

هدف حل مسئله مینیمم سازی $f(x)$ است که در آن $f : \mathbb{R}^n \to \mathbb{R}$ . اگر f یک تابع پیوسته باشد و لزوما مشتق پذیر نباشد لذا از روشهای به اصطلاح مرتبه صفر برای حل مسئله قبل استفاده می کنیم. این روشها فقط براساس محاسبه مقدار تابع کار می کنند. از جمله معروفترین این روشها که در دستورfminsearch متلب نیز از آن استفاده می شود، روش سیمپلکس $Nelder - Mead$ است.

موضوع سمینار:

$$x = f \min search(\text{fun}, \text{x}_0)$$
$$\text{banana} = @(\text{x})100 * (\text{x}_2 - \text{x}_1{}^2)^2 + (1 - \text{x}_1)^2;$$
$$[x,\ fval] = f \min search(\text{banana}, [-1.2, 1])$$

خروجی متلب $x = [1.0000 \quad 1.0000],\ fval = 8.1777e - 010$

قضیه 1: فرض کنید $f : \mathbb{R}^n \to \mathbb{R}$ یک تابع مشتق پذیر باشد. اگر در نقطه $\bar{x}$ جهتی مانند $d$ باشد که $\nabla f(\bar{x})^T.d < 0$ آنگاه $l > 0$ وجود دارد بطوریکه $f(\bar{x} + l\ \text{d}) < \text{f}(\bar{x})$ . این چنین d را یک جهت کاهشی گویند.

اثبات: داریم

$$f(\bar{x} + l\ \text{d}) = \text{f}(\bar{x}) + l \nabla \text{f}(\bar{x})^T d + l \|d\| a(\bar{x}, l\ \text{d})$$

که اگر $l \to 0$ آنگاه $a(\bar{x}, l\ \text{d}) \to 0$ . حال داریم:

$$\frac{f(\bar{x} + l\ \text{d}) - \text{f}(\bar{x})}{l} = \nabla f(\bar{x})^T d + \|d\| a(\bar{x}, l\ \text{d})$$

حال چون $\nabla f\ (\bar{x})^T d < 0$ و $a(\bar{x}, l\ \text{d}) \to 0$ وقتی $l \to 0$ پس $f\ (\bar{x} + l\ \text{d}) - \text{f}(\bar{x}) < 0$.

نتیجه: فرض کنید $f$ در نقطه $\bar{x}$ یک تابع مشتق پذیر باشد.اگر $\bar{x}$ یک مینیمم موضعی باشد آنگاه $\nabla f(\bar{x}) = 0$ .

اثبات به برهان خلف: اگر $\nabla f(\bar{x}) \neq 0$ انگاه $d = -\nabla f(\bar{x})$ یک جهت کاهشی است که با مینیمم موضعی بودن $\bar{x}$ در تناقض است.

## روشهای جهت کاهشی برای حل مساله بهینه سازی نا مقید

1) یک حدس اولیه از جواب انتخاب کنید تا برقراری شرط توقف گامهای زیر را انجام دهید:

2) یک جهت حرکت انتخاب کنید.

3)طول گام مناسب $l_k > 0$ را انتخاب کنید.

4)قرار دهید $x_{k+1} = x_k + l_k d_k$

دو نکته اصلی در الگوریتم قبل ،انتخاب یک جهت حرکت و محاسبه طول گام مناسب است.هم برای انتخاب جهت و هم محاسبه طول گام روشهای متعددی ارائه شده است که در ادامه برخی از Hنها را معرفی می کنیم.

## روش بر مبنای گرادیان:

در این روش الگوریتم از یک نقطه اولیه شروع و با حرکت در جهت منفی گرادیان با طول گام مناسب عمل می کند.

$$x_{k+1} = x_k - a_k \nabla f(\mathrm{x}_k), a_k > 0$$

برای تعیین $a_k$ تابع زیر را برحسب $a_k$ در نظر بگیرید:

$$g(a_k) = \mathrm{f}(\mathrm{x}_k - a_k \nabla f(\mathrm{x}_k))$$

هدف پیدا کردن بهترین $a_k$ است که $g(a_k) < \mathrm{f}(\mathrm{x}_k)$ .

