فصل ۷- بهگزینی نامقید چند متغیّره (جستجو در راستا)، روش‌های مستقیم

مقدمه – در دنیای سیال اندیشه، طرح روش‌های جستجوی چندمتغیره بسیار متنوّع و متکثر است. یک ایده این است که روشهای جستجوی مستقیم یک متغیره نظیر فیبوناچی را به چندمتغیره تعمیم دهیم و یک ایده دیگر تعمیم روشهای مبتنی بر مشتق نظیر نیوتن- رفسون میباشد. یک فلسفه عالی وسوسه میکند که متوسل به روشهای هندسی (در حالت خاص، فضایی- گرافیکی) شویم و یک فکر پست مدرن میگوید از طبیعت الهام بگیریم و روش های رقابتی، تکاملی، الگوریتم بقای انواع، سردشدن فلز و نظایر آن را مطرح کنیم.

به‌هرحال، اگر همچنان به‌طور تاریخی، به تقسم‌بندی روش‌های کلاسیک بپردازیم، در یک تقسیم‌بندی کلی ابتدا می‌توان میزان حرکت مناسب از نقطه جاری جستجو به نقطه بعدی را به دو طایفه کلی، یعنی روش‌های «جستجو در راستا[1]»ی مناسب و روش‌های «مبتنی بر نواحی تقریب[2]» تفکیک کرد. در فامیلی روشهای نوع اول (جستجو در راستا)، می‌توان باز هم به دودسته مستقیم (استفاده از فقط مقادیر تابع هدف) و غیرمستقیم (استفاده مزید از گرادیان یا سایر خصوصیات تابع هدف) قائل شد.

در این مقال، از هندسه (استدلال) کمک می‌گیریم و دسته‌بندی صدها روش کلاسیک بهینه‌سازی را به کمک دو مفهوم گام بهینه و جهت (امتداد) جستجو برای مقدمه‌چینی و معرفی طائفه اول (یعنی روش‌های جستجو در راستا) و از نوع مستقیم آنها انجام می‌دهیم. بدین منظور ابتدا به دو مثال انگیزشی می‌پردازیم.

مثال ۱. یک تابع دو متغیره کوادراتیک (بدون ترم تداخلی) در نظر بگیرید:

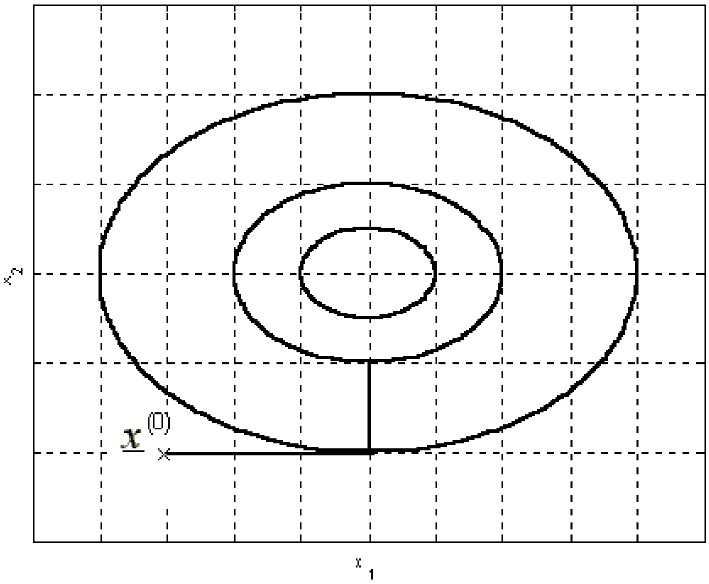
$$y = f(x_1, x_2) = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \qquad \text{یا} \qquad y = f(x_1, x_2) = \frac{(x_1 - x_1^*)}{a^2} + \frac{(x_2 - x_2^*)}{b^2}$$

گراف کنتوری یا منحنیهای تراز این تابع به‌صورت بیضی‌های متحدالمرکز میباشد و چون ترم تداخلی (عبارت $x_1 \times x_2$) ندارد، بیضی‌ها افقی هستند. برای عبارت اول، مینیمم در مبدأ (یعنی $(0,0)$)و برای عبارت دوم، مینیمم در نقطه (x_1^*, x_2^*) قرار دارد.

حال ایده به این شکل است، اگر از یک نقطه (حدس) اولیه مثل $\underline{x}_\circ$ شروع کنیم و در جهت محور x_1، به‌اندازه‌ای پیش رویم که بر یک کنتور مماس شود، عملاً توانسته‌ایم در جهت انتخابی یک گام بهینه برداریم. قبل و بعد از نقطه مماس (در راستای محور x_1) مقدار تابع بیشتر است و مینیمم در نقطه تماس قرار دارد. دقت شود این مینیمم جواب مساله نیست، بلکه فقط در این راستا، بهینه است.

به‌هر حال برای ادامه تکرار، باید کاری کنیم. ایده جستجوی متناوب[3] یا جستجوی تک‌راستایی بر این قاعده استوار است که جهت بعدی، عمود بر جهت قبلی باشد، یعنی بعد از نقطه تماس، جستجو را در جهت x_2 انجام دهیم و دوباره گام بهینه را انتخاب میکنیم. برای حالت کلی (یعنی n متغیره)، راستای جستجو متناوباً عوض میشود، بطور مثال برای سه متغیره، اگر برای اولین تکرار، جهت $[1\,0\,0]^T$ را انتخاب می‌کنید و گام بهینه را در این راستا پیدا می‌کنید، برای تکرار دوم، جهت $[0\,1\,0]^T$ و برای تکرار سوم، جهت $[0\,0\,1]^T$ را انتخاب کنید و هر بار عمل مینیمم سازی را تکرار کنید.

---
[1] Line Search
[2] Trust Regions
[3] Alternating search or univariate search



شکل ۱- گراف کنتوری یک تابع درجه دوم یا مجذوری (بدون ترم تداخل).

لذا، می‌بینید که این روش یک روش جستجوی مستقیم است. برای پروراندن ایده و درک تحلیلی سایر روشهای جستجوی مستقیم از این نکته مشترک استفاده می‌کنیم که اگر از $\underline{x}^{(0)}$ (نقطه اولیه) شروع کنیم، برای نقطه بعدی یعنی $\underline{x}^{(1)}=\underline{x}^{(0)}+\Delta\underline{x}$، چگونه $\Delta\underline{x}$ را انتخاب کنیم؟ اگر $\Delta\underline{x}$ را به شکل $\Delta\underline{x}=\lambda\underline{s}$ نمایش دهیم:

آنگاه برای محاسبه $\Delta\underline{x}$، اوّلاً) به کدام طرف رو کنیم (امتداد خوب کجاست یا بهترین $\underline{s}$ کدامست)؟

ثانیاً) به‌اندازه چه‌قدر در جهت جستجو حرکت کنیم(λ چقدر است؟)

لازم به ذکر است که بحث جهت از موقعی پدیدار می‌شود که از یک بُعدی به چندبُعدی حرکت کنیم. در بحث یک بُعدی، مفهوم بزرگتر و کوچکتر (یا معادلاً چپ و راست، بالا و پائین یا کمتر و بیشتر) مطرح است، درحالی‌که به محض ورود به عالم دوبُعدی، این مفاهیم حذف می‌شوند مگر اینکه به نحوی مساله نورم، قدر، میزان یا اندازه را دخیل کنیم.

نکته: تمامی روشهای کلاسیک بهینه‌سازی چندمتغیّره (از طائفه جستجوی راستائی) ، بر اساس بازی‌کردن با آن $\underline{s}$ (جهت جستجو) و λ (طول گام) می‌باشد.

نکته: بین دو مفهوم λ و $\underline{s}$ ، آنچه که مهمّتر و اولی‌تر است، جهت جستجو ($\underline{s}$) می‌باشد. نتایج تحقیقات ۳۵ سال اخیر نیز روی این مفهوم متمرکز بوده است، چون‌که مفهوم λی بهینه روشن و مسأله یک مسأله بهینه‌سازی یک متغیّره معروف به جستجوی راستا می باشد.

جستجوی راستا – این مفهوم همانطور که در شرح و جزئیات مسائل خواهید دید، عملاً یک بهینه‌سازی یک متغیّره می‌باشد و نوعاً بدین شکل می باشد:

مطلوبست به حداقل رساندن تابع (یک متغیّر تابع اسکالر چند متغیّره) در نقاطی بر روی خطّ $\underline{x}^{(prv.)}+\lambda\underline{s}$ (یا هر نوع نمایشی از راستا که بر حسب طول گام λ پارامتریزه شده‌باشد): $f(\underline{x}^{(prv.)}+\lambda\underline{s})=\Phi(\lambda)$

اگر $\underline{x}^{(prv.)}$ و $\underline{s}$ معلوم باشند، تابع هدف مذکور، فقط تابع λ است، پس کافیست از تکنیک‌های معلوم یک‌متغیّره نظیر تقسیم طلایی و میان‌یابی درجه دوم استفاده‌کنیم. بدیهی است بعد از معلوم شدن $\lambda^{opt.}$ ، می‌توان حدس بعدی را به صورت زیر پیداکرد: $\underline{x}^{(nxt.)}=\underline{x}^{(prv.)}+\lambda^{opt.}\underline{s}$

به‌عبارت دیگر به‌طور خودکار، حدس بعدی معلوم می‌شود و عملاً جمله عمومی تکرار یا موتور جستجو برای روش‌های بهینه سازی نامقیّد چند متغیّره به صورت بالا، طرح می‌شود. نکته دیگری که شایان توجّه است، عدم اصرار بر پیداکردن دقیق $\lambda^{opt.}$ می‌باشد، چرا که لااقل در اوایل کار، محاسبه خیلی دقیق $\lambda^{opt.}$، مهم نیست! فافهم!

به زبان الگوریتمیک و اصطلاح برنامه‌نویسی کامپیوتری، زیرالگوریتم "جستجوی راستا" رایج‌ترین و متداول‌ترین زیربرنامه کاربردی در بهینه‌سازی و روش‌های جستجوی چندمتغیّره می‌باشد:

اگر ورودی زیربرنامه شامل بردار $\underline{x}^{(prv.)}$ (به‌مفهوم نقطه و مختصات) و $\underline{s}$ (به‌مفهوم جهت، امتداد و راستا) باشد، آنگاه خروجی (برگشتی) زیربرنامه باید مقدار (اسکالر) λای باشد که مقدار تابع هدف $f(\underline{x}^{(prv.)}+\lambda\underline{s})$ را مینیمم کند.

