دسترسی در سایتِ http://jnrm.srbiau.ac.ir
سال پنجم، شماره بیست و یکم، آذر و دی ۱۳۹۸
شماره شاپا: ۲۵۸۸-۵۸۸X

JNRM
پژوهش‌های نوین در ریاضی

دانشگاه آزاد اسلامی، واحد علوم و تحقیقات

تصحیح فرمول جستجوی خطی در روش BFGS برای رسیدن به همگرایی سراسری

منصوره حمزه نژاد[۱]، سید علیرضا حسینی دهمیری[۲*]

[۱و۲] گروه ریاضی، دانشگاه ولی‌عصر(عج) رفسنجان، رفسنجان، ایران

تاریخ ارسال مقاله: ۹۶/۱۱/۲۹   تاریخ پذیرش مقاله: ۹۷/۱۲/۰۸

چکیده

مسائل برنامه‌ریزی غیرخطی در گروه مسائل پرکاربرد بهینه‌سازی در دنیای واقعی قرار دارند. تابع هدف این گونه از مسائل، علاوه بر غیرخطی بودن، در بیشتر موارد غیرمحدب است. این در حالی است که برای تضمین همگرایی سراسری در الگوریتم‌هایی که بر اساس روش نیوتن برای حل این مسائل پیشنهاد شده‌اند، عموماً شرط تحدب الزامی است. در این بین روش‌های شبه نیوتن بدلیل استفاده از تقریب ماتریس هسی یا وارون آن دارای محبوبیت بیشتری هستند. هر چند که در این الگوریتم‌ها برای تقریب این ماتریس فقط از اطلاعات گرادیان استفاده می‌شود. یکی از کاربردی‌ترین الگوریتم‌های شبه‌نیوتون در حل مسایل برنامه‌ریزی غیرخطی روش BFGS می‌باشد. این مقاله یک ایده‌ی جدید برای جستجوی خطی در روش BFGS ارائه داده و ثابت می‌کند که استفاده از این تکنیک، همگرایی سراسری را برای مسائل کلی بدون نیاز به هیچ شرط اضافه‌ای به دنبال خواهد داشت. در نهایت، کارایی الگوریتم پیشنهاد شده به صورت عددی مورد ارزیابی قرار گرفته است.

واژه‌های کلیدی: روش BFGS، روش نیوتون، روش‌شبه‌نیوتون، همگرایی سراسری، بهینه‌سازی نامقید.

---

* عهده‌دار مکاتبات: Email: dehmiry@vru.ac.ir

۱. مقدمه

بسیاری از مسائل مهندسی امروزه در رده‌بندی مسائل برنامه‌ریزی غیرخطی قرار می‌گیرند. مسائلی که هدف آنها بهینه‌سازی یک تابع غیرخطی و عموماً غیرمحدب است. دستیابی به مجموعه جواب این مسائل با الگوریتم‌هایی که عموماً دارای پیچیدگی محاسباتی بالا هستند، امکان‌پذیر است. اکثر الگوریتم‌های موجود برای حل مسائل بهینه‌سازی غیرخطی، سعی می‌کنند توابع غیرخطی مساله را با تقریبی از آن‌ها جایگزین کنند که در این بین استفاده از بسط تیلور بیش از بقیه دارای محبوبیت می‌باشد. نقش ماتریس هسی در این‌گونه تقریب‌ها بسیار پر رنگ می‌باشد. به عنوان مثال، معین مثبت متقارن بودن ماتریس هسی در این تقریب تضمینی برای همگرایی روش‌های بر پایه گرادیان می‌باشد. در این مقاله نویسندگان به بررسی یکی از روش‌هایی که بر پایه روش نیوتون بنا شده است، می‌پردازند. این روش به نام روش BFGS شهرت دارد برای تشریح این روش به مقدماتی نیاز است که در ادامه به آن می‌پردازیم.

مسأله بهینه‌سازی تک‌هدفه به فرم

$$\min_{x\in\mathbb{R}^n} f(x) \qquad (۱)$$

که در آن $f:\mathbb{R}^n \to \mathbb{R}$ تابعی دوبار پیوسته مشتق‌پذیر است، را در نظر بگیرید.

اگر تابع $g := \nabla f$ در نظر گرفته شود، در این صورت یافتن صفرهای تابع g معادل با یافتن نقاط اکسترمم تابع f بوده و ماتریس ژاکوبین تابع g همان ماتریس هسی تابع f خواهد بود.

یکی از روش‌های کلاسیک برای حل مسأله (۱)، روش نیوتون است. همگرایی این روش منوط به محدب بودن تابع f می‌باشد. در صورتی که این شرط برقرار نباشد، روش‌های شبه نیوتون جایگزین مناسبی برای روش نیوتون خواهند بود. روش‌های شبه‌نیوتون به دنبال تولید تقریبی از ماتریس هسی با حفظ خاصیت معین مثبت متقارن هستند. در واقع روش‌های شبه‌نیوتون بر پایه روش نیوتون و با تقریبی از ماتریس هسی کار می‌کنند. این روش‌ها با یک تقریب اولیه از ماتریس هسی شروع و در هر گام برای بروزرسانی آن از بردار گرادیان استفاده می‌کنند. روش‌های بروزرسانی عمدتاً دارای تنوع بالایی بوده و هر کدام دارای معایب و محاسن خاص می‌باشند. از جمله روش‌های شبه‌نیوتون می‌توان به روش‌های DFP[۴]، BHHH[۵]، BFGS[۲] و L-BFGS[۱] اشاره کرد.