**مثال** 1: تابع زیر را درنظر بگیرید:

$$f(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3) = (\mathrm{x}_1 - 4)^4 + (\mathrm{x}_2 - 3)^2 + 4(\mathrm{x}_3 + 5)^4$$

با شروع از $x_0 = [4 \quad 2 \quad -1]^T$ داریم:

$$\nabla f(x_0) = \left[4(\mathrm{x}_1-4)^3 \quad 2(\mathrm{x}_2-3) \quad 16(\mathrm{x}_3+5)^3\right]^T$$
$$\nabla f(x_0) = [0 \quad -2 \quad 1024]^T$$
$$x_1 = x_0 - a_0 \nabla f(\mathrm{x}_0)$$
$$a_0 = \arg\min_{a \geq 0} \ f(x_0 - a\nabla f(\mathrm{x}_0))$$
$$= \operatorname*{argmin}_{a \geq 0}(0 + (2 + 2a - 3)^2 + 4(-1 - 1024a + 5)^4)$$
$$= \arg\min_{a \geq 0} f_0(a).$$

با استفاده از روش وتری $a_0 = 3.967 \times 10^{-3}$ و در تکرار جدید نیز بصورت زیر است:

$$x_1 = x_0 - a_0 \nabla f(\mathrm{x}_0) = [4.000 \quad 2.008 \quad -5.062]^T$$

برای تکرار دوم داریم: $\nabla f(\mathrm{x}_1) = [0.000 \quad -1.984 \quad -0.003875]^T$ و

$$a_1 = \arg\min(0 + (2.008 + 1.984a - 3)^2 + 4(-5.062 + 0.003875a + 5)^4)$$

$$= \arg\min_{a \geq 0} f_1(a)$$

که $a_1 = 0.5$ بدست می اید و $x_2 = x_1 - a_1 \nabla f(\mathrm{x}_1) = [4.000 \quad 3.000 \quad -5.060]^T$

برای تکرار سوم داریم: $\nabla f(\mathrm{x}_2) = [0.000 \quad 0.000 \quad -0.003525]^T$ و

$$a_2 = \arg\min_{a \geq 0}(0.000 + 0.000 + 4(-5.060 + 0.003525 + 5)^4)$$
$$= \arg\min_{a \geq 0} f_2(a)$$

که $a_2 = 16.29$ بدست می آید و در نتیجه $x_3 = [4.000 \quad 3.000 \quad -5.002]^T$. توجه کنید که مینیمم واقعی $[4 \quad 3 \quad -5]^T$ است که پس از سه تکرار تقریبا به آن رسیده ایم.

همانطور که در مثال فوق دیدیم با انتخاب یک حدس اولیه و یک جهت کاهشی همچون جهت منفی گرادیان الگوریتم آغاز و با حل یک مسئله مینیمم سازی یک متغییره طول گام یعنی $a$ محاسبه و تکرار جدید بدست می آید. روند قبل تا زمانی که شرایط توقف نقض شوند ادامه خواهد یافت. معروف ترین شرایط توقف عبارتند از :

1)ع حداکثر تکرار 2) حداکثر تعداد محاسبه تابع 3) $e = 10^{-6}, \left\|\nabla f(\mathrm{x}^k)\right\| \leq e$

ولی در مسائل مختلف می تواند متفاوت باشد.

**مثال 2:** تابع درجه دوم $f(\mathrm{x})=\frac{1}{2}x^TQx-b^Tx$ را در نظر بگیرید که در آن $Q\in \mathbb{R}^{n\times n}$ یک ماتریس متقارن ونیمه معین مثبت است. تکرارهای روش گرادیان به صورت زیر است:

$$x_{k+1}=x_k-a_kg_k$$
$$a_k=\arg\min_{a\geq 0} f(\mathrm{x}_k-a\,\mathrm{g}_k)$$

که در آن $\mathrm{g}_k=\nabla f(\mathrm{x}_k)$ و

$$a_k=\arg\min_{a\geq 0}(\frac{1}{2}(\mathrm{x}_k-a\,\mathrm{g}_k)^T\ \mathrm{Q}(\mathrm{x}_k-a\,\mathrm{g}_k)-(\mathrm{x}_k-a\,\mathrm{g}_k)^T\ \mathrm{b})$$