مثال ۲– تابع چند متغیّره زیر را در نظر بگیرید. از آنالیز می‌دانیم که برای تابع اسکالر چندمتغیّره، بیشترین رشد در جهت گرادیان تابع است، پس در جهت منفی گرادیان، بیشترین کاهش را داریم، لذا $\underline{s}$ را منفی گرادیان بگیرید، سپس $\lambda^{opt.}$ را محاصره کرده و بعد مقدار بهینه آن را محاسبه کنید. طول گام اولیه را 0.05 و نقطه اولیه $[1 \quad 2]^T$ را اختیار کنید:

$$f(\underline{x}) = x_1^4 - 2x_1^2x_2 + x_2^2 + x_1^2 - 2x_1 + 5$$

در این مثال، جهت جستجو را منفی گرادیان انتخاب کردیم در حالی‌که در مثال قبلی، راستای محورهای متغیرهای مستقل (متغیرهای تصمیم‌گیری) را متناوباً در هر نوبت در نظر می‌گرفتیم.

$$\underline{s} = -\nabla f = -\begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}^T = -\begin{bmatrix} 4x_1^3 - 4x_1x_2 + 2x_1 - 2 \\ -2x_1^2 + 2x_2^2 \end{bmatrix}$$

$$\underline{x}^{(0)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \rightarrow f(\underline{x}^{(0)}) = 5 \quad , \quad -\nabla f\big|_{\underline{x}^{(0)}} = -\begin{bmatrix} -4 \\ 2 \end{bmatrix}$$

$$\lambda^{(0)} = 0.05 \rightarrow \underline{x}^{(1)} = \underline{x}^{(0)} + 0.05\underline{s} \rightarrow \begin{cases} x_1^{(1)} = x_1^{(0)} + (0.05)(4) = 1.2 \\ x_2^{(1)} = x_2^{(0)} + (0.05)(-2) = 1.9 \end{cases}$$

$$\rightarrow f(\underline{x}^{(1)}) = 4.25 \prec 5 \rightarrow$$ مقدار تابع کم شد.

یک گام دیگر در همان جهت و با طول گام دوبرابر (برخی پیشنهاد می‌کنند ۱۰ برابر باشد):

$$\lambda^{(1)} = 2\lambda^{(1)} = 0.1 \rightarrow \begin{cases} x_1^{(2)} = x_1^{(1)} + (0.1)(4) = 2.6 \\ x_2^{(2)} = x_2^{(1)} + (0.1)(-2) = 1.7 \end{cases}$$

$$\rightarrow f(\underline{x}^{(2)}) = 5.1 \succ 5 \rightarrow$$ مقدار تابع زیاد شد.

نتیجه اینکه $\lambda^{opt.}$ بین صفر و 0.1 است و برای محاسبه آن باید تابع زیر را مینیمم کنیم و به لسان بهینه‌سازی باید در راستای منفی گرادیان، تابع هدف را مینیمم کنیم.

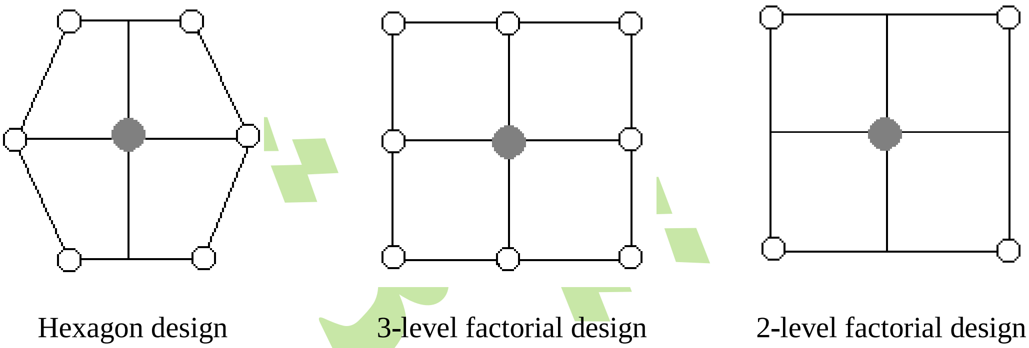
$$\begin{aligned} Minimum\ \Phi(\lambda) &= f(\underline{x}^{(0)} + \lambda(-\nabla f\big|_{\underline{x}=\underline{x}^{(0)}})) \\ &= (x_1^{(0)} + 4\lambda)^4 - 2(x_1^{(0)} + 4\lambda)^2(x_2^{(0)} - 2\lambda) + \\ &= (x_2^{(0)} - 2\lambda)^2 + 2(x_1^{(0)} + 4\lambda) - 2(x_1^{(0)} + 4\lambda) + 5 \\ &= f(x_1^{(0)} + 4\lambda, x_2^{(0)} - 2\lambda) \end{aligned}$$

بعد از محاسبه بهینه‌سازی یک متغیره به جواب $\lambda^{opt.} = 0.04875$ میرسیم.

در ادامه، به معرفی روش‌های جستجوی مستقیم برای بهینه سازی توابع چند متغیّره بر حسب تقسیم‌بندی ذکر شده، یعنی جهت جستجو (جستوی راستا) و گام بهینه، می‌پردازیم.

جستجوی تصادفی[1] [1]: در این روش، اگر از $\underline{x}^{(0)}$ شروع کنیم، $\underline{\Delta x}$ به‌طور رندام انتخاب می‌شود. یعنی هم جهت جستجو و هم طول گام، همزمان انتخاب می‌شوند، سپس مقدار تابع (فراخوانی) حساب شده و مورد ارزیابی قرار می‌گیرد. احتمال رسیدن به مینیمم وقتی به یک میل می‌کند که تعداد تکرار بی‌نهایت باشد.

جستجوی شبکه‌ای[2] [2]: این روش ریشه در روش‌های آماری دارد. دو اصطلاح برای این روش به‌کار می‌رود، یکی نقطه شروع و دیگری طرح ارزیابی، مقدار تابع در نقاط طراحی حساب می‌شوند و سپس نقطه شروع بعدی را نقطه‌ای انتخاب می‌کنیم که کمترین مقدار را در طرح ارزیابی دارد. طرح‌های ارزیابی مختلفی وجود دارند که مشخصه آنها سطح ارزیابی و نقاط طراحی می‌باشد. چند طراحی نمونه در شکل ۲ آمده‌است.



شکل ۲- چند طرح نمونه برای ارزیابی (نقاط هاشورخورده نشانگر نقطه شروع در تکرار k‌ام می‌باشند).

یک عیب عمده این روش تعداد زیاد فراخوانی تابع است. تعداد فراخوانی تابع از فرمول زیر بدست می‌آید:

$$recall = (level)^n - 1$$

به‌طوری‌که $level$ معرّف سطح ارزیابی و n تعداد متغیّرهای تصمیم‌گیری می‌باشد. به‌طور مثال برای تابع هدف 10 متغیّره و طرّاحی سطح 3 نیاز به $3^{10} - 1 = 59048$ فراخوانی در هر تکرار جستجو داریم!

روش جستجوی متناوب: در این روش، $\underline{s}$ همیشه یکی از راستاهای محورهاست و در هر انتخاب یا تناوب، گام بهینه را پیداکرده و نقطه بعدی را اختیار می‌کنیم. یک عیب عمده این روش تعداد تکرار زیاد می‌باشد، گرچه در مثالی که زده شد تعداد تکرار کم بود. علّت آن این است که اوّلاً تابع هدف به فرم مجذوری بود و ثانیاً یک فرم خاصی از توابع کوادراتیک بود. به‌طور مشخّص فرم استفاده‌شده در مثال به‌صورت $f(\underline{x}) = \sum_{i=1}^{n} c_i x_i^2$ یعنی فاقد ترم‌های تداخلی بود و اگر به‌صورت کلی‌تر $f(\underline{x}) = \sum_{j=1}^{n}\sum_{i=1}^{n} d_{ij} x_i x_j$ باشد، آنگاه روش جستجوی متناوب، بسیار کُند عمل می‌کند.

به‌هرحال یک اصلاح برای این روش، روش هوک - جیوز[3] می باشد که در ادامه می‌آید [3].

روش هوک - جیوز – این روش گرچه در سال ۱۹۶۱ تدوین شده ولی روشی کارآ و زیرکانه است که از اطّلاعات مقادیر تابع در حدس های قبلی نیز استفاده می‌کند. روش جستجو بر مبنای یک رشته از مراحل اکتشافی

---
1 Random search
2 Grid search
3 Hook-Jeeves

و ارزیابی[1] حول یک نقطه مبنا می‌باشد که در صورت موفّق‌بودن با حرکات نقشه‌دار ادامه می‌یابد. الگوریتم کامل روش به شرح زیر است:

گام اول – یک نقطه مبنای اولیه مثل $\underline{x}^{(prv.)}$ و یک طول گام h متناظر با هر عنصر به شکل زیر در نظر بگیرید:

$$\underline{h} = [h_1 \quad h_2 \quad \cdots \quad h_n]^T$$

گام دوم – اکتشافی حول نقطه مبنا انجام دهید. هدف از این کار، جمع آوری اطلاعات در مورد رفتار محلی تابع می‌باشد. از این معلومات برای یک جهت احتمالی (جهت $\underline{s}$) برای یک حرکت نقشه‌دار مورد استفاده قرار می‌گیرد، بدین امید که کاهش بیشتری از مقدار تابع در این جهت حاصل شود. اکتشاف حول نقطه مبنای $\underline{x}^{(prv.)}$ به شرح زیر انجام می‌شود:

الف) مقدار تابع هدف را در نقطه مبنا – $f(\underline{x}^{(prv.)})$ – فراخوانی کنید.

ب) اکنون هر متغیر را به تناوب، با افزودن طول گام متناظر با آن تغییر دهید. بدین ترتیب که اوّل مقدار $f(\underline{x}^{(prv.)}+h_1\underline{e}_1)$ را حساب می‌کنیم، که در آن بردار ستونی $\underline{e}_1$، بردار واحد در جهت محور x_1 هاست، یعنی $\underline{e}_1 \triangleq [1 \quad 0 \quad 0 \quad \cdots \quad 0 \quad 0]^T$.

اگر مقدار تابع بدین طریق کاهش یافت آنگاه $\underline{x}^{(prv.)}+h_1\underline{e}_1$ را جایگزین $\underline{x}^{(prv.)}$ (آخرین نقطه مبنای در دسترس) کنید. در غیر اینصورت $f(\underline{x}^{(prv.)}-h_1\underline{e}_1)$ را بیابید و اگر تابع کاهش یافت، $\underline{x}^{(prv.)}-h_1\underline{e}_1$ را جایگزین $\underline{x}^{(prv.)}$ کنید. در صورتی که هیچ یک از این مراحل به کاهش تابع منجر نشد، $\underline{x}^{(prv.)}$ را تغییر نداده و تغییرات در جهت محور x_2 ها را دنبال کنید، یعنی مقدار $f(\underline{x}^{(prv.)}+h_2\underline{e}_2)$ که در آن $\underline{e}_2$ بردار واحد در جهت محورهاست – معروف به $\underline{e}_2 \triangleq [0 \quad 1 \quad 0 \quad \cdots \quad 0 \quad 0]^T$ را پیدا کنید و الی آخر. پس از بررسی تمامی n جهت (متغیر)، یک نقطه مبنای جدید مثل $\underline{x}^{(nxt.)}$ خواهیم داشت.

ج) اگر $\underline{x}^{(nxt.)} = \underline{x}^{(prv.)}$ بود، یعنی به هیچ کاهش در تابع نرسیده باشیم، اکتشاف حول همان نقطه مبنا ($\underline{x}^{(prv.)}$) ولی با طول گام کمتر، تکرار می‌شود. کاهش طول گام‌ها به یک مرتبه عددی پایین‌تر از مقدار قبلی آن، توصیه شده‌است.

د) اگر $\underline{x}^{(nxt.)} \neq \underline{x}^{(prv.)}$، یعنی موفق به یافتن کاهشی در تابع شده‌ایم. لذا با در نظرگرفتن موقعیت $\underline{x}^{(nxt.)}$ نسبت به $\underline{x}^{(prv.)}$، باید بتوانیم جهت جستجوی بهینه محلی را بیابیم. در روش ابداعی هوک – جیوز، محاسبه $\underline{s}$ موسوم به حرکات نقشه‌دار می‌باشد.