در روش‌هایی که بر پایه‌ی روش نیوتون می‌باشند در هر تکرار، جواب فعلی توسط یک جهت بهبود دهنده با طول گام مناسب اصلاح می‌شود. به بیان دیگر اگر x_k جواب تکرار kام روش باشد، نقطه بعدی توسط فرمول

$$x_{k+1} = x_k + \alpha_k d_k, \quad \mathrm{k} = 1,2, \ldots \qquad (۲)$$

به دست می‌آید که در آن α_k طول گام و d_k جهت حرکت می‌باشد. تنها تفاوت بین این‌گونه روش‌ها چگونگی محاسبه طول گام و جهت حرکت است. در ادامه به بررسی دقیق‌تر روش BFGS می‌پردازیم.

توجه داشته باشید که

برای سادگی فرمول‌ها به جای $f(x_k)$ و $g(x_k)$ به ترتیب از f_k و g_k استفاده شده‌است.

در روش BFGS تقریب ماتریس هسی در هر گام به صورت زیر بروزرسانی می‌شود:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^t B_k}{s_k^t B_k s_k} + \frac{y_k y_k^t}{s_k^t y_k}, \quad k = 1,2, \ldots, \qquad (۳)$$

که در آن $s_k = x_{k+1} - x_k = \alpha_k d_k$ و $y_k = g_{k+1} - g_k$. در این روش جهت حرکت $d_k = -B_k^{-1} g_k$ می‌باشد که در آن B_k^{-1} همان معکوس تقریب ماتریس هسی در گام kام است. به سادگی می‌توان دید که دنباله $\{B_k\}$ تولید شده به وسیله روش BFGS در معادله زیر صدق می‌کند:

$$B_{k+1} s_k = y_k. \qquad (۴)$$

برای تعیین طول گام α_k می‌توان به دو شیوه جستجوی خطی دقیق یا تقریبی عمل کرد. پاول نشان داد که اثبات همگرایی سراسری روش BFGS با استفاده از روش جستجوی خطی دقیق ولف امکان پذیر نیست[۷]. پس از آن نشان داده شد که با استفاده از هیچ کدام از روش‌های جستجوی خطی دقیق امکان

اثبات همگرایی سراسری روش BFGS، برای توابع کلی وجود ندارد[۸]. این امر باعث شد که برای پیدا کردن طول گام، به روش‌های جستجوی خطی تقریبی رو آورده شود. از جمله این روش‌ها می‌توان به روش‌های ولف–پاول، ولف–پاول قوی و ولف– پاول ضعیف اشاره کرد[۸].

اثبات همگرایی این روش‌ها نیز در حالت کلی برای مسائل بهینه‌سازی محدب امکان‌پذیر است.

اخیراً روش جستجوی خطی دیگری به‌نام ولف–پاول ضعیف اصلاح شده (MWWP) معرفی و همگرایی آن برای توابع کلی در یک حالت خاص اثبات شده است[۹]. در این روش برای محاسبه طول گام α_k از روابط زیر استفاده می‌شود:

$$f(x_k+\alpha_k d_k) \le f_k + \delta\alpha_k g_k^t d_k + \alpha_k \min\left\{-\delta_1 g_k^t d_k, \delta\frac{\alpha_k}{2}\|d_k\|^2\right\} \qquad (۵)$$

$$g(x_k+\alpha_k d_k)^t d_k \ge \sigma g_k^t d_k + \min\{-\delta_1 g_k^t d_k, \delta\alpha_k\|d_k\|^2\} \qquad (۶)$$

که در آن $\delta \in (0,\frac{1}{2})$، $\delta_1 \in (0,\delta)$ و $\sigma \in (\delta,1)$ پارامترهای مورد نیاز روش می‌باشند.

در این مقاله سعی شده است با تغییر کوچکی در معادلات (۵) و (۶) این همگرایی در حالت کلی اثبات گردد.

این تغییر در روابط (۵) و (۶) به صورت زیر خواهد بود:

$$f(x_k+\alpha_k d_k) \le f_k + \delta\alpha_k g_k^t d_k + \alpha_k \min\left\{-\delta_1 g_k^t d_k, \delta\frac{\alpha_k}{2k}\|d_k\|^2\right\} \qquad (۷)$$

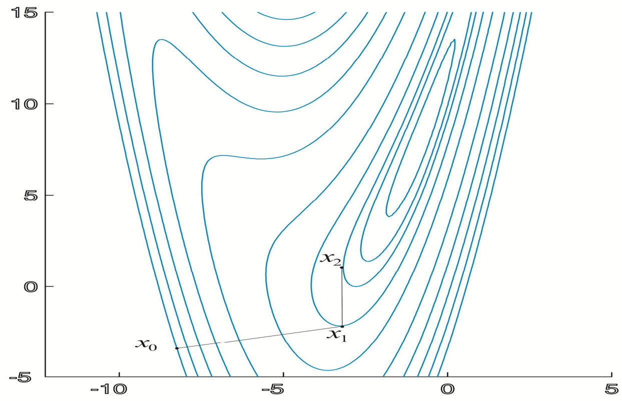
$$g(x_k+\alpha_k d_k)^t d_k \ge \sigma g_k^t d_k + \min\left\{-\delta_1 g_k^t d_k, \delta\frac{\alpha_k}{k}\|d_k\|^2\right\} \qquad (۸)$$

طرح کلی این مقاله به صورت زیر است: در بخش سوم به تصحیح الگوریتم ولف پاول ضعیف اصلاح شده پرداخته شده است. بخش چهارم به ارائه الگوریتم اصلاح شده برای روش BFGS می‌پردازد. بررسی برخی از نتایج اصلی در خصوص روش به بخش پنجم و اثبات همگرایی سراسری آن به بخش ۶ موکول شده است. در نهایت کارایی روش به صورت عددی مورد بررسی قرار گرفته است.