چون تابع محدب است پس اگر $g_k=0$ آنگاه $x_k$ جواب بهینه است والگوریتم خاتمه پیدا می کند. ولی اگر $g_k\neq 0$ انگاه تابع $f_k(a)$ محدب است. پس کمترین مقدار آن در ریشه مشتق اتفاق می افتد: $f'_k(a)=(\mathrm{x}_k-a\,\mathrm{g}_k)^TQ(-\mathrm{g}_k)-\mathrm{b}^T(-\mathrm{g}_k)$

$f'_k(a)=0$ آنگاه $a\,\mathrm{g}_k{}^T\,Qg_k=(\mathrm{x}_k{}^T\mathrm{Q}-b^T)\,\mathrm{g}^k$. اما $x_k{}^TQ-b^T=g_k{}^T$ پس $a_k=\frac{g_k{}^Tg_k}{g_k{}^TQg_k}$. بنابراین $x_{k+1}=x_k-(\frac{g_k{}^Tg_k}{g_k{}^TQg_k})\,\mathrm{g}_k$

محاسبات لازم برای رسیدن به فرمول $a_k$:

$$f(a)=\frac{1}{2}(\mathrm{x}_k-a\,\mathrm{g}_k)^TQ(\mathrm{x}_k-a\,\mathrm{g}_k)-\mathrm{b}^T(\mathrm{x}_k-a\,\mathrm{g}_k)$$
$$=\frac{1}{2}(\mathrm{g}_k{}^T\mathrm{Q}\mathrm{g}_k)a^2+(-\mathrm{x}_k{}^T\mathrm{Q}\mathrm{g}_k+\mathrm{b}^T\mathrm{g}_k)a+\frac{1}{2}x_k{}^TQx_k-b^Tx_k$$

$f$ محدب است پس مینیمم آن در ریشه $f'(a)$ اتفاق می افتد:

$$f'(a)=g_k{}^TQg_ka+(-\mathrm{x}_k{}^T\mathrm{Q}\mathrm{g}_k+\mathrm{b}^T\mathrm{g}_k)=0$$
$$a=\frac{x_k{}^TQg_k-b^Tg_k}{g_k{}^TQg_k}=\frac{(\mathrm{x}_k{}^T\mathrm{Q}-\mathrm{b}^T)\,\mathrm{g}_k}{g_k{}^TQg_k}=\frac{g_k{}^Tg_k}{g_k{}^TQg_k}$$

مثال 3: تابع $f(x_1,x_2)=x_1^2+x_2^2$ را در نظر بگیرید. با شروع از هر نقطه دلخواه در یک تکرار روش تندترین کاهش به جواب بهینه آن یعنی $x^*=0\in\mathbb{R}^2$ می رسد.

$$x^1=x^0-\frac{\begin{pmatrix}2x_1^0 & 2x_2^0\end{pmatrix}\begin{pmatrix}2x_1^0\\2x_2^0\end{pmatrix}}{\begin{pmatrix}2x_1^0 & 2x_2^0\end{pmatrix}\begin{pmatrix}2&0\\0&2\end{pmatrix}\begin{pmatrix}2x_1^0\\2x_2^0\end{pmatrix}}\begin{pmatrix}2x_1^0\\2x_2^0\end{pmatrix}=x^0-\frac{(4x_1^0)^2+(4x_2^0)^2}{(8x_1^0)^2+(8x_2^0)2}\begin{pmatrix}2x_1^0\\2x_2^0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.$$

مثال 4: تابع $f(x_1,x_2)=\frac{x_1^{\ 2}}{5}+x_2^{\ 2}$ را در نظر بگیرید در مقایسه با مثال قبل ،دراین مثال روش تندترین کاهش تکرارهای زیادی برای رسیدن به جواب نیاز دارد.

قضیه همگرایی: فرض کنید $f:\mathbb{R}^n\to\mathbb{R}$ یک تابع مشتق پذیر پیوسته روی مجموعه بسته و کراندار $S=\{x\in\mathbb{R}^n\,|\,f(x)\le f(x^{(0)})\}$ باشد. آنگاه هر نقطه حدی از دنباله تولید شده توسط روش تندترین کاهش در شرط $\nabla f(\bar{x})=0$ صدق می کند.