گام سوم – جهت جستجو را انتخاب کنید، به عبارت دیگر:

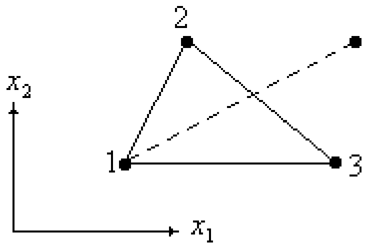
$$\Delta\underline{x} = \underline{x}^{(nxt.)} - \underline{x}^{(prv.)} = \lambda\underline{s} \quad \rightarrow \quad \begin{cases} \underline{s} = \underline{x}^{(nxt.)} - \underline{x}^{(prv.)} \\ \lambda = 1 \end{cases}$$

اگر طول گام (ها)، یعنی نورم $\underline{h}$ از مقدار از پیش‌تعیین‌شده‌ای کوچکتر شد، عملیات را خاتمه دهید. در غیر این‌صورت، نقطه مبنا را $\underline{x}^{(nxt.)}$ قرار داده و به گام دوم الگوریتم بروید.

روش سیمپلکس متوالی[2] [4]، در این روش، مشابه روش جستجوی شبکه‌ای، یک شکل هندسی معیّن (معروف به سیمپلکس) نظیر مثلّث برای دوبعدی و چهاروجهی منتظم برای سه‌بُعدی در نظر گرفته‌شده و فرض

[1] Exploration
[2] Sequential simplex

می‌شود مقدار (حدس) بهینه را محاصره کرده‌است. سپس با فراخوانی تابع در رئوس سیمپلکس و ارزیابی آنها سعی می‌کند از بیشترین مقدار دور شود. بطور مثال برای تابع هدف دو متغیره، سیمپلکس مثلّثی زیر را در نظر بگیرید و فرض کنید رأس 1 دارای بیشترین مقدار است، لذا با انعکاس از آن دور می‌شویم و بدین ترتیب جهت $\underline{s}$ را معلوم می‌کنیم.



شکل ۳- یک سیمپلکس نمونه برای مسئله دومتغیره.

از مشکلات این روش، انتخاب سایز سیمپلکس و چرخه تکرار[1] است و برای توابع هدف 6متغیره به بالا خیلی کُند عمل می‌کند. یک اصلاح برای این روش توسط نلدر[2] و مید[3] در سال ۱۹۶۵ ارائه شده‌است که سیمپلکس با سایز متغیّر می‌باشد [5]. الگوریتم کامل روش به شرح زیر است:

روش نلدر- مید: بنای فکری این روش، مقایسه مقادیر تابع در یک سیمپلکس دارای $n+1$ رأس (برای مسئله n- بعدی) می‌باشد. روش اصلی سیمپلکس ابتدائاً شامل یک سیمپلکس منظّم بود ولی نلدر و مید اصلاحیه‌های متعدّدی از این روش را پیشنهاد کردند که این امکان را می‌دهند که سیمپلکس‌ها نامنظّم باشند. حرکت و تغییر شکل سیمپلکس‌ها با به‌کارگیری سه عمل اساسی «انعکاس»، «انبساط» و «انقباض» انجام می‌پذیرد. الگوریتم روش به شرح زیر است:

گام اوّل – با $n+1$ نقطه در فضای nبعدی شروع می‌کنیم و مقادیر تابع هدف را در این نقاط (رئوس سیمپلکس) بدست می‌آوریم: $f_1 = f(\underline{x}_1) \quad , \quad f_2 = f(\underline{x}_2) \quad , \quad \cdots \quad , \quad f_{n+1} = f(\underline{x}_{n+1})$

گام دوم – از میان نقاط بالا سه نقطه را رکورد می‌کنیم، $\underline{x}_h$ با مقدار f_h که دارای بیشترین مقدار است، نقطه $\underline{x}_g$ با مقدار f_g که بیشترین مقدار بَعدی (کمتر از f_h) را دارد و نقطه $\underline{x}^{(prv.)}$ با مقدار $f^{(prv.)}$ که دارای کمترین مقدار می‌باشد.

گام سوم – مرکز تمام نقاط به جز $\underline{x}_h$ را پیدا می‌کنیم، این نقطه ($\underline{x}_\circ$) و متناظراً مقدار $f_\circ = f(\underline{x}_\circ)$ را محاسبه می‌کنیم: $\underline{x}_\circ = \frac{1}{n}\sum_{\substack{i=1 \\ i\neq n}}^{n} x_i$

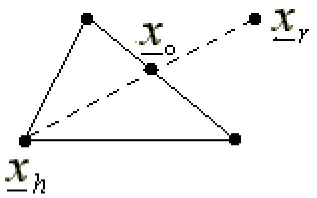
گام چهارم (انعکاس)- معقول به نظر می‌رسد که سعی‌کنیم از $\underline{x}_h$ دور شویم. نقطه $\underline{x}_r$، منعکس نقطه $\underline{x}_h$ نسبت به $\underline{x}_\circ$ را پیدا کرده و مقدار $f_r = f(\underline{x}_r)$ را محاسبه کنیم. انعکاس در شکل (۳-۸) برای یک مسأله دوبُعدی (سیمپلکس سه رأسی) نشان داده شده‌است. اگر $\alpha \succ 0$ را ضریب انعکاس تعریف کنیم:

$$\alpha \equiv \frac{\|\underline{x}_r - \underline{x}_\circ\|}{\|\underline{x}_\circ - \underline{x}_h\|}$$

---
[1] cycling
[2] Nelder
[3] Mead

آنگاه $\underline{x}_r$ را باین شکل یافته‌ایم:

$$\underline{x}_r - \underline{x}_\circ = \alpha(\underline{x}_\circ - \underline{x}_h) \quad \to \quad \underline{x}_r = (1+\alpha)\underline{x}_\circ - \alpha\underline{x}_h$$



شکل ۴– نمایش عمل انعکاس در فضای دومتغیره.

گام پنجم (انبساط)- حال مقدار f_r را با $\underline{x}^{(prv.)}$ مقایسه می‌کنیم.

وهله اول) اگر $f_r \prec f^{(prv.)}$ بود، یعنی کمترین مقدار تابع تا حال را بدست آورده‌ایم. به نظر می‌آید که از $\underline{x}_\circ$ به $\underline{x}_r$ جهت خوبی برای حرکت باشد. لذا، آن را در این جهت بسط می دهیم و $\underline{x}_e$ را پیدا کرده و $f_e = f(\underline{x}_e)$ را محاسبه می‌کنیم. اگر $\gamma \succ 1$ را ضریب انبساط تعریف کنیم: $\gamma \equiv \dfrac{\|\underline{x}_e - \underline{x}_\circ\|}{\|\underline{x}_r - \underline{x}_\circ\|}$

آنگاه $\underline{x}_e$ را با این شکل یافته‌ایم: $\underline{x}_e - \underline{x}_\circ = \gamma(\underline{x}_r - \underline{x}_\circ) \quad \to \quad \underline{x}_e = \gamma\underline{x}_r + (1-\gamma)\underline{x}_\circ$

حال اگر، الف) $f_e \prec f^{(prv.)}$ ، آنگاه $\underline{x}_h$ را به جای $\underline{x}_e$ قرار داده و همگرایی $n+1$ رأس سیمپلکس به نقطه مینیمم را بیازمایید (به گام نهم رجوع کنید). اگر همگرایی حاصل شده باشد، عملیات را خاتمه دهید و الاّ به گام دوم برگردید.

ب) $f_e \succ f^{(prv.)}$ ، آنگاه $\underline{x}_e$ را رها می‌کنیم. بدیهیست که زیاد در جهت $\underline{x}_\circ$ به $\underline{x}_r$ حرکت کرده‌ایم. در عوض $\underline{x}_h$ را به جای $\underline{x}_r$ قرار دهید که می دانیم نتیجه بهتری بدست داد (گام ششم/ وهله اول). سپس آزمون همگرایی را انجام دهید. در صورتیکه همگرا نباشد به گام دوم برگردید.

وهله دوم) اگر $f_r \succ f^{(prv.)}$ بود ولی $f_r \prec f_g$ آنگاه $\underline{x}_r$ از بدترین دو نقطه سیمپلکس بهتر است و $\underline{x}_h$ را به جای $\underline{x}_r$ قرار میدهیم و همگرایی را می‌آزمائیم، در صورت عدم همگرایی به وهله اوّل همین گام، قسمت (ب) بروید.

وهله سوم) اگر $f_r \succ f^{(prv.)}$ و $f_r \succ f_g$ به گام بعدی بروید.

گام ششم (انقباض) - f_r و f_h را مقایسه کنید:

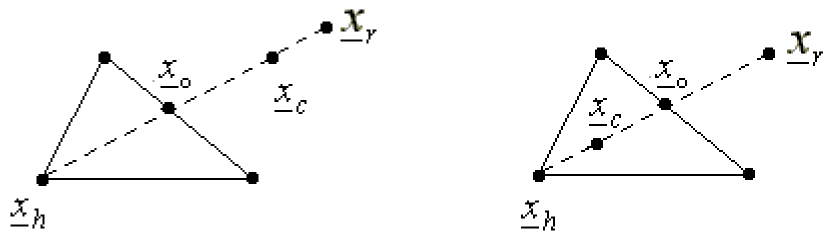
وهله اوّل: اگر $f_r \succ f_h$ ، آنگاه «انقباض» را انجام دهید (وهله دوم همین گام) ولی اگر $f_r \prec f_h$ بود، $\underline{x}_h$ را به جای $\underline{x}_r$ و f_h را به جای f_r قرار دهید. به خاطر داشته‌باشید که طبق گام پنجم/ وهله سوم داریم $f_r \succ f_g$. سپس به وهله بعدی همین گام بروید.

وهله دوم: در این حالت داریم $f_r \succ f_h$ ، یعنی انگار از $\underline{x}_\circ$ به سمت $\underline{x}_h$ زیاد دور شده‌ایم. پس سعی می‌کنیم با انقباض آن را تصحیح کرده و $\underline{x}_c$ را بیابیم: (به شکل ۵ رجوع کنید)

اگر $0 \prec \beta \prec 1$ را به‌صورت زیر و به اسم ضریب انقباض تعریف کنیم: $\beta \equiv \dfrac{\|\underline{x}_c - \underline{x}_\circ\|}{\|\underline{x}_h - \underline{x}_\circ\|}$

آنگاه $\underline{x}_c - \underline{x}_\circ = \beta(\underline{x}_h - \underline{x}_\circ) \quad \to \quad \underline{x}_c = \beta\underline{x}_h + (1-\beta)\underline{x}_\circ \quad \to \quad f_c = f(\underline{x}_c)$

باید دقت‌کرد، اگر $f_r \prec f_h$ بود (همانطور که در بالا اشاره شد)، ابتدا $\underline{x}_r$ را به جای $\underline{x}_h$ قرار داده و سپس منقبض می‌کنیم (به شکل ۵ رجوع شود):



شکل ۵- نمایش عمل انقباض در فضای دومتغیّره.

گام هفتم- حال f_c و f_h را با هم مقایسه می‌کنیم.

وهله اول) اگر $f_c \prec f_h$ ، آن‌گاه $\underline{x}_c$ را به جای $\underline{x}_h$ قرار داده و همگرایی را می‌آزمائیم. در صورت عدم همگرایی به گام دوم بر می‌گردیم.

وهله دوم) اگر $f_c \succ f_h$ ، به نظر میآید که تمام تلاشمان (انعکاس، انبساط و سپس انقباض) جهت پیداکردن مقداری کوچکتر از f_h عقیم مانده است، لذا، امیدوار به همگرایی و رسیدن به تولرانس خطا و توقّف جستجو هستیم، لذا به گام هشتم (کاهش سایز سیمپلکس) بروید.

گام هشتم – در این گام سایز سیمپلکس را با نصف کردن فاصله هر نقطه (رأس) سیمپلکس از نقطه $\underline{x}^{(prv.)}$ (نظیر کمترین مقدار تابع هدف)، کاهش می‌دهیم.

بنابراین به جای $\underline{x}_i$، نقطه $\underline{x}^{(prv.)}+\frac{1}{2}(\underline{x}_i-\underline{x}^{(prv.)})$ یا معادلاً $\frac{1}{2}(\underline{x}_i+\underline{x}^{(prv.)})$ را قرار می‌دهیم. سپس به ازای $i=1,2,\cdots,n+1$ ، f_i ها را محاسبه کرده و همگرایی را می‌آزمائیم، در صورت عدم همگرایی به گام دوم برگردید.

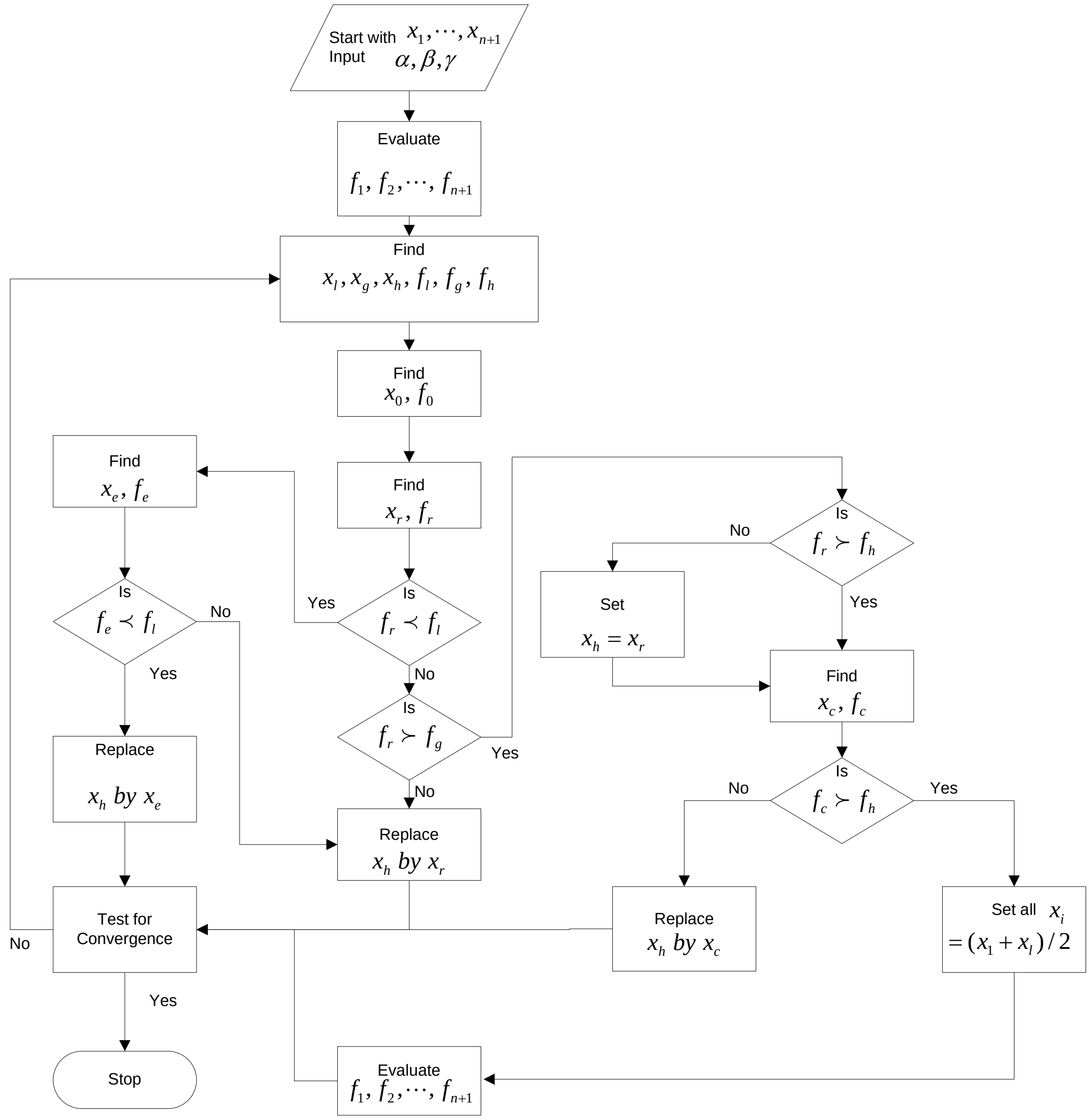
گام نهم – آزمون همگرایی بر اساس انحراف معیار $n+1$ مقدار تابع است که باید از یک مقدار کوچک از قبل تعیین شده مثل ε کوچکتر باشد، بنابر این:

$$\sigma^2=\sum_{i=1}^{n+1}\frac{\left(f_i-\overline{f}\right)^2}{n+1}$$

را محاسبه میکنیم که در آن $\overline{f}=\frac{\sum f_i}{n+1}$ (متوسط) می باشد.

اگر $\sigma \prec \varepsilon$ ، تمام مقادیر تابع خیلی نزدیک به هماند و به خاتمه جستجو رسیده‌ایم.

لازم به ذکر است که مقادیر α، β و γ پیشنهادی خود نلدر و مید به ترتیب معادل $\alpha=1$، $\beta=0.5$ و $\gamma=2$ می‌باشند. فلوچارت روش پیشنهادی نلدر – مید در شکل ۶ آمده است.



شکل ۶. فلوچارت پیشنهادی الگوریتم نلدر- مید.

روش پاول[1] یا جهت‌های جستجوی مزدوج[2] [5]

این روش بر پایه مطالعات نسبتاً عمیق و جدی روی توابع کوادراتیک (مجذوری) بنا شده‌است، به‌طوریکه منجر به یک کلاس بسیار مهمّ و رایج در روش‌های بهینه‌سازی موسوم به برنامه‌ریزی مجذوری شده‌است. نظر به اهمیّت وافر این بحث، در ادامه، سه نکته مهم در ترمینولوژی روش پاول را به‌طور نسبتاً مفصل و در تمام مقدمات و مسلّمات روش پاول مطرح می‌کنیم:

۱- **روش پاول، نمایش فرم مجذوری چندمتغیره** – یک تابع کوادراتیک متغیره به‌طور کلّی به‌شکل زیر نمایش داده می‌شود.

$$\phi(\underline{x}) = a + \underline{x}^T \underline{b} + \frac{1}{2}\underline{x}^T H \underline{x}$$

[1] Powell method
[2] conjugate search directions

همان‌طور که مشهود است عبارت مجذوری دارای سه جمله کلی می‌باشد، عبارت اوّل شامل یک ثابت است و عبارت دوم یک ترکیب خطّی از متغیّرهای مستقل است. عبارت سوم که مشخّصه مهمّ و غیرخطّی‌بودن عبارت مجذوریست، به لسان هندسه دیفرنسیالی، ترم دوخطی[۱]، به زبان هندسه تحلیلی ترم‌های تداخل[۲] مقطع مخروطی و در قاموس آنالیز، تقریب غیرخطی و مرتبه دوم یک تابع پیوسته با بسط تیلور زیر می‌باشد:

$$f\left(\underline{x}^{(k+1)}\right) \cong \phi\left(\underline{x}^{(k)}+\Delta\underline{x}^{(k)}\right) = f\left(\underline{x}^{(k)}\right) + \underline{\nabla^T f}.\Delta\underline{x}^{(k)} + \frac{1}{2}\Delta\underline{x}^{(k),T} H\left(\underline{x}^{(k)}\right)\Delta\underline{x}^{(k)}$$

ایده اساسی این طرز نمایش، برگرفته از بسط تیلور یک تابع اسکالر چندمتغیره می‌باشد:

$$f(\underline{x}) = f(\underline{x}_0) + \sum \frac{\partial f}{\partial x_i} x_i + \sum_i \sum_j \frac{\partial^2 f}{\partial x_i \partial x_j} x_i x_j + \cdots \qquad \text{(نمایش جبری)}$$

$$f(\underline{x}) = a + \underline{b}^T \underline{x} + (1/2)\underline{x}^T H \underline{x} + \cdots \quad , \qquad \text{(نمایش برداری)}$$

$$a \equiv f(\underline{x}_0) \quad , \quad \underline{b} \equiv +\nabla f\big|_{\underline{x}_0} \quad , \quad \left[H_{i,j}\right] \equiv \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{\underline{x}_0}$$

۲- **روش پاول، جستجو در راستای مشخص**- منظور از جستجو در راستای یک خط را در مثال انگیزشی قبلی دیده‌ایم، در اینجا می‌خواهیم بگوئیم برای توابع کوادراتیک، مقدار $\lambda^{opt.}$ را می‌توان بطور تحلیلی بدست آورد، یعنی میان‌یابی درجه دوم را می‌خواهیم تعمیم دهیم:

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \Delta\underline{x}^{(k)} \quad , \quad \Delta\underline{x}^{(k)} = \lambda\underline{s}^{(k)}$$

حال اگر رفتار تابع هدف را در حوالی حدس $\underline{x}^{(k)}$، به‌طور مجذوری و درجه دوم تقریب بزنیم:

$$f\left(\underline{x}^{(k+1)}\right) = f\left(\underline{x}^{(k)} + \lambda\underline{s}^{(k)}\right) \cong f\left(\underline{x}^{(k)}\right) + \underline{\nabla^T f}\left(\underline{x}^{(k)}\right)\lambda\underline{s}^{(k)} + \frac{1}{2}\left(\lambda\underline{s}^{(k)}\right)^T H\left(\underline{x}^{(k)}\right)\left(\lambda\underline{s}^{(k)}\right)$$

$$\frac{df}{d\lambda} = 0 \to \underline{\nabla^T f}\left(\underline{x}^{(k)}\right)\underline{s}^{(k)} + \lambda\left(\underline{s}^{(k)}\right)^T H\left(\underline{x}^{(k)}\right)\left(\underline{s}^{(k)}\right) = 0 \to \lambda^{opt.} = -\frac{\underline{\nabla^T f}\left(\underline{x}^{(k)}\right)\underline{s}^{(k)}}{\left(\underline{s}^{(k)}\right)^T H\left(\underline{x}^{(k)}\right)\left(\underline{s}^{(k)}\right)}$$

به شباهت بسیار زیاد میزان تغییر یا بهبود حدس در روش نیوتن (- رفسون) توابع هدف یک‌متغیّره با رابطه بالا دقّت کنید. در حالت یک متغیّره یک راستا ($\underline{s}$) بیشتر نداریم و آن‌هم واحد است، پس میزان تغییر (طول گام) معادلاً، و مشابهاً می شود $-f'/f''$ که در آخرین حدس در دسترس، فراخوانی می‌شود. در رابطه فوق نیز، افزون بر علامت منفی، در صورت بازنشانی از مشتقّ تابع (در حالت چند متغیّره، گرادیان) و در مخرج با نمادی از مشتق مرتبه دوم (در حالت چندمتغیّره، ماتریس هسیان) روبرو هستیم، فافهم.

۳- **روش پاول، جهت‌های (بردارهای) مزدوج** - به لسان جبری - هندسی، دوبردار $\underline{x}$ و $\underline{y}$ نسبت به ماتریس مربعی و مثبت معیّن Q مزدوجند اگر $\underline{x}^T Q \underline{y} = 0$

همین مفهوم مجرّد، به لسان برداری (فیزیکی)، شکل زیر را می‌گیرد:

دو بردار $\vec{x}$ و $\vec{y}$ نسبت به تبدیل مختصات Q متعامدند اگر: $\vec{x}' \cdot \vec{y}' = 0$ ؛ به‌طوریکه $\vec{x}'$ و $\vec{y}'$ همان بردارهای $\vec{x}$ و $\vec{y}$ در مختصات جدید هستند و علامت نقطه (٠)، نشانه ضرب داخلی دوبردار است.

[1] Bilinear
[2] Interacting term

نکته: اگر در یک مجموعه n تایی از بردارهای n بُعدی، دوبه‌دو مزدوج باشند، آن بردارها مستقل خطّی‌اند و می‌توانند تشکیل یک دستگاه پایه بدهند.

نکته: مفهوم مزدوج‌بودن، تعمیم مفهوم تعامد است مستقل از مختصات دستگاه موردنظر، فافهم.

نکته: مفهوم مزدوج‌بودن، تعمیم استقلال خطی است با ضرائب وزن دار، فافهم.

نکته: در حالت خاص $Q = I$ ، بطوریکه I ماتریس یکه باشد (عناصر روی قطر همه واحد و بقیه عناصر صفر)، مفهوم مزدوج بودن، تبدیل به متعامد بودن می‌شود.

نکته: برای توابع کوادراتیک، می‌توان جهت‌های مزدوج را با استفاده از بردار گرادیان و ماتریس هسیان، به‌طور تحلیلی به‌دست آورد.

نکته: می توان جهت‌های مزدوج را بدون استفاده از گرادیان (مشتق) بدست آورد!!!

نکته: مفهوم جهت‌های مزدوج برای کاندیداتوری جهت جستجوی بهینه ($\underline{s}$) منحصر به روش پاول نیست و بسیاری از روش‌های مدرن جستجوی چندمتغیّره بر اساس جهت‌های مزدوج کار می‌کنند.

نکته: ایده اساسی کار پاول برای بهره‌گیری از جهت‌های مزدوج به جای گرادیان، شاید از سؤال و انتزاع متناظر جبر (آنالیز) و هندسه به‌فرم زیر برگرفته شده باشد؛ از قضایای آنالیز ریاضی می‌دانیم که مقدار دیفرانسیل تابع هدف در نقطه مینیمم باید صفر باشد، یعنی $df = 0$ و این از نظر هندسی به‌معنای صفربودن همزمان تمام عناصر بردار گرادیان میباشد، یعنی $\underline{\nabla f} = \underline{0}$. در طرف مقابل، راستا یا جهت گرادیان، معرّف شدیدترین تغییر مقدار تابع، چه کاهش و یا افزایش می‌باشد. بنابراین وسوسه‌انگیز است اگر بخواهیم بدانیم به‌طور کلی فیلد برداری (یعنی مقدار تک‌تک عناصر بردار) در یک راستای دلخواه چگونه تغییر می‌کند؟ پس اگر ساختار f را مجذوری بگیریم، آنگاه گرادیان تابع به این شکل است؛ $\underline{\nabla f} = H\underline{x} + \underline{b}$ (دقّت کنید نقطه اکسترمم از حلّ دستگاه $H\underline{x} = \underline{b}$ به‌دست می‌آید، چون در اکسترمم $\underline{\nabla f} = \underline{0}$) و میزان تغییرات به‌فرم زیر محاسبه می‌شود؛ $\underline{\Delta(\nabla f)} = H\underline{\Delta x}$. حال فرض کنید در یک جهت مثل $\underline{u}$ حرکت کرده‌ایم و نقطه مینیمم را در این راستا یافته‌ایم، یعنی $\underline{\Delta x} = \lambda^* \underline{u}$، از طرفی این راستا (چه $\underline{u}$ و چه $\underline{\Delta x}$) باید عمود بر جهت گرادیان باشد (به گرافهای کنتوری تابع دومتغیره رجوع کنید، چون این راستا مماس بر ترازها یا کنتورها می‌باشد). حال فرض کنید می‌خواهیم در یک جهت دیگر مثل $\underline{v}$ حرکت کنیم ($\underline{\Delta x} = \delta^* \underline{v}$) و به یک نقطه مینیمم در آن راستا برسیم (یعنی یک مماس دیگر بر کنتور رسم کنیم) ولی این قید را داشته باشیم که همچنان گرادیان عمود بر $\underline{u}$ بماند، یعنی شرط زیر حفظ شود: (دقت شود در نقطه اپتیمم محلی، بردار گرادیان مقدارش صفر است - یک بردار بدیهی صفر است - به همین خاطر بردار تغییرات گرادیان را در نظر می‌گیریم)

$$\underline{u} \cdot \underline{\Delta(\nabla f)} = 0 \to \underline{u} \cdot (H\underline{\Delta x}) = 0 \to \underline{u} \cdot (H\delta^* \underline{v}) = 0 \to \underline{u} \cdot H \cdot \underline{v} = 0 \quad or \quad \underline{u}^T H \underline{v} = 0$$

شرط آخرِ معروف به شرط مزدوج‌بودن دو بردار $\underline{u}$ و $\underline{v}$ نسبت به ماتریس H می‌باشد.

در ادامه برای مأنوس شدن و فهم عمیق‌تر، به دو مثال انگیزشی، یکی «مفهوم جهت مزدوج» و دیگری محاسبه جهت‌های مزدوج بدون نیاز به مشتق (اساس کار پاول) می‌پردازیم.

مثال ۳- تابع هدف مجذوری $f(\underline{x}) = 2x_1^2 + x_2^2 - 3$ را با نقطه اولیه $\underline{x}^{(0)} = [1 \quad 1]^T$ در نظر بگیرید. جهت اولیه را $\underline{s}^{(0)} = [-4 \quad -2]^T$ در نظر گرفته و یک جهت مزدوج با فرض معادل ماتریس هسیان، نسبت به آن پیدا کنید.

علت صفر بودن عناصر غیرقطری، عدم وجود ترم های تداخلی می باشد.

$$\underline{s}^{(0)} = -\begin{bmatrix} 4 \\ 2 \end{bmatrix} \quad , \quad H(\underline{x}) = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

فرض می کنیم جهت مزدوج مورد سوال، به شکل زیر باشد:

$$\underline{s}^{(1)} = \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix}$$

در نتیجه هدف و مطلوب مساله، دو مجهول $s_1^{(1)}$ و $s_2^{(1)}$ می‌باشد.

شرط مزدوج‌بودن را می‌نویسیم:

$$(\underline{s}^{(0)})^T Q(\underline{s}^{(0)}) = 0\,,\, Q = H \rightarrow -\begin{bmatrix}4 & 2\end{bmatrix}\begin{bmatrix}4 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}s_1^{(1)}\\ s_2^{(1)}\end{bmatrix} = 0$$

چون به دستگاه معادله جبری همگن برخورده‌ایم، عملاً با بی‌نهایت جواب روبرو هستیم، لذا یکی از مجهولات را فیکس میگیریم یا معادلاً، یکی را بر حسب دیگری حلّ می‌کنیم. برای این‌کار $s_1^{(1)}$ را معادل واحد (محض سهولت) در نظر گرفته و $s_2^{(1)}$ را محاسبه می‌کنیم:

$$-\begin{bmatrix}4 & 2\end{bmatrix}\begin{bmatrix}4 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}1\\ s_2^{(1)}\end{bmatrix} = 0 \;\rightarrow\; s_2^{(1)} = -4 \;\rightarrow\; \underline{s}^{(1)} = \begin{bmatrix}1\\ -4\end{bmatrix}$$

مثال ۴- همان مسأله مثال قبل را در نظر گرفته و جهت های جستجو را متناوباً بصورت مزدوج در نظر بگیرید تا مینیمم را پیدا کنید.

حل: از $\underline{x}^{(0)}$ شروع کنید و در جهت $\underline{s}^{(0)}$ جستجو کنید تا مینیمم f را بیابید، یعنی $\lambda^{*(0)}$ را بیابید. از رابطه تحلیلی محاسبه طول گام بهینه اگر استفاده کنیم:

$$\lambda^{*(0)} = \frac{\begin{bmatrix}4 & 2\end{bmatrix}\begin{bmatrix}-4\\ -2\end{bmatrix}}{\begin{bmatrix}-4 & -2\end{bmatrix}\begin{bmatrix}4 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}-4\\ -2\end{bmatrix}} = \frac{20}{72}$$

$$\underline{x}^{(1)} = \underline{x}^{(0)} + \lambda^{*(0)}\underline{s}^{(0)} = \begin{bmatrix}1\\ 1\end{bmatrix} + \frac{20}{72}\begin{bmatrix}-4\\ -2\end{bmatrix} = \begin{bmatrix}-0.1111\\ 0.4444\end{bmatrix}$$

در گام بعدی، باید از $\underline{x}^{(1)}$ شروع کرده و در جهت مزدوج $\underline{s}^{(0)}$ ، یعنی $\underline{s}^{(1)}$ جستجو کنیم تا مینیمم را بیابیم.

$$\lambda^{*(1)} = \frac{\begin{bmatrix}-0.4444 & 0.8888\end{bmatrix}\begin{bmatrix}1\\ -4\end{bmatrix}}{\begin{bmatrix}1 & -4\end{bmatrix}\begin{bmatrix}4 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}1\\ -4\end{bmatrix}} = 0.1111$$

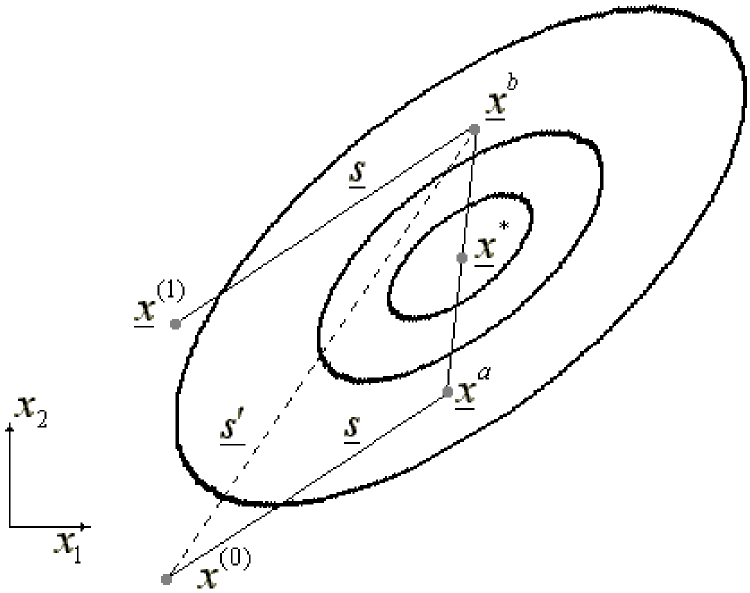
$$\underline{x}^{(2)} = \underline{x}^{(1)} + \lambda^{*(1)}\underline{s}^{(1)} = \begin{bmatrix}-0.1111\\ 0.4444\end{bmatrix} + 0.1111\begin{bmatrix}1\\ -4\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

یعنی در دو تکرار به جواب رسیدیم. قضیه پاول نیز همین را می‌گوید:

برای مینیمم‌سازی توابع مجذوری n بُعدی، اگر از جهت‌های مزدوج استفاده کنیم، حداکثر تا n تکرار (سعی و خطا) به جواب میرسیم.

اثبات قضیه پاول در این مقال نمی‌گنجد و اثبات‌شده آن را می‌پذیریم ولی برای مأنوس‌شدن با قضیه، برای حالت دو متغیره نمایش می‌دهیم که چرا جهات مزدوج این‌طور عمل می‌کنند و همزمان نشان می‌دهیم که جهات مزدوج را می‌توان بدون مشتق نیز بدست آورد. اصولاً انگیزه پاول برای ابداع این روش (به‌کارگیری جهات مزدوج) نیز بر این نیاز (عدم نیاز به مشتق‌گیری!) مبتنی بوده‌است. لازم به یادآوریست که به مثال مقایسه‌ای روش‌های وتری، یا روش شبه‌نیوتن برگردید. می‌خواهیم بگوئیم که روش پاول، عملاً تعمیم روش وتری به حالت n متغیره است!

منحنی‌های تراز یک تابع هدف دومتغیره مجذوری به‌طور کلّی در شکل زیر نمایش داده شده‌است. اگر تابع هدف شامل جملات تداخلی نباشد، منحنی های تراز بصورت بیضی‌های افقی یا عمودی ظاهر می‌شوند ولی در حالت کلّی به‌صورت زاویه دار و کج[1] ظاهر می‌شوند. محض یادآوری ذکر دو روش انگیزشی برای انتخاب جهت مناسب جستجو (جهت $\underline{s}$) خالی از فایده نخواهد بود. در روش تغییر جهت به تناوب، در صورتی که منحنی‌های تراز افقی یا عمودی باشند، با تغییر یا حرکت در محورهای افقی و عمودی (در حالت دومتغیّره) یا کلّی‌تر با تغییر متناوب جهت‌های عمود بر هم، می‌توانستیم با مینیمم تکرار و مستقل از حدس اولیه به جواب نهایی برسیم. در روش گرادیان، به‌طور شهودی و حسّی، حدس زدیم که بهترین جهت جستجو، جهت منفی گرادیان باشد. به‌هرحال، به‌نظر می‌رسد که پاول همان ایده جستجو به تناوب را در نظر گرفت و گفت اگر بیضی‌ها، کج بودند، آنگاه جهت‌های جستجوی مناسب نیز باید کج باشند ولی در عین حال متعامد تعمیم‌یافته یا به‌عبارت ریاضی مزدوج باشند، چون با تغییر مختصات موقعی فضای برداری محقق[2] تشکیل می‌شود که پایه‌های مولّد مستقل خطّی از هم باشند.



شکل ۷. منحنی‌های تراز (گراف کنتوری) یک تابع مجذوری به‌فرم کلی.

به شکل دقت کنید. از نقطه $\underline{x}^{(0)}$شروع کرده و در جهت جستجوی دلخواهی مثل $\underline{s}$ ، مینیمم تابع را پیدا میکنیم (با $\lambda^{opt.}$). این مینیمم میانی در نقطه $\underline{x}^a$ اتفاق میافتد. حال فرض کنید از یک نقطه دیگر مثل $\underline{x}^{(1)}$ که مخالف $\underline{x}^{(0)}$ است شروع کنیم و مینیمم تابع را در جهت $\underline{s}$ پیدا کنیم. فرض کنید این نقطه بهینه میانی $\underline{x}^b$ باشد. حال ادعا میکنیم برای توابع کوادراتیک، اگر نقطه بهینه نهایی (در شکل با $\underline{x}^*$ نمایش داده شده‌است) روی خط فاصل $\underline{x}^a$ و $\underline{x}^b$ قرار داشته‌باشد، آنگاه جهت $\underline{x}^b-\underline{x}^a$ ، نسبت به $\underline{s}$ مزدوج بوده‌است که این اتفاق برقرار شده‌است.

نکته: این طرز برخورد بسیار شبیه روش وتری میباشد، فتأمّل.

نکته: در روش وتری یا شبه‌نیوتن، وقتی تابع هدف یک متغیره از نوع درجه دوم (کوادراتیک) بود، آنگاه در حلّ غیرخطّی $f'(x)=0$ ، نقطه مینیمم روی خطّ $f'(x)$ وجود داشت، فافهم.

---

[1] skewed
[2] Realized

نکته: اگر بتوان ادّعای بالا را ثابت کرد، آنگاه توانسته‌ایم بدون نیاز به مشتق‌گیری (بدون نیاز به روش نیوتن) و البتّه فقط برای فرم درجه دوم (کوادراتیک)، نقطه بهینه را با مینیمم تکرار بدست آوریم (درست شبیه روش وتری که برای توابع درجه دوم مثل روش نیوتن، با مینیمم تکرار، بهینه نهایی بدست می‌آید)، فاصبروا.

اوّل، اثبات می‌کنیم، گرادیان تابع در $\underline{x}^a$ (نقطه‌ای که تابع هدف، مینیمم محلّی شده‌است) بر جهت $\underline{s}$ عمود است. می‌دانیم نقطه $\underline{x}^a$ یا $\underline{x}^b$ از روی مینیمم‌سازی تابع هدف در جهت $\underline{s}$ به‌دست آمده‌است:

فرض کنید در حالت کلی، $f(x)$ را می توان در این شعاع همسایگی با درجه دوم تقریب زد، پس با بسط تیلور:

$$f(\underline{x}) \cong f\left(\underline{x}^{(0)}\right) + \underline{\nabla}^T f\left(\underline{x}^{(0)}\right)\Delta\underline{x}^{(0)} + \frac{1}{2}\Delta\underline{x}^{(0),T} H \Delta\underline{x}^{(0)}$$

به‌طوریکه H ماتریس هسیان است، قبلاً دیده‌ایم که به‌طور تحلیلی میتوان λ بهینه را بدست‌آورد:

$$\lambda^{opt.} = -\frac{\underline{\nabla}^T f(\underline{x}^{(0)})\underline{s}}{\underline{s}^T H \underline{s}}$$

حال تابع را به شکل یا به زبان کوادراتیک مینویسیم تا قادر به جایگزینی در رابطه بالا باشیم.

$$f(\underline{x}) = a + \underline{x}^T \underline{b} + (1/2)\underline{x}^T H \underline{x} \quad \rightarrow \quad \underline{\nabla} f(\underline{x}) = \underline{b} + H\underline{x} \quad \rightarrow$$

$$\underline{\nabla} f(\underline{x}^{(0)}) = \underline{b} + H\underline{x}^{(0)} \quad \rightarrow \quad \lambda^{opt.} = -\frac{(\underline{b} + H\underline{x}^{(0)})^T \underline{s}}{\underline{s}^T H \underline{s}} \quad \rightarrow$$

$$(\underline{b} + H\underline{x}^{(0)})^T \underline{s} + \underline{s}^T H \underline{s}\lambda^{opt.} = 0$$

از طرفی با توجه به اتحاد $(AB)^T = B^T A^T$ ، می‌توان عبارات متّحد زیر را نوشت:

$$(\underline{b} + H\underline{x}^{(0)})^T \underline{s} = \underline{s}^T (\underline{b} + H\underline{x}^{(0)})$$

همچنین از تعریف جستجو در راستا (محاسبه $\lambda^{opt.}$) داریم:

$$\lambda^{opt.}\underline{s} = \underline{x}^a - \underline{x}^{(0)}$$

در نتیجه با ترکیب تعریف فوق و اتحاد مافوق، آخرین عبارت را می توان به شکل زیر نوشت:

$$\underline{s}^T (\underline{b} + H\underline{x}^{(0)}) + \underline{s}^T H(\underline{x}^a - \underline{x}^{(0)}) = 0 \quad \rightarrow$$

$$\underline{s}^T \underline{b} + \underline{s}^T H\underline{x}^{(0)} + \underline{s}^T H\underline{x}^a - \underline{s}^T H\underline{x}^{(0)} = 0 \quad \rightarrow$$

$$\begin{cases} \underline{s}^T (\underline{b} + H\underline{x}^a) = 0 \\ \underline{\nabla} f(\underline{x}) = \underline{b} + H\underline{x} \end{cases} \quad \rightarrow \quad \underline{s}^T \underline{\nabla} f(\underline{x}^a) = 0$$

و این یعنی گرادیان تابع در نقطه $\underline{x}^a$ بر جهت جستجو عمود است. لذا، اوّلین نتیجه یا لمّ حاصل از مطالعه توابع کوادراتیک به این شکل درآمد:

اگر $f(x)$ یک تابع به فرم مجذوری باشد، آنگاه گرادیان $f(x)$ در نقطه حداقلش در یک جهت جستجو مثل $\underline{s}$ ، عمود می‌باشد.

در ادامه، همین نتیجه بالا را می‌توانیم برای نقطه $\underline{x}^b$ نیز بگیریم:

$$\begin{cases} \underline{s}^T \underline{\nabla} f(\underline{x}^b) = 0 = \underline{s}^T (\underline{b} + H\underline{x}^b) \\ \underline{s}^T \underline{\nabla} f(\underline{x}^a) = 0 = \underline{s}^T (\underline{b} + H\underline{x}^a) \end{cases} \quad \rightarrow \quad (subtraction) \quad \rightarrow \quad \underline{s}^T H(\underline{x}^b - \underline{x}^a) = 0$$

و ادّعای اخیرالدعوی اثبات شد.

به‌هرحال، دومین لمّ یا نکته در بررسی و مطالعه توابع کوادراتیک چندمتغیره به شکل زیر می‌باشد.

دو جهت $\underline{x}^b - \underline{x}^a$ و $\underline{s}$ نسبت به ماتریس هسیان مزدوج هستند.

به‌طور خلاصه از همین شکل و بحث (دو نتیجه یا لمّ ذکرشده) اخیر، می توان این ایده را دنبال کرد، که اگر از یک جهت جستجوی اوّلیه و دلخواه مثل $\underline{s}$ شروع کنیم، آنگاه بهترین جهت جستجوی بعدی، جهت مزدوج با $\underline{s}$ است. این جهت مزدوج را برای توابع کوادراتیک می توان از روی $\underline{x}^b - \underline{x}^a$ به‌دست آورد. این ایده بعداً به شکل مدوّن و الگوریتمیک درآمده و به افتخار پژوهشگر آن، یعنی پاول بامسمّا شد و امروزه به‌عنوان قوی‌ترین روش جستجوی چندمتغیره مستقیم و بدون نیاز به گرادیان از آن نام برده می‌شود.

روش پاول- در این روش از یک نقطه (حدس اوّلیه) شروع می‌کند و در هر مرحله تکرار، در راستای جهت‌های مزدوج جستجو کرده تا به نقطه بعدی برسد. آن‌قدر این کار ادامه پیدا می‌کند تا به مینیمم تابع هدف برسد. دونکته مهمّ در این روش وجود دارد:

۱- در اولین مرحله، جهت‌های مزدوج اوّلیه، محورهای مختصات می‌باشد، یعنی در اوّلین مرحله تکرار، روش پاول، یک روش جستجوی متناوب می‌باشد.

۲- برخی جهت‌های مزدوج که در هر مرحله بدست می‌آید، ممکنست متناظر با $\lambda^{opt.}$ نزدیک به صفر باشد، لذا عاقلانه نیست که جهت جستجو را عوض‌کنیم. در روش پاول، یک آزمون برای چک‌کردن این نکته برقرار است.

برای تبیین روش و همچنین مأنوس شدن با روش پاول، به‌طور موازی تابع دومتغیّره را نیز مورد بررسی قرار می‌دهیم. فرض کنید یک نقطه اوّلیه مثل $\underline{x}^{(0)}$ داریم (برای حالت دو متغیّره: $\underline{x}^{(0)} = \begin{bmatrix} x_1^{(0)} & x_2^{(0)} \end{bmatrix}^T$)

جهت های اوّلیه را به‌صورت $\underline{s}_1^{(0)}, \underline{s}_2^{(0)}, \cdots, \underline{s}_n^{(0)}$ در نظر می‌گیریم یا برای راحتی، ماتریس $R^{(0)}$ که ستونهای آن، جهت‌های جستجو هستند را تعریف می‌کنیم:

$$R_{n\times n}^{(0)} = \begin{bmatrix} \underline{s}_1^{(0)} & \underline{s}_2^{(0)} & \cdots & \underline{s}_n^{(0)} \end{bmatrix}$$

برای حالت دو متغیّره، جهت‌های جستجو عبارتند از $\underline{s}_1^{(0)}$ و $\underline{s}_2^{(0)}$. چون در مرحله صفرم هستیم، جهت‌های مزدوج (مستقل خطّی) را در راستای محورهای مختصات می‌گیریم: $\underline{s}_1^{(0)} = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ و $\underline{s}_2^{(0)} = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$

و برای حالت کلی: $$R_{n\times n}^{(0)} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

برای طی کردن مرحله تکرار صفرم و رسیدن به مرحله یکم، بدین شکل عمل می‌کنیم:

از نقطه $\underline{x}^{(0)}$ شروع کرده و در جهت $\underline{s}_1^{(0)}$، تابع را مینیمم می‌کنیم تا به نقطه میانی بعدی برسیم:

$$\underline{x}_{(1)}^{(0)} = \underline{x}_{(0)}^{(0)} + \lambda_{(1)}^{*(0)} \underline{s}_1^{(0)}$$

لازم به‌ذکرست که زیرنویسها، معرّف تکرارهای میانی هستند، در حالی‌که بالانویسها، معرّف مراحل تکرار (اصلی) الگوریتم است.

سپس از نقطه جدید بدست‌آمده در جهت $\underline{s}_2^{(0)}$ حرکت می‌کنیم تا به نقطه مینیمم بعدی برسیم.

$$\underline{x}_{(2)}^{(0)} = \underline{x}_{(1)}^{(0)} + \lambda_{(2)}^{*(0)} \underline{s}_2^{(0)} = \underline{x}_{(0)}^{(0)} + \lambda_{(1)}^{*(0)} \underline{s}_1^{(0)} + \lambda_{(2)}^{*(0)} \underline{s}_2^{(0)}$$

در حالت دو متغیّره، $\underline{x}_{(1)}^{(0)}$ همان $\underline{x}^a$ می‌باشد و $\underline{x}_{(2)}^{(0)} - \underline{x}_{(0)}^{(0)}$ همان $\underline{x}^b$ می‌باشد. لذا برای حالت دومتغیّره، باید جهت بعدی که مینیمم در آن قرار دارد، در راستای $\underline{x}_{(2)}^{(0)} - \underline{x}_{(0)}^{(0)}$ بوده و به اندازه طول گام $\underline{\Delta x}^{opt.}$ باشد، یعنی

$$\underline{x}^{(1)} = \underline{x}^{(0)} + \underline{\Delta x}^{opt.}$$

$$\underline{x}_{(0)}^{(1)} = \underline{x}_{(2)}^{(0)} + \underline{x}_{(2)}^{(0)} - \underline{x}_{(0)}^{(0)} = 2\underline{x}_{(2)}^{(0)} - \underline{x}_{(0)}^{(0)}$$

برای حالت کلّی، دقت کنید $\underline{\Delta x}^{opt.}$ به‌این شکل بدست می‌آید:

$$\underline{\Delta x}^{opt.} = \underline{x}_{(2)}^{(0)} - \underline{x}_{(1)}^{(0)} = \sum_{i=1}^{n} \lambda_{(i)}^{*(0)} \underline{s}_i^{(0)}$$

نکته: دقّت شود که نقطه قبلی را همیشه آخرین نقطه‌ای می‌گیریم که در مراحل میانی مرحله تکرار بدست آورده‌ایم، یعنی در مثال بالا، $\underline{x}^{(0)}$ را با $\underline{x}_{(0)}^{(0)}$ نمی‌گیریم، بلکه آخرین نقطه را مورد توجه قرار می‌دهیم، یعنی نقطه $\underline{x}_{(2)}^{(0)}$. بطور کلی در روش پاول داریم: $\underline{x}_{(0)}^{(k)} = \underline{x}_{(n)}^{(k-1)}$

پس بطور خلاصه گام اوّل روش پاول را بطور کلّی می‌توان به این شکل بیان نمود:

گام اول (روش پاول)، حرکت از نقطه اوّلیه قبلی به نقطه اولیه بعدی

از $\underline{x}_{(0)}^{(k)}$ شروع کرده و در جهت $\underline{s}_1^{(k)}$ جستجو کنید و مینیمم تابع را بیابید($\lambda_{(1)}^{*(k)}$):

$$\underline{x}_{(1)}^{(k)} = \underline{x}_{(0)}^{(k)} + \lambda_{(1)}^{*(k)} \underline{s}_1^{(k)}$$

از نقطه مینیمم به‌دست آمده ($\underline{x}_{(1)}^{(k)}$) شروع کرده و در جهت دوم ($\underline{s}_2^{(k)}$)جستجو کنید و مینیمم تابع را بیابید:

$$\underline{x}_{(2)}^{(k)} = \underline{x}_{(1)}^{(k)} + \lambda_2^{*(k)} \underline{s}_2^{(k)}$$
$$= \underline{x}_{(0)}^{(k)} + \lambda_1^{*(k)} \underline{s}_1^{(k)} + \lambda_2^{*(k)} \underline{s}_2^{(k)}$$

و همین‌طور این کار را ادامه دهید تا همه جهات جستجو انجام شده‌باشد.

$$\vdots$$

$$\underline{x}_{(n)}^{(k)} = \underline{x}_{(n-1)}^{(k)} + \lambda_n^{(k)} \underline{s}_n^{(k)} = \underline{x}_{(0)}^{(k)} + \sum_{i=1}^{n} \lambda_{(i)}^{*(k)} \underline{s}_i^{(k)}$$

حال مقدار تابع را در $\underline{x}_{(n)}^{(k)}$ (آخرین نقطه‌ای که در این مرحله رسیده‌ایم) فراخوانی می‌کنیم و با محک اختتام آن‌را امتحان می‌کنیم، اگر به مینیمم رسیدیم که هیچ، یعنی $\underline{x}^{(*)} = \underline{x}_{(n)}^{(k)}$، بعد از k بار تکرار به مینیمم رسیده‌ایم ولی اگر محک اختتام، ارضاء نشد، آنگاه نقطه شروع تکرار بعدی همانست که الان رسیده‌ایم، یعنی: $\underline{x}_{(0)}^{(k+1)} = \underline{x}_{(n)}^{(k)}$

نکته قابل توجّه در اینجاست که در حال حاضر نقطه تکرار بعدی را می‌دانیم، ولی جهات جستجو چه؟ آیا همان قبلی‌ها را نگهداریم؟ اگر بلی در آن‌صورت روش همان روش جستجو به تناوب است. با توجه به مباحث قبلی، جهت مزدوج $\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)}$ را برای شروع تکرار بعدی انتخاب می‌کنیم. اگر تابع کوادراتیک باشد، باید مینیمم مسأله در جهت $\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)}$ قرار داشته باشد. افزون بر این، اگر بخواهیم همچنان جهات جستجو را مزدوج نسبت به هم بگیریم، امکان دارد در برخی مراحل میانی، بعضی λ^*ها نزدیک به صفر درآیند. بدین ترتیب، مجموعه $\underline{s}$ها، مستقل نخواهند بود و به‌عبارتی آن ماتریس R، دارای نقص رتبه[1] می‌شود. لذا، پاول پیشنهادکرد که قبل از ورود به مرحله تکرار بعدی و حفظ جهات جستجو یا تغییر آنها (لااقل یکی از آنها)، یک گام دیگر نیز (باندازه $(\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)})$ و مقدار تابع را در $2\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)}$ بیابیم که می‌شود جهت مزدوج جدید نسبت به آخرین جهت جستجو و البتّه کاندیدا برای گام دوم. سپس یک تستی انجام دهیم (گام سوم) آیا جهاتِ قدیمِ جستجو، مستقل خطّی مانده‌اند یا خیر. این آزمون را مفصلاً در گام سوم بررسی خواهیم‌کرد. پس به‌طور خلاصه گام بعدی به شرح زیر است:

گام دوم- حرکت در راستای $\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)}$ و محاسبه مقدار تابع در $2\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)}$، جهت سهولت، نمادگذاریهای زیر را انجام می‌دهیم:

مقدار تابع در نقطه شروع تکرار kام: $f_1 = f(\underline{x}_{(0)}^{(k)})$

مقدار تابع در نقطه آخر تکرار kام: $f_2 = f(\underline{x}_{(n)}^{(k)})$

[1] Rank Deficiency

مقدار تابع در نقطه محتمل مینیمم تابع هدف در تکرار kام: $f_3 = f(2\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)})$

گام سوم: معیار و یا تست برای تغییر دادن یا ندادن جهات جستجو.

فرض کنید در مراحل میانی مرحله تکرار kام، هر بار که به نقطه دیگر می‌رویم، مقدار کاهش f را بیابیم: (این کار را در گام اول باید انجام دهیم و در حافظه ذخیره کنیم).

$$\begin{aligned}
&\underline{x}_{(1)}^{(k)} = \underline{x}_{(0)}^{(k)} + \lambda_{(1)}^{*(k)}\underline{s}_1^{(k)} \quad \rightarrow \quad \Delta_{(1)}^{(k)} = f(\underline{x}_{(1)}^{(k)}) - f(\underline{x}_{(0)}^{(k)})\\
&\underline{x}_{(2)}^{(k)} = \underline{x}_{(1)}^{(k)} + \lambda_{(2)}^{*(k)}\underline{s}_2^{(k)} \quad \rightarrow \quad \Delta_{(2)}^{(k)} = f(\underline{x}_{(2)}^{(k)}) - f(\underline{x}_{(1)}^{(k)})\\
&\vdots\\
&\underline{x}_{(n)}^{(k)} = \underline{x}_{(n-1)}^{(k)} + \lambda_{(n)}^{*(k)}\underline{s}_n^{(k)} \quad \rightarrow \quad \Delta_{(n)}^{(k)} = f(\underline{x}_{(n)}^{(k)}) - f(\underline{x}_{(n-1)}^{(k)})
\end{aligned}$$

آنگاه روی همه Δهای فوق جستجو می‌کنیم، تا ببینیم کدام‌یک بیشترین کاهش را داده‌اند:

$$\Delta^{(k)} = \max_{i=1,\cdots,n}\left\{f(\underline{x}_{(i-1)}^{k}) - f(\underline{x}_{(i)}^{k})\right\}$$

جهت جستجوی متناظر با این $\Delta^{(k)}$ را با $\underline{s}_m^{(k)}$ نمایش می‌دهیم.