۲. تصحیح الگوریتم ولف پاول ضعیف اصلاح شده

مسأله بهینه‌سازی تک‌هدفه نامقید (۱) را در نظر بگیرید. برای رسیدن به یکی از نقاط بحرانی تابع f لازم است که یکی از صفرهای تابع g را بیابیم. فرض کنید B_1 یک تقریب اولیه از ماتریس هسی f باشد که معین مثبت متقارن است. برای راحتی کار می‌توان B_1 را ماتریس همانی در نظر گرفت. از نقطه‌ی دلخواه $x_1 \in \mathbb{R}^n$ شروع می‌کنیم. اگر فاصله‌ی این نقطه تا صفر تابع g از تلورانس در نظر گرفته شده ($\|g(x)\| \le \varepsilon \in (0,1)$) کمتر بود متوقف می‌شویم. نقطه‌ی فعلی همان جواب مطلوب ماست. در غیر این صورت بایستی با یک طول گام مناسب در یک جهت مناسب تا جایی حرکت کنیم که دیگر امکان بهبود تابع g در آن جهت نباشد. با توجه به روش نیوتن، جهت حرکت از این نقطه توسط فرمول $d_1 = -B_1^{-1} g_1$ بدست می‌آید. چون B_1 یک ماتریس معین مثبت متقارن است لذا $g_1^t d_1 = -d_1^t B_1 d_1 < 0$ خواهد بود. حال باید طول گام مناسب برای حرکت در امتداد این جهت را بیابیم. برای این منظور باید α_1 را به گونه‌ای تعیین کنیم که مقدار تابع f در نقطه‌ی جدید $x_1+\alpha_1 d_1$ کمتر از مقدار تابع f در نقطه x_1 بوده و همچنین تابع g به ریشه‌ی خود تا جای امکان نزدیک شود.

نقطه‌ی جدید را x_2 نامیده و بعد از بروزرسانی ماتریس هسی روند فوق را با این نقطه تکرار می‌کنیم. این پروسه در شکل زیر به تصویر کشیده شده است.



شکل ۱: روند حرکت الگوریتم BFGS اصلاح شده

فرض کنید تابع $\varphi(\alpha)$ را به صورت زیر تعریف شده باشد:

$$\varphi(\alpha) := f(x_k+\alpha d_k) - f_k - \delta\, \alpha {g_k}^t d_k - \min\{-\delta_1 g_k^t d_k\,, \delta\frac{\alpha}{2k}\|d_k\|^2\}.$$

بنا به توضیحات داده شده الگوریتم این روند تکراری را می‌توان به صورت زیر نوشت.

۳. الگوریتم BFGS اصلاح شده

گام ۱) نقطه شروع x_1 و ثابت‌های $\varepsilon، \in (0,1)$ $\delta \in (0,\frac{1}{2})$ $\delta_1 \in (0,\delta)$ و $\sigma\in(\delta,1)$ و ماتریس اولیه معین مثبت متقارن B_1 را انتخاب کن و قرار بده $k := 1$.

گام ۲) اگر $\|g_k\| \le \varepsilon$ بود، متوقف شو.

گام ۳) معادله $B_k d_k + g_k = 0$ را برای محاسبه جهت جستجوی d_k حل کن.

گام ۴) طول گام α_k را با استفاده از الگوریتم A محاسبه کن.

گام ۵) $x_{k+1} = x_k + \alpha_k d_k$ قرار بده و B_{k+1} را با استفاده از فرمول (۳) بروز رسانی کن.

گام ۶) $k:= k+1$ و به گام ۲ برگرد.

الگوریتم A:

(ارزش‌دهی آغازین) $\alpha_0 = 0$، $\alpha_u > 0$، $i := 1$ و $\alpha_1 \in (0,\alpha_u)$ را در نظر بگیر.

(حلقه اصلی)

الف) مقدار $\varphi(\alpha_i)$ را محاسبه کن.

ب) اگر $\varphi(\alpha_i) > 0$

۱) طول گام دلخواه اولیه α را بین α_i و α_{i-1} در نظر بگیر.

۲) مقدار $\varphi(\alpha)$ را محاسبه کن.

۳) اگر $\varphi(\alpha) > 0$ یا $\varphi(\alpha) \ge \varphi(\alpha_{i-1})$ در اینصورت $\alpha_i := \alpha$ به مرحله ۱ برگرد.

در غیر اینصورت

۳-الف) مقدار $\varphi'_+(\alpha)$ را محاسبه کن.

۳-ب) اگر $\varphi'_+(\alpha) \ge 0$ آنگاه α همان خروجی مورد نظر است و متوقف شو.

۳-ج) اگر $\varphi'_+(\alpha)(\alpha_i - \alpha_{i-1}) \ge 0$ آنگاه $\alpha_{i-1} := \alpha_i$.

۳-د) $\alpha_i := \alpha$ و به مرحله ۱ برگرد.

ج) مقدار $\varphi'_+(\alpha_i)$ را محاسبه کن اگر $\varphi'_+(\alpha_i) \ge 0$ آنگاه α_i خروجی مورد نظر است و متوقف شو.

د) $\alpha_{i+1} \in (\alpha_i,\alpha_u)$ را انتخاب کن.

ه) $i:= i+1$ و به گام الف برگرد.

اگر یکی از شرایط زیر برقرار باشد، بازه‌ی کراندار تولید شده توسط پارامترهای α_{i-1} و α_i در الگوریتم A شامل طول گام مناسب خواهد بود که در شرایط (۷) و (۸) صدق می‌کند.

- α_i شرط (۷) را نقض کند؛
- $\varphi(\alpha_i) \ge \varphi(\alpha_{i-1})$؛

- $\varphi'_{+}(\alpha_i) \geq 0$.

در هر تکرار از الگوریتم A طول گام α بین α_{i-1} و α_i انتخاب و یکی از این نقاط ابتدایی یا انتهایی با α جایگزین می‌شود به طوری که همچنان سه خاصیت زیر حفظ شود:

- بازه کراندار تولید شده توسط پارامترهای α_{i-1} و α_i شامل طول گامی باشد که در شرایط (۷) و (۸) صدق کند؛
- α_{i-1} بیشترین فاصله را تا طول گام مورد نظر داشته باشد و کمترین مقدار تابع در این نقطه اتفاق بیفتد؛
- α_i به‌گونه‌ای انتخاب شود که $\varphi'_{+}(\alpha_{i-1})(\alpha_i - \alpha_{i-1}) < 0$ باشد.

۴. برخی از نتایج اصلی در خصوص الگوریتم BFGS اصلاح شده

در ابتدا نشان می‌دهیم که طول گام $\alpha_k>0$ موجود است که در روابط (۷) و (۸) صدق می‌کند.

قضیه ۱: فرض کنیم f تابعی دوبار پیوسته مشتق‌پذیر و از پایین کراندار باشد. اگر $g_k^t d_k \leq 0$ باشد آنگاه ثابت $\alpha_k \in [\alpha_l, \alpha_u]$ با شرط $0 < \alpha_l < \alpha_u$ وجود دارد که در روابط (۷) و (۸) صدق می‌کند.

اثبات. با استفاده از فرض از پایین کراندار بودن f_k و $f(x_k + \alpha_k d_k)$ و اینکه $\delta_1 < \delta$ و $g_k^t d_k \leq 0$، می‌توان گفت که اگر $\alpha \to +\infty$ آنگاه

$$\min\{-\delta_1 g_k^t d_k \,, \delta \frac{\alpha}{2k} \|d_k\|^2\} = -\delta_1 g_k^t d_k$$

و $\varphi(\alpha)$ به سمت بی‌نهایت میل می‌کند. واضح است که $\varphi(0) = 0$. حال عدد مثبت α وجود دارد که

$$\begin{aligned}
&\varphi(\alpha) \\
&= f(x_k + \alpha d_k) - f_k - \delta\, \alpha {g_k}^t d_k \\
&- \alpha \min\{-\delta_1 g_k^t d_k \,, \delta \frac{\alpha}{2k} \|d_k\|^2\} \\
&= \left(f_k + \alpha {g_k}^t d_k + o(\alpha)\right) - f_k \\
&- \delta\, \alpha {g_k}^t d_k \\
&- \alpha \min\left\{-\delta_1 g_k^t d_k \,, \delta \frac{\alpha}{2k} \|d_k\|^2\right\} \\
&= \alpha(1 - \delta + \delta_1) g_k^t d_k + o(\alpha) < 0.
\end{aligned}$$

چون φ یک تابع پیوسته است لذا ثابت $\rho > 0$ وجود دارد که $\varphi(\rho) = 0$ است. فرض کنیم $\tilde{\alpha}$ کمینه سراسری $\varphi(\alpha)$ در بازه $[0, \rho]$ باشد.

کمینه سراسری نمی‌تواند در نقاط مرزی اتفاق بیفتد، چون همانطور که دیدیم مقدار تابع φ در نقاط مرزی برابر صفر است در حالی که نشان دادیم $\alpha \in (0, \rho)$ وجود دارد که $\varphi(\alpha) < 0$. بنابراین می‌توان نتیجه گرفت که حداقل یک کمینه موضعی $\tilde{\alpha} \in (0, \rho)$ وجود دارد که $\varphi(\tilde{\alpha}) < 0$ و $\varphi'_{+}(\tilde{\alpha}) \geq 0$. در این صورت

$$\begin{aligned}
&\varphi(\tilde{\alpha}) \\
&= f(x_k + \tilde{\alpha} d_k) - f_k - \delta \tilde{\alpha} {g_k}^t d_k \\
&- \min\left\{-\delta_1 g_k^t d_k \,, \delta \frac{\tilde{\alpha}}{2k} \|d_k\|^2\right\} < 0.
\end{aligned}$$

در نتیجه

$$\begin{aligned}
&f(x_k + \tilde{\alpha} d_k) \\
&< f_k + \delta \tilde{\alpha} {g_k}^t d_k \\
&+ \min\{-\delta_1 g_k^t d_k \,, \delta \frac{\tilde{\alpha}}{2k} \|d_k\|^2\}.
\end{aligned}$$

بنابراین با انتخاب $\alpha = \tilde{\alpha}$، رابطه (۷) برقرار است. علاوه براین

$$\varphi'_{+}(\tilde{\alpha}) = \lim_{\alpha \to (\tilde{\alpha})^+} \frac{\varphi(\alpha) - \varphi(\tilde{\alpha})}{\alpha - \tilde{\alpha}}$$

$$\begin{aligned}
&= \lim_{\alpha \to (\tilde{\alpha})^+} \{\frac{f(x_k + \alpha d_k) - f_k - \delta \alpha {g_k}^t d_k - \alpha \min\left\{-\delta_1 g_k^t d_k \,, \delta \frac{\alpha}{2k} \|d_k\|^2\right\}}{\alpha - \tilde{\alpha}} \\
&- \frac{f(x_k + \tilde{\alpha} d_k) - f_k - \delta \tilde{\alpha} {g_k}^t d_k - \tilde{\alpha} \min\{-\delta_1 g_k^t d_k \,, \delta \frac{\tilde{\alpha}}{2k} \|d_k\|^2\}}{\alpha - \tilde{\alpha}}\}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{\alpha \to (\tilde{\alpha})^+} \{\frac{f(x_k + \alpha d_k) - f(x_k + \tilde{\alpha} d_k)}{\alpha - \tilde{\alpha}} \\
&- \frac{\delta(\alpha - \tilde{\alpha}) {g_k}^t d_k}{\alpha - \tilde{\alpha}} \\
&- \frac{\min\{-\delta_1 g_k^t d_k \,, \delta \frac{\alpha^2}{2k} \|d_k\|^2\}}{\alpha - \tilde{\alpha}} \\
&+ \frac{\min\{-\delta_1 g_k^t d_k \,, \delta \frac{\tilde{\alpha}^2}{2k} \|d_k\|^2\}}{\alpha - \tilde{\alpha}}\}
\end{aligned}$$

حال چون $\alpha \to (\tilde{\alpha})^+$، پس مینیمم در مؤلفه یکسانی از هر دو جمله رخ خواهد داد. بنابراین:

$$\begin{aligned}
\varphi'_{+}(\tilde{\alpha}) = \lim_{\alpha \to (\tilde{\alpha})^+} \{&\frac{f(x_k + \alpha d_k) - f(x_k + \tilde{\alpha} d_k)}{\alpha - \tilde{\alpha}} \\
&- \frac{\delta(\alpha - \tilde{\alpha}) {g_k}^t d_k}{\alpha - \tilde{\alpha}} \\
- \frac{\min\{-\delta_1 (\alpha - \tilde{\alpha}) g_k^t d_k \,, \delta \frac{(\alpha^2 - \tilde{\alpha}^2)}{2k} \|d_k\|^2\}}{\alpha - \tilde{\alpha}}\}&
\end{aligned}$$

$$\begin{aligned}
&= \lim_{\alpha\to(\tilde{\alpha})^+}\{\frac{(\alpha-\tilde{\alpha})g(x_k+\tilde{\alpha}d_k)^t d_k+o((\alpha-\tilde{\alpha})^2)}{\alpha-\tilde{\alpha}}\\
&-\delta\, g_k{}^t d_k\\
&-\min\{-\delta_1 g_k^t d_k\,,\delta\frac{\alpha+\tilde{\alpha}}{2k}\|d_k\|^2\}\}\\
&= \lim_{\alpha\to(\tilde{\alpha})^+}\{\frac{(\alpha-\tilde{\alpha})g(x_k+\tilde{\alpha}d_k)^t d_k+o((\alpha-\tilde{\alpha})^2)}{\alpha-\tilde{\alpha}}\\
&-\delta\, g_k{}^t d_k\\
&-\min\{-\delta_1 g_k^t d_k\,,\delta\frac{\tilde{\alpha}}{k}\|d_k\|^2\}\}\\
&= g(x_k+\tilde{\alpha}d_k)^t d_k-\delta\, g_k{}^t d_k-\\
&\min\{-\delta_1 g_k^t d_k\,,\delta\tfrac{\tilde{\alpha}}{k}\|d_k\|^2\}\}.
\end{aligned}$$

حال از اینکه $\varphi'_+(\tilde{\alpha})\geq 0$ داریم:

$$\begin{aligned}
&\varphi'_+(\tilde{\alpha})\\
&= g(x_k+\tilde{\alpha}d_k)^t d_k-\delta\, g_k{}^t d_k\\
&-\min\{-\delta_1 g_k^t d_k\,,\delta\frac{\tilde{\alpha}}{k}\|d_k\|^2\}\}\geq 0.
\end{aligned}$$

پس

$$\begin{aligned}
&g(x_k+\tilde{\alpha}d_k)^t d_k\\
&\geq\ \delta\, g_k{}^t d_k+\min\{-\delta_1 g_k^t d_k\,,\delta\frac{\tilde{\alpha}}{k}\|d_k\|^2\}\\
&\geq\ \sigma\, g_k{}^t d_k+\min\{-\delta_1 g_k^t d_k\,,\delta\frac{\tilde{\alpha}}{k}\|d_k\|^2\}.
\end{aligned}$$

بنابراین با انتخاب $\alpha=\tilde{\alpha}$ رابطه (۸) نیز برقرار است.

۵. بررسی همگرایی سراسری روش BFGS اصلاح شده

ابتدا برخی ازمفروضات اساسی را که در ادامه نیاز داریم را در نظر می‌گیریم:

۱. فرض کنیم مجموعه تراز $L_0=\{x\in\mathbb{R}^n: f(x)\leq f(x_1)\}$ کراندار باشد.

۲. فرض کنیم f تابعی دوبار پیوسته مشتق‌پذیر و از پایین کراندار باشد. و تابع گرادیان g در شرط پیوستگی لیپشیتس با ثابت لیپشیتس L صدق کند، یعنی $\|g(x)-g(y)\|\leq L\|x-y\|$.

لم زیر نشان می‌دهد که دنباله $\{B_k\}$ حاصل از فرمول بروزرسانی(۳) شرط معین مثبت بودن را داراست.

لم ۱. فرض کنیم $\{B_k\}$ دنباله تولید شده به وسیله فرمول بروزرسانی(۳) باشد. در این صورت به ازای هر k، B_k معین مثبت خواهد بود.

اثبات. برای اثبات کافی است که ثابت کنیم شرط $s_k^t y_k>0$ به ازای هر k برقرار است. اثبات به وسیله استقرا انجام می‌شود. برای $k=1$، فرض کردیم B_1 یک ماتریس معین مثبت باشد. با استفاده از رابطه (۸) و معادله $B_k d_k+g_k=0$ داریم:

$$\begin{aligned}
&s_k^t y_k\\
&=(g(x_k+\alpha_k d_k)-g_k)^t(x_k+\alpha_k d_k\\
&-x_k)\ =\alpha_k g(x_k+\alpha_k d_k)^t d_k-\alpha_k g_k^t d_k\\
&\geq\alpha_k\sigma g_k^t d_k\\
&+\alpha_k\min\left\{-\delta_1 g_k^t d_k\,,\delta\frac{\alpha_k}{k}\|d_k\|^2\right\}-\alpha_k g_k^t\\
&=\ \ -\alpha_k(1-\sigma)g_k^t d_k\\
&+\alpha_k\min\left\{-\delta_1 g_k^t d_k\,,\delta\frac{\alpha_k}{k}\|d_k\|^2\right\}\\
&\geq-\alpha_k(1-\sigma)g_k^t d_k>0\,.
\end{aligned}$$

بنابراین معین مثبت بودن ماتریس B_{k+1} در رابطه (۳) حاصل می‌شود.

نتیجه ۱. از لم ۱و همچنین رابطه (۳) می‌توان نتیجه گرفت که برای هر $k\geq 1$، $g_k^t d_k<0$.

لم ۲. فرض کنیم $\{x_k,\alpha_k,d_k,g_k\}$ توسط الگوریتم ۱۴ تولید شده باشند. همچنین مفروضات اساسی بیان شده در ابتدای این بخش نیز برقرار باشند. در این صورت

$$\sum_{k=1}^{\infty}\frac{(g_k^t d_k)^2}{\|d_k\|^2}<\infty\,.$$

اثبات. با استفاده از رابطه (۷) داریم:

$$\begin{aligned}
&f_k-f(x_k+\alpha_k d_k)\\
&\geq-\delta\alpha_k g_k^t d_k\\
&-\alpha_k\min\left\{-\delta_1 g_k^t d_k\,,\delta\frac{\alpha_k}{2k}\|d_k\|^2\right\}\\
&\geq-(\delta-\delta_1)\alpha_k g_k^t d_k.
\end{aligned}$$

با گرفتن سیگما از دو طرف نامساوی و با استفاده از فرض ۲ داریم:

$$\begin{aligned}
\infty>f_1-\lim_{k\to\infty}f_k&=\sum_{k=1}^{\infty}\big(f_k\\
&\quad-f(x_k+\alpha_k d_k)\big)\\
&\geq\sum_{k=1}^{\infty}-(\delta-\delta_1)\alpha_k g_k^t d_k.
\end{aligned}$$

فرض ۲ وجود حد فوق را تضمین می‌کند. حال از روابط (۸)، پیوستگی لیپشیتس تابع g، نتیجه ۱و نامساوی کشی شوارتز داریم:

$$\alpha_k L\|d_k\|^2 \geq \alpha_k\left(\frac{\|g_{k+1}-g_k\|}{\|\alpha_k d_k\|}\right)\|d_k\|^2 = \|g_{k+1}-g_k\|\|d_k\| \geq |(g_{k+1}-g_k)^t d_k| \geq (g_{k+1}-g_k)^t d_k = \sigma g_k^t d_k + \min\left\{-\delta_1 g_k^t d_k\ ,\delta\frac{\alpha_k}{k}\|d_k\|^2\right\} - g_k^t d_k \geq -(\sigma-\delta)g_k^t d_k + \min\left\{-\delta_1 g_k^t d_k\ ,\delta\frac{\alpha_k}{k}\|d_k\|^2\right\} \geq (\sigma-1)g_k^t d_k.$$

بنابراین $\alpha_k \geq \frac{(\sigma-1)g_k^t d_k}{L\|d_k\|^2}$ که با جایگذاری در رابطه بالا نتیجه می‌شود:

$$\frac{(\delta-\delta_1)(1-\sigma)}{L}\sum_{k=1}^{\infty}\frac{(g_k^t d_k)^2}{\|d_k\|^2} < \infty$$

و حکم برقرار است.

نتیجه۲. با استفاده از نتیجه ۱و همچنین رابطه (۷) دنباله $\{f_k\}$ موجود یک دنباله نزولی است.