برای توابع درجه دوم محدب داریم: $\frac{f(x')-f(x^*)}{f(x)-f(x^*)}\le(\frac{A/a-1}{A/a+1})^2$

که در آن $A$ بزرگترین مقدار ویژه هسیان تابع است و $x'=x+a\,d$ .

- نامساوی قبل به نامساوی کانترویچ معروف است. کانترویچ برنده جایزه نوبل در سال 1975بخاطر ارائه کاربردهای برنامه ریزی خطی در اقتصاد شد.
- متاسفانه این کران در حل مسائل اغلب اتفاق می افتد که نشان از بد بودن روش تندترین کاهش در حل مسائلی است که نسبت مقادیر ویژه بزرگ به کوچک آن بسیار بزرگ است.
- چون در نزدیکی جواب بهینه توابع غیر درجه دو نیز همانند درجه دو هستند پس چنین رفتاری برای آن توابع نیز اتفاق می افتد.

مثال 5: تابع $f(x)=\frac{1}{2}x^TQx-c^Tx+10$ را در نظر بگیرید. خروجی الگوریتم تندترین کاهش برای این مساله به صورت زیر است.

$$c=\begin{pmatrix}14\\6\end{pmatrix},Q=\begin{pmatrix}20&5\\5&2\end{pmatrix}$$

مثال 6: تابع $f(\mathrm{x}) = \mathrm{x}_1 - 0.6x_2 + 4x_3 + 0.25x_4 - \sum_{i=1}^{4}\log(\mathrm{x}_i) - \log(5 - \sum_{i=1}^{4} x_i)$ را درنظر بگیرید.

همانطور که با آزمون روش تندترین کاهش روی مثال های متعددی اعم از درجه2،غیر درجه 2 دیدیم،این روش می تواند بسیار کند باشد هرچند از نظر محاسباتی روش کم هزینه ای است لذا استفاده ازاین روش به تنهایی برای حل مسائل بهینه سازی نامقید توصیه نمی شود. ولی میتوان در روشهای ترکیبی یا مسائل خاص از توانایی این روش بهره برد.

مثال 6: فرض کنید $f(\mathrm{x}) = \mathrm{x}_1 + \dfrac{x_2}{x_1 x_2 - 1}, x \in \mathbb{R}^2$ باشروع از $x^{(\mathrm{k})} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$ داریم:

$$\nabla f(\mathrm{x}^{(\mathrm{k})}) = \begin{pmatrix} \frac{37}{64} \\ \frac{21}{64} \end{pmatrix}, x^{(\mathrm{k+1})} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} - a^{(\mathrm{k})} \begin{pmatrix} \frac{37}{64} \\ \frac{21}{64} \end{pmatrix}$$

$$g(a^{(\mathrm{k})}) = 3 - a^{(\mathrm{k})}\frac{37}{64} + \frac{(3 - \frac{37}{64}a^{(\mathrm{k})})^2}{(3 - \frac{37}{64}a^{(\mathrm{k})})(3 - \frac{21}{64}a^{(\mathrm{k})}) - 1}$$

ملاحظه می کنید که تابع یک متغیره که مینیمم آن بایستی برای تعیین طول گام محاسبه شود می تواند بسیار پیچیده باشد.

## روش نیوتن:

دراین روش یک تقریب درجه دوم از تابع در نقطه $x^{(\mathrm{k})}$ بصورت زیر استفاده می شود:

$$\mathrm{f(x)} \approx \mathrm{f(x}^{(k)}) + (\mathrm{x} - \mathrm{x}^{(\mathrm{k})})^T g^{(\mathrm{k})} + \frac{1}{2}(\mathrm{x} - \mathrm{x}^{(\mathrm{k})})^T F(\mathrm{x}^{(k)})(\mathrm{x} - \mathrm{x}^{(\mathrm{k})}) \triangleq q(\mathrm{x})$$

سپس با استفاده از شرط لازم مرتبه اول داریم: $0 = \nabla q(x) = g^{(k)} + F(x^{(k)})(x - x^{(k)})$

حال اگر هسیان معین مثبت باشد داریم: $x^{(k+1)} = x^k - F(x^{(k)})^{-1} g^{(k)}$

طول گام در این حالت 1 فرض شده است ولی در حالت کلی لزوما اینطور نیست.