حال پاول پیشنهاد کرده‌است که دو شرط زیر را چک کنیم:

اگر $f_3 \succ f_1$ یا $(f_1 - 2f_2 + f_3)(f_1 - f_2 - \Delta^{(k)})^2 \geq 0.5\Delta^{(k)}(f_1 - f_3)^2$ آنگاه به مشکل برخورده‌ایم و خطر وابستگی خطّی هست، لذا برای جهات جستجوی بعدی و همچنین نقطه آولیه پاس بعدی باید تصمیم بگیریم؛ پاول پیشنهاد کرده‌است که جهات جستجو همان قبلی‌ها را در نظر بگیریم و نقطه اوّلیه را آخرین نقطه‌ای که رسیده‌ایم در نظر بگیریم:

$$\underline{s}_i^{(k+1)} = \underline{s}_i^{(k)} \quad , \quad \underline{x}_{(0)}^{(k+1)} = \underline{x}_{(n)}^{(k)}$$

در صورتی که شرط بالا برقرار نباشد، یعنی مسیر تکرار به اشکال برنخورده باشد، آنگاه جهت جستجوی جدید را موازی $\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)}$ انتخاب می‌کنیم:

$$\begin{bmatrix} \underline{s}_1^{(k+1)} \\ \underline{s}_2^{(k+1)} \\ \vdots \\ \underline{s}_n^{(k+1)} \end{bmatrix} = \begin{bmatrix} \underline{s}_1^{(k)} \\ \underline{s}_2^{(k)} \\ \vdots \\ \underline{s}_{m-1}^{(k)} \\ \vdots \\ \underline{s}_n^{(k)} \\ \underline{s}_{new}^{(k)} \end{bmatrix}$$

و نقطه اوّلیه بعدی $\underline{x}_{(0)}^{(k+1)}$ (یا همان $\underline{x}_{(n)}^{(k)}$) یا $2\underline{x}_{(n)}^{(k)} - \underline{x}_{(0)}^{(k)}$ ، هر کدام که مقدار کوچکتری از تابع بدهد را انتخاب می‌کنیم.

نکته: علّت انتخاب جهت جستجوی بعدی به شکل بالا، بخاطر نتیجه یا لمّ سومی است که پاول در مطالعات بیست‌ساله خود به آن رسیده‌است. پاول اثبات کرد که برای توابع کوادراتیک، اگر جهات جستجو نرمالیزه باشند (نرمالیزاسیون نسبت به ماتریس هسیان: $\left(\underline{s}_i^{(k)}\right)^T H\left(\underline{s}_i^{(k)}\right)=1$)، آنگاه دترمینان ماتریس جهات جستجو (همان R)، بیشترین مقدار خود را می‌گیرد اگر و تنها اگر $\underline{s}_i^{(k)}$ دوبه‌دو نسبت به H مزدوج باشند. لازم به ذکر است که در این قضیه، سایز ماتریس با دترمینان آن مشخصه‌سازی شده است، که خیلی فرض درستی نیست، فافهم.

گام چهارم – آخرین گام، معیار اختتام می‌باشد که نوعاً جنبه سلیقه‌ای دارد. به‌هرحال، پاول پیشنهاد کرده است که برای اختتام، میزان تغییر تک‌تک متغیرها از یک حدی کمتر باشد،

$$\left|x_i^{(k)} - x_i^{(k-1)}\right| \prec \varepsilon_i \quad i = 1,\cdots,n$$

به‌طوریکه x_i مولفه یا عناصر بردار $\underline{x}$ می‌باشد.

یک پیشنهاد دیگر، اندازه گیری نورم نقطه n بعدی در آخر هر مرحله تکرار می‌باشد:

$$\left\| x_{(n)}^{(k)} - x_{(0)}^{(k-1)} \right\| \prec \varepsilon$$

مثال۵. روش پاول را برای تابع نمونه آزمون[1] روزنبروک[2] بکار ببرید، حدس اولیه: $\underline{x}^{(0)} = \begin{bmatrix} -1.2 \\ 1.0 \end{bmatrix}$

$$Minimize \quad f(\underline{x}) = 100(x_2 - x_1^2)^2 + (1-x_1)^2$$

(دقت کنید، تابع روزنبروک کوادراتیک نیست)

حل: گام اول:

$$\underline{x}^{(0)} = \begin{bmatrix} -1.2 \\ 1.0 \end{bmatrix} \quad , \quad \underline{s}_1^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad , \quad \underline{s}_2^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad , \quad f(\underline{x}^{(0)}) = 24.2$$

$$\underline{x}_{(1)}^{(0)} = \begin{bmatrix} -1.2 \\ 1.0 \end{bmatrix} + \lambda_{(1)}^{*(0)} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rightarrow \lambda_{(1)}^{*(0)} \text{ محاسبات بهگزینی برای } \rightarrow \underline{x}_{(1)}^{(0)} = \begin{bmatrix} -0.995 \\ 1.000 \end{bmatrix} \rightarrow f(\underline{x}_{(1)}^{(0)}) = 3.990$$

$$\underline{x}_{(2)}^{(0)} = \underline{x}_{(1)}^{(0)} + \lambda_{(2)}^{*(0)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.995 \\ 1.000 \end{bmatrix} + \lambda_{(2)}^{*(0)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rightarrow \lambda_{(2)}^{*(0)} \text{ محاسبات بهگزینی برای } \rightarrow$$

$$\rightarrow \underline{x}_{(2)}^{(0)} = \begin{bmatrix} -0.995 \\ 0.999 \end{bmatrix} \rightarrow f(\underline{x}_{(2)}^{(0)}) = 3.980$$

گام دوم:

$$\left(\underline{x}_{(3)}^{(0)}\right)\tilde{\underline{x}}^{*(0)} = 2\underline{x}_{(2)}^{(0)} - \underline{x}_{(0)}^{(0)} = 2\begin{bmatrix} -0.955 \\ 0.99 \end{bmatrix} - \begin{bmatrix} -1.2 \\ 1.0 \end{bmatrix} = \begin{bmatrix} -0.79 \\ 0.98 \end{bmatrix} \rightarrow f(\tilde{\underline{x}}^{*(0)}) = 15.872$$

گام سوم:

$$\begin{cases} \Delta_{(1)}^{(0)} = 24.2 - 3.990 = 20.21 \\ \Delta_{(2)}^{(0)} = 3.99 - 3.980 = 0.01 \end{cases} \rightarrow \Delta^{(0)} = \max(20.21, 0.01) = 20.21$$

شرط اول : $(f_3 \succ f_1) \rightarrow 15.872 \nsucc 24.2$ برقرار نیست

یا شرط دوم : $(f_1 - 2f_2 + f_3)(f_1 - f_2 - \Delta^{(k)})^2 \geq 0.5\Delta^{(k)}(f_1 - f_3)^2$

$(24.2 - 2\times 3.98 + 15.872)(24.2 - 3.98 - 20.21)^2 \ngeq 0.5(20.21)(24.2 - 15.872)^2$ برقرار نیست

چون هر دو شرط برقرار نیست، لذا، باید جهت را عوض کنیم. جهت متناظری که باید عوض شود. اوّلین جهت است، چون متناظر با $\Delta_{(1)}^{(0)}$ است، بهرحال جهات جستجوی بعدی عبارتند از:

$$\underline{s}_1^{(1)} = \underline{s}_2^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\underline{s}_2^{(1)} = \underline{x}_{(2)}^{(0)} - \underline{x}_{(0)}^{(0)} = \begin{bmatrix} -0.995 \\ 0.990 \end{bmatrix} - \begin{bmatrix} -1.2 \\ 1.0 \end{bmatrix} = \begin{bmatrix} 0.205 \\ -0.010 \end{bmatrix}$$

باید نرمالیزه کنیم $\rightarrow \left\| \underline{s}_2^{(1)} \right\| = \sqrt{\underline{s}_2^{(1)T} \underline{s}_2^{(1)}} = 0.206 \rightarrow \underline{s}_2^{(1)} = \begin{bmatrix} 0.0205 \\ -0.010 \end{bmatrix} \Big/ 0.206 = \begin{bmatrix} 0.999 \\ -0.0488 \end{bmatrix}$

نقطه شروع بعدی:

$$\begin{cases} \underline{x}_{(0)}^{(1)} = 2\underline{x}_{(2)}^{(0)} - \underline{x}_{(0)}^{(0)} = 2\times\begin{bmatrix} -0.995 \\ 0.990 \end{bmatrix} - \begin{bmatrix} -1.2 \\ 1.0 \end{bmatrix} = \begin{bmatrix} -0.790 \\ 0.980 \end{bmatrix} \quad , \quad f = 15.872 \\ \underline{x}_{(0)}^{(1)} = \underline{x}_{(2)}^{(0)} = \begin{bmatrix} -0.995 \\ 0.990 \end{bmatrix} \quad , \quad f = 3.98 \end{cases}$$

چون $3.98 \prec 15.872$ ، پس نقطه شروع بعدی را $\begin{bmatrix} -0.995 \\ 0.990 \end{bmatrix}$ اختیار می‌کنیم.

---

[1] Benchmark
[2] Rosenbrook

در جدول زیر، برخی از مراحل میانی کارآورده شده است.

مرحله k ام تکرار	$\underline{s}_1^{(k)}$	$\underline{s}_2^{(k)}$	$\underline{x}^{(k)}$	$f(\underline{x}^{(k)})$
1	$[0 \quad 1]^T$	$[0.205 \quad -0.01]^T$	$\underline{x}_{(0)}^{(1)} = [-0.955 \quad 0.99]^T$	3.969
			$\underline{x}_{(1)}^{(1)} = [-0.990 \quad 0.99]^T$	3.959
			$\underline{x}_{(2)}^{(1)} = [-0.984 \quad 0.979]^T$	3.948
2	$[0 \quad 1]^T$	$[0.453 \quad -0.892]^T$	$\underline{x}_{(0)}^{(2)} = [-0.761 \quad 0.54]^T$	3.257
			$\underline{x}_{(1)}^{(2)} = [-0.761 \quad 0.579]^T$	3.101
			$\underline{x}_{(2)}^{(2)} = [-0.702 \quad 0.462]^T$	2.986
3	$[0.453 \quad -0.892]^T$	$[0.608 \quad -0.794]^T$	$\underline{x}_{(0)}^{(3)} = [-0.503 \quad 0.203]^T$	2.510
			$\underline{x}_{(1)}^{(3)} = [-0.538 \quad 0.273]^T$	2.396
			$\underline{x}_{(2)}^{(3)} = [-0.466 \quad 0.178]^T$	2.301

تعداد فراخوانی تابع = $1562 \quad \rightarrow$ (تقسیم بر 3) تعداد تکرار $\approx 500 \quad \rightarrow$

$\underline{x}^* = [1.0000 \quad 1.0000]^T \quad , \quad f(\underline{x}^*) = 1.34 \times 10^{-16} \cong 0$

مراجع

[1]. Dixon & James, 1980

[2]. Box & Hunter, 1962

[3]. Hook, R. ; Jeeves, T.A.; “Direct Search Solution of Numerical and Statistical Problems”, J. ass. Comp. Mach., 8. 212-229, 1961.

[4]. Spendy, W.; Hext, G.R.; Himsworth, F.R.; “Sequential Applications of Simplex Designs in Optimization and Evolutionary Operation”, Technometrics, 4., 441-461, 1962.

[5]. Nelder, J.A.; Mead, R.; “A Simplex Method for Function Minimization”, The Comp. Journal, 7, 308-313, 1965.

[6]. Powell, M.J.D.; “An Iterative Method for Finding Stationary Values of a Function of Several Variables”, The Comp. Journal, 5, 147-151, 1962.