اثبات. برای اثبات دو حالت را در نظر می‌گیریم:

۱. حالتی که $\min\{-\delta_1 g_k^t d_k,\ \delta\frac{\alpha_k}{2k}\|d_k\|^2\} = -\delta_1 g_k^t d_k$؛

۲. حالتی که $\min\{\delta_1 g_k^t d_k,\ \delta\frac{\alpha_k}{2k}\|d_k\|^2\} = \delta\frac{\alpha_k}{2k}\|d_k\|^2$.

در حالت اول با استفاده از رابطه (۷) داریم:

$$f(x_k+\alpha_k d_k) \leq f_k + \delta\alpha_k g_k^t d_k - \alpha_k\delta_1 g_k^t d_k \leq f_k + \alpha_k(\delta-\delta_1)g_k^t d_k$$

چون $g_k^t d_k \leq 0$ و $(\delta-\delta_1) > 0$ لذا حکم برقرار است.

برای حالت دوم نیز از رابطه (۷) داریم:

$$f(x_k+\alpha_k d_k) \leq f_k + \delta\alpha_k g_k^t d_k + \delta\frac{\alpha_k}{2k}\|d_k\|^2 \leq f_k + \delta\alpha_k(g_k^t d_k + \frac{1}{2k}\|d_k\|^2) \leq f_k$$

و حکم برقرار است.

ملاحظه ۱. برای اثبات همگرایی سراسری الگوریتم اصلاح شده، کافیست جستجوی خطی مورد نظر را برای حالتی که

$$\min\left\{-\delta_1 g_k^t d_k\ ,\delta\frac{\alpha_k}{k}\|d_k\|^2\right\} = \delta\frac{\alpha_k}{k}\|d_k\|^2,$$

در نظر بگیریم. چون وقتی $k\to\infty$ مینیمم در مؤلفه دوم رخ خواهد داد.

لم ۳. فرض کنیم ملاحظه ۱ و همچنین فرض ۲ برقرار باشد. در این صورت

$$s_k^t y_k \geq \delta\|s_k\|^2, \qquad \frac{\|y_k\|^2}{s_k^t y_k} \leq \frac{L^2}{\delta}.$$

اثبات. با استفاده از فرض ۲ داریم:

$$\|y_k\|^2 = \|g_{k+1}-g_k\|^2 < L^2\|s_k\|^2$$

از ملاحظه ۱و همچنین رابطه (۸) نتیجه می‌شود:

$$s_k^t y_k = (g_{k+1}-g_k)^t\alpha_k d_k \geq \alpha_k g_{k+1}^t d_k - \alpha_k g_k^t d_k \geq \alpha_k\sigma g_k^t d_k + \alpha_k\min\left\{-\delta_1 g_k^t d_k\ ,\delta\frac{\alpha_k}{k}\|d_k\|^2\right\} - \alpha_k g_k^t d_k$$
$$\geq (\sigma-1)g_k^t\alpha_k d_k + \alpha_k\min\left\{-\delta_1 g_k^t d_k\ ,\delta\frac{\alpha_k}{k}\|d_k\|^2\right\} = (\sigma-1)g_k^t s_k + \frac{\delta}{k}\|s_k\|^2, \quad \forall\, k\geq 1$$

که نامساوی آخر از $g_k^t s_k \leq 0$ نتیجه می‌شود. بنابراین داریم:

$$\frac{s_k^t y_k}{\|s_k\|^2} \geq \max\left\{\frac{\delta}{k}, k\geq 1\right\} = \delta\ ,\ \|y_k\|^2 \leq L^2\|s_k\|^2 \leq \frac{L^2 s_k^t y_k}{\delta} \Longrightarrow \frac{\|y_k\|^2}{s_k^t y_k} \leq \frac{L^2}{\delta}$$

قضیه۲. ([۳]) فرض کنیم دنباله $\{B_k\}$ همان دنباله تولید شده به وسیله روش BFGS باشد. و B_1 یک ماتریس معین مثبت متقارن و شرایط لم ۳ برقرار باشد. آنگاه ثابت‌های $b_1 \geq b_2 > 0$ وجود دارند به طوری که روابط

$$\|B_k s_k\| \leq b_1\|s_k\|, \qquad (۹)$$
$$s_k^t B_k s_k \geq b_2\|s_k\|^2 \qquad (۱۰)$$

برای حداقل $[\frac{t}{2}]$ مقادیر $k \in \{1,2,\dots,t\}$، برقرار هستند که در آن t عدد صحیح دلخواهی است.

قضیه ۳. فرض کنیم شرایط لم ۳ برقرار باشد. در این صورت:

$$\lim_{k\to\infty} inf\, \|g_k\| = 0.$$

که معادل با همگرایی سراسری الگوریتم اصلاح شده می‌باشد.

اثبات. از لم ۲ نتیجه می‌شودکه هرگاه $k \to \infty$ در این صورت جمله عمومی سری، یعنی $\frac{(g_k^t d_k)^2}{\|d_k\|^2}$، به صفر میل می‌کند. حال از رابطه $B_k d_k + g_k = 0$ داریم:

$$\frac{(g_k^t d_k)^2}{\|d_k\|^2} = \frac{((-B_k d_k)^t d_k)^2}{\|d_k\|^2} = \frac{({d_k}^t B_k d_k)^2}{\|d_k\|^2} \to 0,$$

و همچنین از ترکیب روابط (۹) و (۱۰) هرگاه $k \to \infty$ در این صورت:

$$b_2^2\|\alpha_k d_k\|^4 \le \left((\alpha_k d_k)^t B_k(\alpha_k d_k)\right)^2 \Rightarrow$$
$$0 \le b_2^2\|d_k\|^2 \le \frac{\left({d_k}^t B_k d_k\right)^2}{\|d_k\|^2} \to 0.$$

بنابراین هرگاه $k \to \infty$، آنگاه $\|d_k\| \to 0$. لذا از $g_k = -B_k d_k$ و لم ۳ نتیجه می‌شود:

$$b_2\|d_k\|^2 \le \frac{{d_k}^t B_k d_k}{\|d_k\|} = \frac{-d_k^t g_k}{\|d_k\|} \le \frac{\|d_k\|\|g_k\|}{\|d_k\|} \Rightarrow b_2\|d_k\| \le \|g_k\|.$$

از طرفی

$$\|g_k\| = \|-B_k d_k\| \le b_1\|d_k\|.$$

لذا

$$b_2\|d_k\| \le \|g_k\| \le b_1\|d_k\|.$$

چون این رابطه تنها به ازای حداقل $[\frac{t}{2}]$ مقادیر $k \in \{1,2,\dots,t\}$، برقرار است، لذا زیردنباله‌ای از دنباله $\{\|g_k\|\}$ وجود دارد که همگرا به صفر است. بنابراین حکم برقرار است.

مثال. مسائل زیر را در نظر بگیرید:

$P1)\quad \min f(x,y) = 10(y-x^2)^2 + (x-1)^2$

$P2)\quad \min f(x,y) = x^4 - 6x^2 + 4y^2 + 12$

الگوریتم‌های BFGS-WWP، BFGS-MWWP و الگوریتم پیشنهادی با پارامترهای ورودی زیر برای این دو مساله اجرا گردید. همانگونه که انتظار می‌رفت نتایج هر سه الگوریتم برای این مسائل یکسان می‌باشد.

$$\varepsilon = 0.01, \delta = \frac{1}{3}, \sigma = \frac{2}{3}.$$

خروجی الگوریتم پیشنهادی برای این دو مساله در جدول ۱ آورده شده است.

مساله P1				مساله P2			
k	x_k	f_k	$\|\|g_k\|\|$	k	x_k	f_k	$\|\|g_k\|\|$
۱	(۰ ,۰)	۱	۲	۱	(−۱,۱)	۱۱	۱۱/۳۱
۲	(۰/۲۵,۰)	۰/۶۰	۱/۵۲	۲	(−۲ ,۰)	۴	۸
۳	(۰/۵۸,۰/۲۷)	۰/۲۱	۱/۳۴	۳	(−۱/۷۳,−۰/۴۰)	۳/۶۴	۳/۲۲
۴	(۰/۵۹,۰/۳۱)	۰/۱۸	۰/۸۴	۴	(−۱/۷۳,−۰/۳۳)	۳/۴۴	۲/۶۸
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۲۵	(۰/۹۴,۰/۸۹)	۰/۰۰۳۱	۰/۱۰	۲۵	(−۱/۷۳,−۰/۰۲)	۳/۰۰۲	۰/۱۶
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۴۴	(۰/۹۹,۰/۹۹)	۰/۰۰۰۰۱	۰/۰۰۹	۴۶	(−۱/۷۳,۰)	۳	۰/۰۰۹

جدول ۱: خروجی الگوریتم تصحیح شده برای مسائل P1 وP2.

نتیجه‌گیری

این مقاله فرمول جستجوی خطی در روش BFGS را به گونه‌ای تصحیح می‌نماید که به واسطه‌ی آن بتوان همگرایی سراسری این روش را برای مسائل کلی اثبات کرد. تصحیح انجام شده سرعت همگرایی را تغییر نمی‌دهد ولی در عوض شرایط مورد نیاز برای اثبات تئوری همگرایی سراسری را ایجاب می‌کند. اصلاح دوباره‌ی این فرمول برای رسیدن به سرعت همگرایی بالاتر، می‌تواند یک ایده مناسب برای مقالات آتی باشد.

فهرست منابع

[1] Branke, J., et al. (2005). Practical Approaches to Multi-Objective Optimization, Internationales Begegnungs- und Forschungszentrum (IBFI).

[2] Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations.IMA Journal of Applied Mathematics,6(1): 76-90.

[3] Byrd, R. H. and J. Nocedal (1989). A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. SIAM Journal on Numerical Analysis, 26(3): 727-739.

[4] Davidon, W. C. (1991). Variable metric method for minimization. SIAM Journal on Optimization, 1(1): 1-17.

[5] Henningsen, A. and O. Toomet (2011). maxLik: A package for maximum likelihood estimation in R.Computational Statistics,26(3): 443-458.

[6] Powell, M. J. (1976). Some global convergence properties of a variable metric algorithm for minimization without exact line searches. Nonlinear programming, 9(1): 53-72.

[7] Mascarenhas, W. F. (2004). The BFGS method with exact line searches fails for non-convex objective functions.Mathematical Programming, 99(1): 49-61.

[8] Nocedal, J. and S. Wright (2006). Numericaloptimization, Springer Science & Business Media.

[9] G. Yuan, Z. Sheng, B. Wang, W. Hu, and C. Li (2018). The global convergence of a modified BFGS method for nonconvex functions, Journal of Computational and Applied Mathematics, 327: 274–294.