مثال 1: تابع زیر را در نظر بگیرید.

$$f(x_1,x_2,x_3,x_4) = (x_1+10x_2)^2 + 5(x_3-x_4)^2 + (x_2-2x_3)^4 + 10(x_1-x_4)^4$$

با شروع از نقطه $x^{(0)} = \begin{pmatrix} 3 & -1 & 0 & 1 \end{pmatrix}^T$ داریم:

$$\nabla f(x) = \begin{bmatrix} 2(x_1+10x_2)+40(x_1-x_4)^3 \\ 20(x_1+10x_2)+4(x_2-x_3)^3 \\ 10(x_3-x_4)-8(x_2-2x_3)^3 \\ -10(x_3-x_4)-40(x_1-x_4)^3 \end{bmatrix}$$

$$F(X) = \begin{bmatrix} 2+120(x_1-x_4)^2 & 20 & 0 & -120(x_1-x_4)^2 \\ 20 & 200+12(x_2-2x_3)^2 & -24(x_2-2x_3)^2 & 0 \\ 0 & -24(x_2-2x_3)^2 & 10+48(x_2-2x_3)^2 & -10 \\ -120(x_1-x_4)^2 & 0 & -10 & 10+120(x_1-x_4)^2 \end{bmatrix}$$

تکرار اول:

$$g^{(0)} = \begin{bmatrix} 306 & -144 & -2 & -310 \end{bmatrix}^T$$

$$F(x^{(0)}) = \begin{bmatrix} 482 & 20 & 0 & -480 \\ 20 & 212 & -24 & 0 \\ 0 & -24 & 58 & -10 \\ -480 & 0 & -10 & 490 \end{bmatrix}, F(x^{(0)})^{-1} = \begin{bmatrix} 0.1126 & -0.0089 & 0.0154 & 0.1106 \\ -0.0089 & 0.0057 & 0.0008 & -0.0087 \\ 0.0154 & 0.0008 & 0.0203 & 0.0155 \\ 0.1106 & -0.0087 & 0.0155 & 0.1107 \end{bmatrix}$$

$$F(x^{(0)})^{-1} g^{(0)} = \begin{bmatrix} 1.4127 & -0.8413 & -0.2540 & 0.7460 \end{bmatrix}^T$$

$$x^{(1)} = x^{(0)} - F(x^{(0)})^{-1} g^{(0)} = \begin{bmatrix} 1.5873 & -0.1587 & 0.2540 & 0.2540 \end{bmatrix}^T$$

$$f(x^{(1)}) = 318$$

تکرار دوم:

$$g^{(1)}=\begin{bmatrix}94.81 & -1.179 & 2.371 & -94.81\end{bmatrix}^T$$

$$F(\mathrm{x}^{(1)})=\begin{bmatrix}215.3 & 20 & 0 & -213.3\\ 20 & 205.3 & -10.67 & 0\\ 0 & -10.67 & 31.34 & -10\\ -213.3 & 0 & -10 & 223.3\end{bmatrix}$$

$$F(\mathrm{x}^{(1)})^{-1}g^{-1}=\begin{bmatrix}0.5291 & -0.0529 & 0.0846 & 0.0846\end{bmatrix}^T$$

$$x^{(2)}=x^{(1)}-F(\mathrm{x}^{(1)})^{-1}g^{(1)}=\begin{bmatrix}1.0582 & -0.1058 & 0.1694 & 0.1694\end{bmatrix}^T$$

$$f(\mathrm{x}^{(2)})=6.28$$

$$\mathrm{F}(\mathrm{x}^{(1)})^{-1}g^{(1)}=\begin{bmatrix}0.5291 & -0.0529 & 0.0846 & 0.0846\end{bmatrix}^T$$

تکرارسوم:

$$g^{(2)}=\begin{bmatrix}28.09 & -0.3475 & 0.7031 & -28.08\end{bmatrix}^T$$

$$F(\mathrm{x}^{(2)})=\begin{bmatrix}96.80 & 20 & 0 & -94.80\\ 20 & 202.4 & -4.744 & 0\\ 0 & -4.744 & 19.49 & -10\\ -94.80 & 0 & -10 & 104.80\end{bmatrix}, x^{(3)}=\begin{bmatrix}0.7037 & -0.0704 & 0.1121 & 0.1111\end{bmatrix}^T$$

$$f(\mathrm{x}^{(3)})=1.24$$

جستجوی خطی:

تابع $j$ در الگوریتم گرادیان و دیگر الگوریتمها که بحث خواهیم کرد نیاز به پیدا کردن مینیمم یک تابع یک متغییره هستیم. اگر این تابع مشتق پذیر باشد آنگاه می توان از روشهایی چون تنصیف ،نیوتن یا وتری برای پیدا کردن ریشه $j$ استفاده کرد.

$j(a)=\mathrm{f}(\mathrm{x}+a\,\mathrm{d}),\mathrm{d}$ جهت کاهشی

اگر تابع مشتق پذیر نباشد یا محاسبه مشتق آن سخت باشد می توان از روش دو بخشی یا تقسیم طلایی استفاده کرد.

الگوریتم روش نیوتن با طول گام یک:

1)دستگاه $F(\mathrm{x}^{(k)})\mathrm{d}^{(k)}=-g^{(k)}$ را حل کنید.

2)قرار دهید: $x^{(k+1)}=x^{(k)}+d^{(k)}$

مثال: تابع درجه دوم زیر را در نظر بگیرید: $f(\mathrm{x})=\frac{1}{2}x^TQx-x^Tb$

گرادیان وهسیان آن بصورت زیر است:

$g(\mathrm{x})=\nabla \mathrm{f}(\mathrm{x})=\mathrm{Qx}-\mathrm{b}$ ، $F(\mathrm{x})=\mathrm{Q}$

اگر هسیان معین مثبت باشد آنگاه روش نیوتن برای مساله بصورت زیر است:

$$\begin{aligned}x^{(1)}&=x^{(0)}-F(\mathrm{x}^{(0)})^{-1}g^{(0)}\\&=x^{(0)}-Q^{-1}(\mathrm{Qx}^{(0)}-\mathrm{b})=\mathrm{Q}^{-1}b=x^*\end{aligned}$$

به عبارت دیگر روش نیوتن در یک گام به جواب بهینه می رسد.

*همانطور که در روند تکراری نیوتن میبینیم برخلاف روش تندترین کاهش علاوه بر محاسبه ماتریس هسیان یک دستگاه معادلات نیز حل شود که در ابعاد بزرگ می تواند بسیار زمان بر باشد.

قضیه 1: اگر هسیان تابع $f$ معین مثبت باشد آنگاه جهت $d_k=-H(\mathrm{x}_k)^{-1}\nabla f(\mathrm{x}_k)$ یک جهت کاهشی است.

مثال: تابع $f(\mathrm{x})=-\ln(1-\mathrm{x}_1-\mathrm{x}_2)-\ln \mathrm{x}_1-\ln x_2$ را در نظر بگیرید.

گرادیان، هسیان و جواب بهینه آن بصورت زیر است:

$$f(\mathrm{x})=\begin{bmatrix}\dfrac{1}{1-x_1-x_2}-\dfrac{1}{x_1}\\\dfrac{1}{1-x_1-x_2}-\dfrac{1}{x_2}\end{bmatrix}$$

$$H(\mathrm{x}) = \begin{bmatrix} (\frac{1}{1-x_1-x_2})^2 + (\frac{1}{x_1})^2 & (\frac{1}{1-x_1-x_2})^2 \\ (\frac{1}{1-x_1-x_2})^2 & (\frac{1}{1-x_1-x_2})^2 + (\frac{1}{x_2})^2 \end{bmatrix}, x^* = (\frac{1}{3}, \frac{1}{3}), \mathrm{f}(\mathrm{x}^*) = 3.295836866$$

تکرارهای روش نیوتن در جدول زیر آمده است: