

روش بهینه‌سازی نیوتن در فضاهای ژئودزیک

بیژن احمدی کاکاوندی

چکیده

بسیاری از فضاهای مهم در مسائل مربوط به علوم داده یا مهندسی، فضاهای غیرخطی‌اند. از این رو، در سال‌های اخیر روش‌های عددی برای بهینه‌سازی تابع‌های تعریف‌شده روی خمینه‌های ریمانی، مورد توجه و پژوهش بسیاری قرار گرفته است. از طرف دیگر، هندسه‌دانانی چون آلیکساندر ف و گرومف، با ابداع فضاهای ژئودزیک، دریچه تازه‌ای به مطالعه اشیاء هندسی گشودند. این فضاها تعمیم خمینه‌های ریمانی‌اند و، علاوه بر مزیت‌های دیگر، فاقد پیچیدگی‌های تانسوری این خمینه‌ها هستند. این فضاها بسیاری از اشیاء ناهموار ریاضی، از جمله گراف‌ها یا خمینه‌های توپولوژیک را هم شامل می‌شوند. در این مقاله، روش نیوتن برای یافتن نقطه مینیم یک تابع خودسازگار روی فضاهای متری ژئودزیک ارائه می‌شود. از مزیت‌های مهم این نوع بررسی در قیاس با روش نیوتن روی خمینه‌های ریمانی، سادگی بسیار نظریه و کاهش حجم محاسبات است. علی‌رغم نبود ساختار هموار و جبر تانسوری روی خمینه‌ها، صرفاً با استفاده از مفهوم «خم ژئودزی» نشان می‌دهیم که می‌توان روش نیوتن را به شکلی موفق و حتی ساده‌تر روی رده وسیعی از ساختارهای معمول ریاضی طراحی و اجرا کرد.

۱. مقدمه

امروزه مسائل بهینه‌سازی از جمله مهم‌ترین و پردامنه‌ترین مسائل در ریاضی کاربردی، آمار، علوم داده، و یادگیری ماشین است. با اینکه بهینه‌سازی تاریخ نسبتاً طولانی‌ای دارد و نخستین پژوهش‌های

عبارات و کلمات کلیدی. بهینه‌سازی روی خمینه‌ها، فضای ژئودزیک، روش نیوتن، تابع خودسازگار.

این حوزه به ریاضی‌دانان بزرگی مانند نیوتن، فرما، و لاگرانژ می‌رسد، در نیم‌قرن اخیر و همگام با گسترش ریاضی کاربردی مدرن، به یکی از پایه‌های اصلی در این زمینه‌ها تبدیل شده است. به‌طورکلی می‌توان یک مسئله بهینه‌سازی را به‌شکل زیر صورت‌بندی کرد

$$\min_{x \in D} f(x) \quad (1.1)$$

که در آن، معمولاً، D یک زیرمجموعه از فضای اقلیدسی با بُعد متناهی یا (به‌ندرت) از یک فضای برداری توپولوژیک با بُعد نامتناهی X و f یک تابع حقیقی‌مقدار روی D است. یکی از وضعیت‌های نسبتاً خوش‌رفتار - که خوشبختانه در عمل هم بسیار پیش می‌آید - زمانی است که D زیرمجموعه‌ای محدب و f تابعی محدب باشد. خواننده علاقه‌مند می‌تواند، مثلاً، در کتاب [۷] شرح مفصلی درباره «نظریه بهینه‌سازی محدب» از جنبه نظری و عملی و همچنین ارائه الگوریتم‌های عددی بیابد. همچنین در کتاب [۲۴] می‌توان کاربردهای مهمی از بهینه‌سازی در نظریه یادگیری ماشین را دید.

یکی از روش‌های مهم برای حل عددی مسئله بهینه‌سازی، روش نیوتن است. روش نیوتن یک روش عددی گام‌به‌گام و مبتنی بر این ایده است که به جای یافتن جواب مستقیم مسئله (۱.۱)، می‌توان صفر تابع گرادیان ∇f را تقریب زد. برای این منظور، ضروری است که تابع f دوبار مشتق‌پذیر باشد؛ به‌علاوه باید ماتریس مشتقات مرتبه دوم آن، یعنی ماتریس هسهای $\nabla^2 f$ ، دست‌کم در یک همسایگی از جواب مسئله، ناتکین باشد. در این صورت، گام تکرار در الگوریتم نیوتن به این صورت است

$$x_{n+1} = x_n - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n).$$

مزیت بزرگ روش نیوتن در سرعت همگرایی بالای آن است. اما این مزیت، هزینه سنگینی دارد. علاوه بر آنکه باید در هر گام، وارون ماتریس هسین را محاسبه کرد - که معادل با حل یک دستگاه از معادلات خطی است - باید شرایط نسبتاً زیادی هم برای همگرایی روش نیوتن فراهم باشد. برای نخستین بار نستیرف^۲ و نیمروفسکی^۳ با معرفی رده تابع‌های خودسازگار^۴ نشان دادند که روش نیوتن برای این دسته از تابع‌ها بسیار خوش‌رفتار است و همگرایی آن نیز تضمین می‌شود [۱۸]. تابع $f: \mathbb{R} \rightarrow \mathbb{R}$ را خودسازگار می‌نامیم هرگاه دست‌کم سه بار مشتق‌پذیر باشد، اکیداً محدب باشد (یعنی برای هر عدد حقیقی x ، $f''(x) > 0$)، و داشته باشیم

$$|f'''(x)| \leq 2 (f''(x))^{\frac{3}{2}} \quad (x \in \mathbb{R}).$$

^۱Hessian ^۲Nesterov ^۳Nemirovskii ^۴self-concordant

به همین ترتیب، تابع $f : \mathbb{R}^n \rightarrow \mathbb{R}$ خودسازگار است هرگاه برای هر $x, v \in \mathbb{R}^n$ تابع اسکالر $t \mapsto f(x + tv)$ خودسازگار باشد؛ یعنی تحدید f به هر خط راست خودسازگار باشد.

یک وضعیت پیچیده‌تر در مسئله بهینه‌سازی (۱.۱) زمانی رخ می‌دهد که یا خود X یا ناحیه D که متناظر با قید مسئله است خمینه باشد. در واقع، از یک نظر، این وضعیتی تقریباً متداول در مسائل بهینه‌سازی است. توجه کنیم که خود فضای \mathbb{R}^n یا هر زیرفضای مستوی^۱ آن، به‌طور طبیعی خمینه‌های همواری هستند. اما مشکل دیگر زمانی رخ می‌دهد که خمینه موردنظر یک خمینه غیرتخت باشد؛ مثلاً یک کره یا یک سهمی‌گون یا هر خمینه دلخواه دیگری. یک مورد بسیار مهم، حالتی است که X خمینه ماتریسی باشد. مثلاً می‌توان دید که مجموعه همه ماتریس‌های متقارن معین مثبت از مرتبه n یک خمینه ریمانی از بُعد $n(n+1)/2$ است. از طرفی، چنین خمینه‌هایی در بسیاری از مسائل بهینه‌سازی در مهندسی، فیزیک یا علوم داده ظاهر می‌شوند و از طرف دیگر، روش‌های کلاسیک مسائل بهینه‌سازی، چون مبتنی بر فضای اقلیدسی هستند، تناسب چندانی با ساختار چنین خمینه‌هایی ندارند. اغلب این روش‌ها در این موارد بسیار پرهزینه هستند و حتی گام‌های تکرار را در آن خمینه حفظ نمی‌کنند، و بنابراین عملاً مفید نیستند. به همین دلیل در سال‌های اخیر، موضوع بهینه‌سازی روی خمینه‌ها به موضوع پژوهشی پراهمیتی در رشته‌های ریاضی کاربردی، مهندسی، و علوم داده تبدیل شده است. خواننده می‌تواند در این زمینه به [۱، ۲۲، ۲۵] مراجعه کند. بحث روش بهینه‌سازی نیوتن و نیز تابع‌های خودسازگار روی خمینه‌های ریمانی هم اخیراً مورد توجه بسیاری از جمله خود نستیرف قرار گرفته است [۱۹]. علاوه بر [۶، ۱۱، ۱۲، ۱۴، ۱۷]، رساله دکتری [۱۳] هم مرجع مناسبی در این موضوع است.

از سوی دیگر، برپایه کارهای هندسه دانان بزرگی چون گاوس، آدامار^۲، و کارتان^۳ و کسانی چون آلکسیساندرف^۴، توپونوگف^۵، و گرومف^۶ با انگیزه‌هایی که ارتباط چندانی به بهینه‌سازی ندارد، دست به ابداع یک مفهوم هندسی جدید زده‌اند که از آن به هندسه متریکی یا هندسه آلکسیساندرف نام می‌برند و ساختارهای مورد مطالعه را فضاهای طولی، فضاهای ژئودزیک، فضاهای $CAT(\kappa)$ ، و یا فضاهای آدامار می‌نامند.^۷ امروزه این مفهوم هندسی به‌طور وسیعی در سراسر ریاضی مورد توجه ریاضی‌دانان است. از جمله می‌توان به آثار پرلمان در این زمینه [۲۱] و نیز به تأثیر این شاخه جدید در برخی نتایج ویلانی [۱۶] - هر دو از برندگان جایزه فیلدز - اشاره کرد. در بخش دوم مقاله حاضر شرح کوتاهی از این فضاها را می‌آوریم. آثار [۸، ۹، ۱۰] و نیز کتاب تازه منتشر شده [۴] که بسیار خواندنی و سودمند

^۷ تعریف این فضاها با هم فرق دارد؛ به بخش ۲ در ادامه مقاله و نیز رابطه (۶.۲) مراجعه شود.

^۱affine ^۲Hadamard ^۳Cartan ^۴Alexandrov ^۵Toponogov ^۶Gromov

است، از منابع معروف در این زمینه هستند. در این فضاها، با آنکه تعمیم خمینه‌های ریمانی محسوب می‌شوند، محاسبات، به دلیل فقدان ساختار تانسوری، اغلب ساده‌تر از محاسبات نظیر در خمینه‌های ریمانی است. به علاوه، بسیاری از ساختارهای ناهموار ریاضی را می‌توان در شمار چنین فضاهایی آورد. چنین ملاحظاتی به موضوع بهینه‌سازی روی این فضاها انجامیده است؛ برای نمونه [۵] و [۲۰] را مطالعه کنید.

در بخش ۲ی این مقاله پس از معرفی فضاهای ژئودزیک، به اختصار هندسهٔ متریکی را مرور می‌کنیم. در بخش ۳، به تعمیم مفهوم تابع خودسازگار روی این فضاها می‌پردازیم و از جمله فرمول تقریب مرتبهٔ دوم (۹.۳) را برای این تابع‌ها بیان می‌کنیم. سرانجام در بخش ۴، الگوریتم بهینه‌سازی نیوتن برای تابع‌های خودسازگار روی فضاهای ژئودزیک ارائه و همگرایی آن اثبات می‌شود. این مقاله، دراصل، کوششی است برای معرفی حوزهٔ جدید پژوهشی که هم از جنبهٔ نظری و هم کاربردی بسیار غنی و سرشار از مسائل عمیق و جذاب است. بحث دربارهٔ روش نیوتن صرفاً نقطهٔ شروعی برای رسیدن به این هدف است.

۲. فضاهای ژئودزیک

فرض کنید (X, d) یک فضای متریک باشد. یک خم، تابعی پیوسته مانند $\gamma : I_\gamma \rightarrow X$ است که در آن I_γ بازه‌ای ناتهی در \mathbb{R} است. طول خم γ به صورت زیر تعریف می‌شود

$$L(\gamma) = \sup \sum_{k=1}^n d(\gamma(t_{k-1}), \gamma(t_k))$$

که در آن n عددی طبیعی و سوپریم روی همهٔ نقاط $t_0, \dots, t_n \in I_\gamma$ با شرط $t_0 < \dots < t_n$ گرفته می‌شود. خم γ را طول‌پذیر نامیم هرگاه $L(\gamma) < +\infty$. گوئیم فضای متریک (X, d) یک فضای طولی^۱ است اگر برای هر دو نقطهٔ $x, y \in X$ داشته باشیم^۲

$$d(x, y) = \inf \{L(\gamma) \mid \gamma : [0, 1] \rightarrow X, \gamma(0) = x, \gamma(1) = y\}. \quad (1.2)$$

به عبارت دیگر، فضای طولی یک فضای متریک است که فاصلهٔ میان نقاط آن توسط طول خم‌ها به دست آید.

^۲ به جای این روش، می‌توان ابتدا فضای طولی را به صورت یک مجموعه به همراه یک «ساختار طولی» تعریف کرد سپس این ساختار یک متریک القاء می‌کند که رابطه (۱.۲) برای آن برقرار است؛ مراجعه کنید به فصل دوم [۹].

^۱length space

مثال ۱.۲. مجموعه نقاطی از صفحه که خارج از ربع اول باشند، یعنی

$$X = \{(x, y) \in \mathbb{R}^2 \mid x \leq 0 \text{ یا } y \leq 0\}$$

با متریک اقلیدسی، به وضوح، یک فضای طولی نیست؛ زیرا، مثلاً، طول هر خم بین دو نقطه $(0, 1)$ و $(1, 0)$ در X حداقل برابر با ۲ است. با این حال، اگر رابطه (۱.۲) را به عنوان تعریف متریکی جدید در نظر بگیریم، یک فضای طولی به دست می آید. با استفاده از همین ایده، می توان از هر فضای متریک همبند مسیری، یک فضای طولی ساخت؛ [۹] را ببینید.

خمی مانند $\gamma: I_\gamma \rightarrow X$ را در نظر بگیرید. می توان دید که شرایط زیر معادل اند:

(۱) برای هر $a, b, c \in I_\gamma$ با شرط $a < b < c$ داریم

$$d(\gamma(a), \gamma(c)) = d(\gamma(a), \gamma(b)) + d(\gamma(b), \gamma(c)).$$

(۲) برای هر $a, b \in I_\gamma$ داریم

$$L(\gamma|_{[a,b]}) = d(\gamma(a), \gamma(b)).$$

(۳) یک عدد حقیقی ثابت و نامنفی s وجود دارد به طوری که برای هر $a, b \in I_\gamma$ داریم

$$d(\gamma(a), \gamma(b)) = s|a - b|.$$

اگر خم γ در این شرطها صدق کند، می گوییم یک ژئودزی با سرعت s است.

تعریف ۲.۲. فضای متریک (X, d) را فضای ژئودزیک نامیم هرگاه بین هر دو نقطه آن دست کم یک ژئودزی وجود داشته باشد. گوییم (X, d) یک فضای به طور یکتا ژئودزیک^۱ است هرگاه بین هر دو نقطه متمایز آن تنها یک ژئودزی وجود داشته باشد.

اگر (X, d) یک فضای به طور یکتا ژئودزیک و $x, y \in X$ دو نقطه متمایز آن باشند آنگاه ژئودزی یکتای γ_{xy} وجود دارد به طوری که $I_{\gamma_{xy}} \subset [0, 1]$ و نیز $\gamma_{xy}(0) = x$ و $\gamma_{xy}(1) = y$. توجه کنید که γ_{xy} یک ژئودزی با سرعت $d(x, y)$ است. روشن است که هر فضای ژئودزیک، فضای طولی است. در فضاهای طولی کامل و موضعاً فشرده ویژگی آشنای هاینه-بورل نیز برقرار است. در واقع، قضیه بسیار سودمند زیر را داریم.

^۱uniquely geodesic space

قضیه ۳.۲ (قضیه هویف-رینو^۱). فرض کنید (X, d) یک فضای طولی موضعاً فشرده باشد. در این صورت گزاره‌های زیر هم‌رزنند:

(۱) (X, d) یک فضای متریک کامل است، یعنی هر دنباله کوشی در آن همگراست.

(۲) هر زیرمجموعه بسته و کراندار از X در (X, d) فشرده است.

به علاوه، در صورت برقراری این دو شرط، (X, d) یک فضای ژئودزیک است.

خوب است به این موضوع توجه کنیم که در اینجا هم -مشابه با فضاهای برداری توپولوژیکی- بین موضعاً فشرده بودن فضای ژئودزیک و متناهی بودن بُعد آن ارتباط نزدیکی برقرار دارد. در ادامه این بخش، مثال‌های متنوعی از فضاهای ژئودزیک ارائه خواهیم کرد.

فضاهای مدل. هر خمینه ریمانی همبند فضایی ژئودزیک است. با این حال، سه خمینه ریمانی زیر که به ترتیب دارای خمیدگی‌های مقطعی^۲ ثابت صفر، مثبت، و منفی هستند، به‌دلیلی که در ادامه روشن خواهد شد، اهمیت مضاعفی دارند. ابتدا، به فضای اقلیدسی می‌پردازیم که یک خمینه ریمانی تخت^۳ است. جزئیات و اثبات‌های مربوط به این مثال‌ها را می‌توان، مثلاً، در [۸] یافت.

مثال ۴.۲. فرض کنید n عددی صحیح و مثبت باشد. مجموعه

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, \dots, x_n) \mid x_i \in \mathbb{R}\}$$

را به حاصل ضرب داخلی استاندارد^۴ $\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^n x_k y_k$ ، نرم اقلیدسی $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ ، و متریک ناشی از آن مجهز می‌کنیم. این فضای متریک را با \mathbb{E}^n نشان می‌دهیم. به‌سادگی می‌توان دید که \mathbb{E}^n یک فضای به‌طور یکتا ژئودزیک است. در واقع، برای هر دو نقطه $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ داریم

$$\gamma_{xy}(t) = (1-t)\mathbf{x} + t\mathbf{y}, \quad (t \in \mathbb{R}).$$

مثال دوم، کره^۵ n -بُعدی واحد است که دارای خمیدگی مقطعی ثابت 1 است.

مثال ۵.۲. فرض کنید n عددی صحیح و مثبت باشد. تعریف کنید

$$S^n = \{\mathbf{x} \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\| = 1\}.$$

متریک ذاتی^۶ روی این فضا به‌صورت زیر تعریف می‌شود

$$d_{S^n}(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\mathbf{x} \cdot \mathbf{y}) \in [0, \pi], \quad (\mathbf{x}, \mathbf{y} \in S^n). \quad (۲.۲)$$

^۱Hopf-Rinow Theorem ^۲sectional curvature ^۳flat ^۴intrinsic

به سادگی می توان دید که (\mathbb{S}^n, d) یک فضای متریک ژئودزیک کامل و فشرده است. در واقع، همه ژئودزی‌ها آن با سرعت یک به صورت زیر توصیف می‌شوند: گیریم $\mathbf{x} \in \mathbb{S}^n$ یک نقطه دلخواه و $\mathbf{u} \in \mathbb{R}^{n+1}$ برداری یکه و عمود بر \mathbf{x} باشد، یعنی $\|\mathbf{u}\| = 1$ و $\mathbf{u} \cdot \mathbf{x} = 0$ ؛ در این صورت

$$\gamma(t) = (\cos t)\mathbf{x} + (\sin t)\mathbf{u} \quad (۳.۲)$$

یک ژئودزی با سرعت واحد آغاز شده از \mathbf{x} است. $\gamma(0) = \mathbf{x}$ در واقع، برای هر $t, s \in \mathbb{R}$ با شرط $|t - s| \leq \pi$ داریم

$$d(\gamma(t), \gamma(s)) = |t - s|.$$

اگر $\mathbf{x}, \mathbf{y} \in \mathbb{S}^n$ دو نقطه متمایز و غیرممتقاطع باشند، یعنی $\mathbf{y} \notin \{\mathbf{x}, -\mathbf{x}\}$ ، با تعریف

$$\theta := d(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\mathbf{x} \cdot \mathbf{y}) \in (0, \pi)$$

و

$$\mathbf{u} := \frac{\mathbf{y} - (\cos \theta)\mathbf{x}}{\sin \theta}$$

ضابطه ژئودزی زیر به دست می‌آید

$$\gamma_{\mathbf{x}\mathbf{y}}(t) = (\cos \theta t)\mathbf{x} + (\sin \theta t)\mathbf{u} \quad (۴.۲)$$

که مطابق انتظار ما، $\gamma_{\mathbf{x}\mathbf{y}}(0) = \mathbf{x}$ و $\gamma_{\mathbf{x}\mathbf{y}}(1) = \mathbf{y}$.

سومین مثال، فضای هذلولوی n -بعدی با خمیدگی مقطعی ثابت -1 است.

مثال ۶.۲. ابتدا صورت دوخطی زیر را در نظر بگیرید

$$\langle \mathbf{x}, \mathbf{y} \rangle = -x_{n+1}y_{n+1} + \sum_{k=1}^n x_k y_k, \quad (\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n+1}).$$

تعریف می‌کنیم

$$\mathbb{H}^n = \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{x} \rangle = -1, x_{n+1} > 0\}.$$

به سادگی می‌توان نشان داد که برای هر دو نقطه $\mathbf{x}, \mathbf{y} \in \mathbb{H}^n$ داریم $\langle \mathbf{x}, \mathbf{y} \rangle \leq -1$ و تساوی تنها وقتی برقرار است که $\mathbf{x} = \mathbf{y}$. به کمک این مطلب می‌توان ثابت کرد که تابع زیر یک متریک تعریف می‌کند

$$d_{\mathbb{H}^n}(\mathbf{x}, \mathbf{y}) = \cosh^{-1}(-\langle \mathbf{x}, \mathbf{y} \rangle), \quad (\mathbf{x}, \mathbf{y} \in \mathbb{H}^n). \quad (۵.۲)$$

فضای متریک (\mathbb{H}^n, d) نیز یک فضای متریک ژئودزیک کامل و موضعاً فشرده است. در واقع، همه ژئودزی‌های آن با سرعت یک به صورت زیر توصیف می‌شوند: فرض کنید $\mathbf{x} \in \mathbb{H}^n$ یک نقطه دلخواه و $\mathbf{u} \in \mathbb{R}^{n+1}$ برداری باشد که $\langle \mathbf{u}, \mathbf{u} \rangle = 1$ و $\langle \mathbf{u}, \mathbf{x} \rangle = 0$. در این صورت

$$\gamma(t) = (\cosh t)\mathbf{x} + (\sinh t)\mathbf{u}$$

یک ژئودزی با سرعت واحد آغاز شده از $\mathbf{x} = \gamma(0)$ است.

اکنون فضای مُدل \mathbb{M}_k را تعریف می‌کنیم. این فضا یک خمینه ریمانی دو بُعدی با خمیدگی مقطعی ثابت k است.

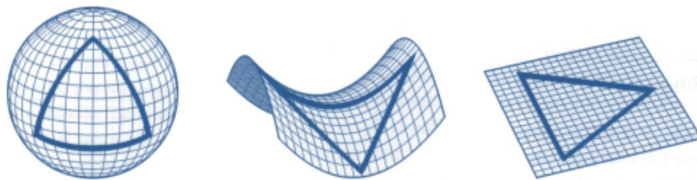
تعریف ۷.۲. برای عدد حقیقی k ، فضای \mathbb{M}_k چنین تعریف می‌شود:

$$(1) \quad \mathbb{M}_0 = \mathbb{E}^2 \text{ صفحه اقلیدسی است؛}$$

(۲) اگر k مثبت باشد، \mathbb{M}_k همان \mathbb{S}^2 است که متریک استانداردش (۲.۲) در $1/\sqrt{k}$ ضرب شده است؛

(۳) اگر k منفی باشد، \mathbb{M}_k همان \mathbb{H}^2 است که متریک استانداردش (۵.۲) در $1/\sqrt{-k}$ ضرب شده است.

قطر فضای مُدل $D_k = \text{diam } \mathbb{M}_k$ پارامتری بسیار مفید است. در واقع، D_k تنها هنگامی که $k > 0$ یک عدد متناهی و برابر $D_k = \pi/\sqrt{k}$ است.



شکل ۱. فضای مدل متناظر با (به ترتیب از راست به چپ) $k = 0$ ، $k < 0$ و $k > 0$ منفی، و $k > 0$ مثبت

خمیدگی بدون تانسور. اندیشه اولیه بیان خمیدگی برحسب خم‌ها به اوائل دوران کشف هندسه‌های ناقلیدسی و تحقیقات گاوس می‌رسد. در شکل کنونی هندسه متریک، به زبان نادقیق، گفته می‌شود که یک فضای ژئودزیک دارای خمیدگی آلکساندرف کمتر یا مساوی κ است هرگاه هر مثلث آن از مثلث نظیر در \mathbb{M}_{κ} «چاق‌تر» نباشد؛ همچنین دارای خمیدگی آلکساندرف بیشتر یا مساوی κ است هرگاه هر مثلث آن از مثلث نظیر در \mathbb{M}_{κ} «لاغرتر» نباشد. این مفاهیم را به زبان دقیق ریاضی بیان خواهیم کرد (تعریف ۸.۲ را ببینید)؛ برای این کار نخست باید با مفاهیم مثلث ژئودزیک و مثلث قیاسی آشنا شویم.

فرض کنید (X, d) یک فضای ژئودزیک و $x_1, x_2, x_3 \in X$ سه نقطه متمایز در آن باشند. منظور از مثلث ژئودزیک $\Delta(x_1, x_2, x_3)$ اجتماع سه ژئودزی واصل میان این نقاط در X است (همان محیط مثلث). بنابر قضیه‌ای از آلکساندرف، برای هر عدد حقیقی κ و هر مثلث ژئودزیک $\Delta(x_1, x_2, x_3)$ در X که محیط آن از $2D_{\kappa}$ کمتر باشد، یک مثلث ژئودزیک $\bar{\Delta}(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ در \mathbb{M}_{κ} وجود دارد که تاحد طولپایی یکتاست و

$$d_X(x_i, x_j) = d_{\mathbb{M}_{\kappa}}(\bar{x}_i, \bar{x}_j), \quad (i, j \in \{1, 2, 3\}).$$

مثلث $\bar{\Delta} \subset \mathbb{M}_{\kappa}$ را مثلث قیاسی نظیر مثلث ژئودزیک $\Delta \subset X$ می‌نامیم.

تعریف ۸.۲. گوئیم فضای ژئودزیک (X, d) دارای خمیدگی کمتر یا مساوی κ است هرگاه برای هر مثلث ژئودزیک $\Delta \subset X$ با محیط کوچک‌تر از $2D_{\kappa}$ این شرط برقرار باشد: اگر $\bar{\Delta}$ مثلث قیاسی متناظر و $p, q \in \Delta$ دو نقطه دلخواه باشند، برای نقطه‌های متناظر $\bar{p}, \bar{q} \in \bar{\Delta}$ داشته باشیم

$$d_X(p, q) \leq d_{\mathbb{M}_{\kappa}}(\bar{p}, \bar{q}).$$

خمیدگی بزرگ‌تر یا مساوی κ نیز به‌طور مشابه با نامساوی در جهت برعکس تعریف می‌شود.

معمولاً به پیروی از گروه [۱۰]، رده فضاهای ژئودزیک با خمیدگی کوچک‌تر یا مساوی κ را با $\text{CAT}(\kappa)$ نشان می‌دهند. حروف C, A, T ، به ترتیب، حروف نخست اسامی کارتان، آلکساندرف، و توپونوگف‌اند. در کتاب [۸] به مطالعه این فضاها و برخی کاربردهای آن در نظریه گروه‌ها و توپولوژی جبری پرداخته شده است. یک فضای ژئودزیک را که خمیدگی آن از بالا یا پایین کراندار باشد «فضای آلکساندرف» می‌نامند. کتاب‌های [۴] و [۹] به‌طور مفصل به این فضاها می‌پردازند. در حالت خاص، یک فضای $\text{CAT}(0)$ کامل را یک «فضای آدامار» می‌نامند. فضاهای

^۱comparison triangle

آدامار را می‌توان فضاهاى هیلبرت «خمیده» تصور کرد! در واقع، این فضاها دارای ویژگی‌های فراوانی هستند که آن‌ها را تبدیل به چارچوب مناسبی برای آنالیز غیرخطی، آنالیز محدب، و نظریهٔ بهینه‌سازی می‌کند. از جملهٔ این ویژگی‌ها می‌توان به وجود تابع تصویر روی هر زیرمجموعهٔ محدب، مفهومی مشابه با حاصل ضرب داخلی به نام شبه‌خطی‌سازی^۱ [۲]، توپولوژی‌های ضعیف [۳]، خاصیتی مشابه قضیهٔ باناخ-آلاقلو^۲، و بسیاری موارد دیگر اشاره کرد. خواننده علاقه‌مند می‌تواند منابع [۵] و [۱۵] را در این زمینه مطالعه کند. مهم‌ترین مثال‌های فضای آدامار عبارت‌اند از فضاهاى هیلبرت، خمینه‌های ریمانی همبند ساده با خمیدگی مقطعی نامشبت، درخت‌ها (گراف‌های همبند و فاقد دور)، هر زیرمجموعهٔ بسته و محدب از یک فضای آدامار، حاصل ضرب‌های مستقیم یا شبه‌مستقیم از فضاهاى آدامار، توابع از ردهٔ L^2 با مقادیر در یک فضای آدامار، فضاهاى حاصل از چسباندن به شیوهٔ ریشتنیاک^۳ یا خارج‌قسمت گرفتن از فضاهاى آدامار، و موارد بسیار دیگر. با اینکه فضاهاى آدامار بیشترین قابلیت را برای نظریهٔ بهینه‌سازی دارند، ما در این مقاله بحث را در چارچوبی بسیار کلی‌تر عرضه می‌کنیم. چارچوب ما، در ادامهٔ مقاله، یک فضای ژئودزیک کامل است. این فضاها همهٔ فضاهاى آلیکساندراف و حتی فراتر از آن‌را شامل می‌شوند. در واقع، ارتباط میان فضاهاى مطرح در هندسهٔ متریکی را می‌توان به‌صورت زیر خلاصه کرد.

(۶.۲) فضای متریک \Rightarrow فضای طولی \Rightarrow فضای ژئودزیک \Rightarrow فضای آلیکساندراف \Rightarrow فضای $CAT(\kappa)$ \Rightarrow فضای آدامار

این بخش را با ذکر دو قضیه، که ارتباط بین خمینه‌های ریمانی و فضاهاى آلیکساندراف را روشن می‌کنند، به پایان می‌بریم. خوانندهٔ علاقه‌مند می‌تواند اثبات آن‌را در کتاب‌های [۹، ۱۰] بیابد.

قضیه ۹.۲. فرض کنید M یک خمینهٔ ریمانی کامل و همبند و κ عددی حقیقی باشد. خمیدگی مقطعی M همه‌جا کمتر یا مساوی (بیشتر یا مساوی) κ است اگر و تنها اگر به‌طور موضعی خمیدگی آلیکساندراف آن (با تعریف ۸.۲) کمتر یا مساوی (بیشتر یا مساوی) κ باشد.

به‌عبارت‌دیگر، خمیدگی آلیکساندراف تعمیم مفهوم خمیدگی مقطعی در خمینه‌های ریمانی است. اکنون فرض کنید κ عدد حقیقی، $D > 0$ عددی حقیقی و مثبت، و n عدد صحیح و نامنفی داده شده باشند. ردهٔ همهٔ فضاهاى ژئودزیک با خمیدگی بیشتر یا مساوی κ ، قطر کمتر یا مساوی D ، و بُعد هاوسدورف کمتر یا مساوی n را با $\mathfrak{M}(n, \kappa, D)$ نشان می‌دهیم.

قضیه ۱۰.۲ (قضیهٔ فشردگی گرومف). ردهٔ $\mathfrak{M}(n, \kappa, D)$ نسبت به متریک گرومف-هاوسدورف فشرده است.

^۱quasilinearization ^۲Banach-Alaoglu Theorem ^۳Reshetnyak's gluing

یادآوری کنیم که متریک گرومف-هاوسدورف، یک تابع فاصله بین فضاهای متریک است که تعمیم متریک هاوسدورف بین زیرمجموعه‌های فشرده یک فضای متریک است. همچنین بُعد هاوسدورف، تعمیم بُعد خمینه‌ای به فضاهای متریک است. این تعریف‌های کلاسیک را برای مثال در فصل اول و فصل دهم [۹] می‌توان مطالعه کرد. توجه کنید که رده همه خمینه‌های ریمانی با خمیدگی مقطعی بیشتر یا مساوی κ ، قطر کمتر یا مساوی D ، و بُعد کمتری مساوی n زیرده $\mathcal{M}(n, \kappa, D)$ است. بنابراین از قضیه بالا نتیجه می‌شود که حد هر دنباله همگرا از چنین خمینه‌هایی در متریک گرومف-هاوسدورف یک فضای آلکساندر است؛ درحالی‌که، ممکن است دیگر یک خمینه ریمانی نباشد و حتی ساختار هموار نداشته باشد. یک کاربرد جذاب و غیرمنتظره از این مطلب، اثبات پایداری مدل فریدمن^۱ در کیهان‌شناسی است [۲۳].

۳. تابع‌های خودسازگار

از این پس همواره فرض می‌کنیم (X, d) یک فضای متریک ژئودزیک کامل باشد. برای هر نقطه $x \in X$ مجموعه همه راستاهای آغاز شده از x را با علامت $S_x(X)$ نشان می‌دهیم. راستای $\gamma \in S_x(X)$ یک ژئودزی مانند $\gamma : I_\gamma \rightarrow X$ است که در آن $I_\gamma \subset \mathbb{R}$ بازه‌ای باز با شرط $\gamma(0) = x$ و $0 \in I_\gamma$ است. مجموعه $S(X) := \bigcup_{x \in X} S_x(X)$ را نیز مجموعه همه راستاهای روی X می‌نامیم. خود این مجموعه‌ها را هم می‌توان به یک متریک طبیعی مجهز کرد که آن‌ها را به فضاهای متریک ژئودزیک جدیدی تبدیل می‌کند. این فضاها را می‌توان، به ترتیب، تعمیم فضای مماسی و کلاف مماسی بر خمینه مشتق‌پذیر تلقی کرد. چون در این مقاله تنها از خود راستاها استفاده می‌کنیم، به تعریف متریک اخیر نمی‌پردازیم؛ خواننده علاقه‌مند می‌تواند به کتاب‌های [۴، ۸، ۹] مراجعه کند. اکنون دو مفهوم مهم در این زمینه را تعریف می‌کنیم. یادآوری می‌کنیم که $D \subset X$ محدب است هرگاه شامل همه ژئودزی‌های واصل بین تمام نقاط خود باشد.

تعریف ۱.۳. فرض کنید $D \subset X$ مجموعه‌ای ناتهی، باز، و محدب باشد.

- (۱) تابع پیوسته $f : D \rightarrow \mathbb{R}$ را اکیداً محدب می‌نامند هرگاه برای هر راستای $\gamma \in S(X)$ در D ، یعنی $\gamma(I_\gamma) \subset D$ ، تابع $f \circ \gamma : I_\gamma \rightarrow \mathbb{R}$ دوبار مشتق‌پذیر باشد و $(f \circ \gamma)''(0) > 0$.
- (۲) تابع اکیداً محدب $f : D \rightarrow \mathbb{R}$ را خودسازگار گوییم هرگاه عدد ثابت و مثبتی مانند M_f موجود باشد به طوری که برای هر راستای γ در D تابع $f \circ \gamma$ سه بار مشتق‌پذیر باشد و

$$|(f \circ \gamma)'''(0)| \leq M_f ((f \circ \gamma)''(0))^{\frac{3}{2}}. \quad (1.3)$$

^۱Friedmann

ضریب M_f را می‌توان با ضرب کردن تابع f در اعداد مثبت تغییر داد؛ در واقع، اگر تابع f رابطه (۱.۳) صدق کند و $a > 0$ عدد حقیقی مثبتی باشد آنگاه تابع af در شرط فوق، با ضریب M_f/\sqrt{a} صدق می‌کند. اغلب، بدون کاستن از کلیت، فرض می‌کنیم $M_f = 2$. مثال زیر، یک مثال نابديهی برای تابعی خودسازگار روی کره است.

مثال ۲.۳. کُنج مثبت کره واحد، یعنی $D = \{x \in \mathbb{S}^n | x > 0\}$ را در نظر بگیرید که در آن منظور از نماد $x > 0$ ، مثبت بودن همه مؤلفه‌های x است. نشان می‌دهیم که تابع $f : D \rightarrow \mathbb{R}$ با ضابطه

$$f(x) = - \sum_{k=1}^{n+1} \ln x_k$$

یک تابع خودسازگار است. برای این کار، با استفاده از ضابطه ژئودزی‌ها روی \mathbb{S}^n ، یعنی (۳.۲) که از D آغاز شده باشند، و اندکی محاسبه خواهیم داشت

$$(f \circ \gamma)'(0) = - \sum_{k=1}^{n+1} \frac{u_k}{x_k}, \quad (f \circ \gamma)''(0) = \sum_{k=1}^{n+1} \left(1 + \left(\frac{u_k}{x_k}\right)^2\right)$$

و

$$(f \circ \gamma)'''(0) = -2 \sum_{k=1}^{n+1} \left(\frac{u_k}{x_k}\right) \left(1 + \left(\frac{u_k}{x_k}\right)^2\right).$$

به این ترتیب، f یک تابع اکیداً محدب است، زیرا $(f \circ \gamma)''(0) > 0$ ؛ همچنین با استفاده از نابرابری کوشی-شوارتس و نیز نابرابری $\sum_k a_k^2 \leq \left(\sum_k a_k\right)^2$ برای اعداد حقیقی مثبت، داریم

$$\begin{aligned} ((f \circ \gamma)'''(0))^2 &\leq 4 \sum_{k=1}^{n+1} \left(\frac{u_k}{x_k}\right)^2 \sum_{k=1}^{n+1} \left(1 + \left(\frac{u_k}{x_k}\right)^2\right)^2 \\ &\leq 4 \sum_{k=1}^{n+1} \left(1 + \left(\frac{u_k}{x_k}\right)^2\right) \left(\sum_{k=1}^{n+1} \left(1 + \left(\frac{u_k}{x_k}\right)^2\right)\right)^2 = 4((f \circ \gamma)''(0))^3. \end{aligned}$$

پس تعریف خودسازگاری (۱.۳) برای تابع f با همان ضریب $M_f = 2$ برقرار است.

برای هر زیرمجموعه به‌طور یکتا ژئودزیک D ، مشابه با حالت هموار، می‌توان بیضی‌گون‌هایی از نوع دیکین^۱ به‌شکل زیر تعریف کرد. فرض کنید $x \in D$ و $r > 0$ ، تعریف کنید

$$W(x; r) := \{\gamma \in S_x(X) | [(f \circ \gamma)'''(0)]^{\frac{1}{3}} < r\}$$

^۱Dikin-type ellipsoids

و

$$W^\circ(x; r) := \{y \in X \mid [(f \circ \gamma_{xy})''(\circ)]^{\frac{1}{\gamma}} < r\}.$$

توجه کنیم که برای هر راستای $\gamma \in S_x(X)$ داریم $\gamma \in W(x; r)$ اگر و تنها اگر داشته باشیم $\gamma(1) \in W^\circ(x; r)$. از این پس، همیشه فرض می‌کنیم که مجموعه D یک زیرمجموعه به طور یکتا ژئودزیک از X باشد.

در ادامه مقاله لم زیر بسیار به کار خواهد آمد.

لم ۳.۳. فرض کنید تابع خودسازگار $f: D \rightarrow \mathbb{R}$ و جهت γ در D داده شده باشند. تعریف کنید

$$\phi(t) := ((f \circ \gamma)''(t))^{-\frac{1}{\gamma}}, \quad (t \in I_\gamma).$$

در این صورت

$$(1) \text{ برای هر } t \in I_\gamma \text{ داریم } |\phi'(t)| \leq 1.$$

$$(2) \text{ اگر } \phi(\circ) > 0 \text{ آنگاه } (-\phi(\circ), \phi(\circ)) \subset I_\gamma.$$

اثبات. با مشتق گرفتن از تابع ϕ داریم

$$\phi'(t) = -\frac{1}{\gamma} ((f \circ \gamma)''(t))^{-\frac{\gamma+1}{\gamma}} (f \circ \gamma)'''(t), \quad (t \in I_\gamma).$$

اکنون از تعریف (۱.۳) نتیجه می‌شود که $|\phi'(\circ)| \leq 1$ ، اما چون γ می‌تواند هر راستای آغازشده از نقطه‌ای دلخواه باشد پس حکم (۱) نتیجه می‌شود.

برای اثبات حکم (۲) فرض کنید بازه I_γ برابر با کل اعداد حقیقی نباشد، زیرا در غیر این صورت حکم واضح است. اگر $T > 0$ را بزرگ‌ترین عددی بگیریم که $(-T, T) \subset I_\gamma$ ، آنگاه دست‌کم یکی از دو انتهای این بازه در I_γ نیست، مثلاً $T \notin I_\gamma$. اکنون کافی است ثابت کنیم $\phi(\circ) \leq T$. برای این کار با برهان خلف عمل می‌کنیم. فرض کنید $\phi(\circ) < T$. برای هر $t \in (-T, T)$ از (۱) نتیجه می‌شود که $|\phi(t) - \phi(\circ)| \leq |t|$ پس

$$\phi(t) \geq \phi(\circ) - |t| \geq \phi(\circ) - T.$$

بنابراین از فرض خلف و تعریف تابع ϕ نتیجه می‌شود که

$$(f \circ \gamma)''(t) = \phi(t)^{-\gamma} \leq \frac{1}{(\phi(\circ) - T)^\gamma}$$

در نتیجه برای هر $t \in (-T, T)$ داریم

$$\begin{aligned}(f \circ \gamma)(t) &= (f \circ \gamma)(\circ) + \int_{\circ}^t ((f \circ \gamma)'(\circ) + \int_{\circ}^{\tau} (f \circ \gamma)''(s) ds) d\tau \\ &\leq (f \circ \gamma)(\circ) + (f \circ \gamma)'(\circ)t + \frac{t^2}{2(\phi(\circ) - T)^2}.\end{aligned}$$

□ اما این یعنی $(f \circ \gamma)(T) < +\infty$ ، پس $T \in I_{\gamma}$ که با انتخاب T در تناقض است.

برای هر ژئودزی $\gamma : I_{\gamma} \rightarrow X$ و هر نقطه ثابت $t \in I_{\gamma}$ ، خم «بازیرمایش» شده‌ای مانند $\gamma^t : [0, 1] \rightarrow X$ را چنین تعریف می‌کنیم

$$\gamma^t(s) = \gamma(st), \quad (s \in [0, 1]). \quad (2.3)$$

در این صورت اگر خم γ در دامنه تابع f قرار گرفته باشد و $f \circ \gamma$ نیز مشتق پذیر باشد، آنگاه واضح است که برای هر $s \in [0, 1]$ داریم $(f \circ \gamma^t)'(s) = t(f \circ \gamma)'(st)$.

گزاره ۴.۳. فرض کنید $f : D \rightarrow \mathbb{R}$ یک تابع خودسازگار، $x, y \in D$ ، $\gamma := \gamma_{xy}$ ، و $r := ((f \circ \gamma)''(\circ))^{-1} > 0$. در این صورت داریم

$$(f \circ \gamma)''(1) \geq \frac{r^2}{(1+r)^2}, \quad (3.3)$$

$$(f \circ \gamma)'(1) - (f \circ \gamma)'(\circ) \geq \frac{r^2}{1+r}, \quad (4.3)$$

$$f(y) \geq f(x) + (f \circ \gamma)'(\circ) + r - \ln(1+r). \quad (5.3)$$

اثبات. بنابه قسمت (۱) از لم ۳.۳ داریم $\phi(1) \leq \phi(\circ) + 1$ و این همان رابطه (۳.۳) است. اکنون با استفاده از این رابطه و تعریف ۲.۳، رابطه (۴.۳) به دست می‌آید:

$$\begin{aligned} (f \circ \gamma)'(1) - (f \circ \gamma)'(\circ) &= \int_{\circ}^1 (f \circ \gamma)''(t) dt \\ &= \int_{\circ}^1 \frac{1}{t^2} (f \circ \gamma^t)''(1) dt \\ &\geq \int_{\circ}^1 \frac{1}{t^2} \frac{(f \circ \gamma^t)''(\circ)}{(1 + ((f \circ \gamma^t)''(\circ))^{\frac{1}{r}})^2} dt \\ &= \int_{\circ}^1 \frac{r^2}{(1 + rt)^2} dt = \frac{r^2}{1 + r}. \end{aligned}$$

و سرانجام اینکه

$$\begin{aligned} f(y) - f(x) - (f \circ \gamma)'(\circ) &= \int_{\circ}^1 ((f \circ \gamma)'(t) - (f \circ \gamma)'(\circ)) dt \\ &= \int_{\circ}^1 \frac{1}{t} ((f \circ \gamma^t)'(1) - (f \circ \gamma^t)'(\circ)) dt \\ &\geq \int_{\circ}^1 \frac{1}{t} \frac{(f \circ \gamma^t)''(\circ)}{1 + ((f \circ \gamma^t)''(\circ))^{\frac{1}{r}}} dt \\ &= \int_{\circ}^1 \frac{r^2 t}{1 + rt} dt = r - \ln(1 + r). \end{aligned}$$

□

برای نابرابری‌هایی در جهت عکس، باید فرض کنیم $r < 1$. در واقع، گزاره زیر را داریم.

گزاره ۵.۳. فرض کنید $f : D \rightarrow \mathbb{R}$ یک تابع خودسازگار، $x \in D$ ، $y = \gamma \in W(x; 1)$ ، $r := ((f \circ \gamma)''(\circ))^{\frac{1}{r}} \in (\circ, 1)$ و $\gamma(1)$ در این صورت داریم

$$(1 - r)^2 r^2 \leq (f \circ \gamma)''(1) \leq \frac{r^2}{(1 - r)^2}, \quad (6.3)$$

$$(1 - r + \frac{r^2}{3}) r^2 \leq (f \circ \gamma)'(1) - (f \circ \gamma)'(\circ) \leq \frac{r^2}{1 - r}, \quad (7.3)$$

$$f(y) \leq f(x) + (f \circ \gamma)'(\circ) - r - \ln(1 - r). \quad (8.3)$$

اثبات. تابع $\psi : I_\gamma \rightarrow (0, +\infty)$ را با ضابطه $\psi(t) = (f \circ \gamma)''(t)$ در نظر بگیرید. اگر تعریف تابع خودسازگار (۱.۳) را برای $s \mapsto \gamma(s+t)$ به کار ببریم، با استفاده از (۲.۳)، برای $t \in I_\gamma$ داریم

$$\psi'(t) \leq 2((f \circ \gamma)''(t))^{\frac{1}{2}} (f \circ \gamma)''(t) = \frac{2}{t} ((f \circ \gamma^t)''(1))^{\frac{1}{2}} \psi(t).$$

با استفاده از قسمت (۱) لم ۳.۳ برای γ^t داریم $\phi(1) \geq \phi(0) - 1$ که به رابطه زیر می‌رسد

$$((f \circ \gamma^t)''(1))^{\frac{1}{2}} \leq \frac{((f \circ \gamma^t)''(0))^{\frac{1}{2}}}{1 - ((f \circ \gamma^t)''(0))^{\frac{1}{2}}},$$

بنابراین

$$\frac{\psi'(t)}{\psi(t)} \leq \frac{2r}{1 - rt}.$$

حال با انتگرال‌گیری از این رابطه، از $t = 0$ تا $t = 1$ ، و توجه به این مطلب که $\psi(0) = r^2$ رابطه (۶.۳) نتیجه می‌شود.

اکنون از این رابطه برای اثبات نابرابری‌های (۷.۳) استفاده می‌کنیم. نخست برای نشان دادن درستی نابرابری سمت راست، می‌نویسیم

$$\begin{aligned} (f \circ \gamma)'(1) - (f \circ \gamma)'(0) &= \int_0^1 (f \circ \gamma)''(t) dt \\ &= \int_0^1 \frac{1}{t^2} (f \circ \gamma^t)''(t) dt \\ &\leq \int_0^1 \frac{1}{t^2} \frac{(f \circ \gamma^t)''(0)}{\left(1 - ((f \circ \gamma^t)''(0))^{\frac{1}{2}}\right)^2} dt \\ &= \int_0^1 \frac{r^2}{(1 - rt)^2} dt = \frac{r^2}{1 - r}. \end{aligned}$$

و سپس برای اثبات نابرابری سمت چپ، می‌نویسیم

$$\begin{aligned} (f \circ \gamma)'(1) - (f \circ \gamma)'(0) &= \int_0^1 (f \circ \gamma)''(t) dt \\ &= \int_0^1 \frac{1}{t^2} (f \circ \gamma^t)''(t) dt \\ &\geq \int_0^1 \frac{1}{t^2} \left(1 - ((f \circ \gamma^t)''(0))^{\frac{1}{r}}\right)^2 (f \circ \gamma^t)''(0) dt \\ &= \int_0^1 r^2 (1 - rt)^2 dt = (1 - r + \frac{r^2}{3})r^2. \end{aligned}$$

سرانجام رابطه (۸.۳) را از رابطه (۷.۳) به ترتیب زیر به دست می‌آوریم

$$\begin{aligned} f(y) - f(x) - (f \circ \gamma)'(0) &= \int_0^1 ((f \circ \gamma)'(t) - (f \circ \gamma)'(0)) dt \\ &= \int_0^1 \frac{1}{t} ((f \circ \gamma^t)'(1) - (f \circ \gamma^t)'(0)) dt \\ &\leq \int_0^1 \frac{1}{t} \frac{(f \circ \gamma^t)''(0)}{1 - ((f \circ \gamma^t)''(0))^{\frac{1}{r}}} dt \\ &= \int_0^1 \frac{r^2 t}{1 - rt} dt = -r - \ln(1 - r). \end{aligned}$$

□

اگر رابطه‌های (۵.۳) و (۸.۳) را با یکدیگر مقایسه کنیم و به خاطر داشته باشیم که برای عدد دلخواه $r \in (0, 1)$ سری تیلر

$$r - \ln(1 + r) = \frac{r^2}{2} - \frac{r^3}{3} + \frac{r^4}{4} - \dots$$

معتبر است، آنگاه می‌توان نوشت

$$f(y) = f(x) + (f \circ \gamma)'(0) + \frac{(f \circ \gamma)''(0)}{2} + O(r^3). \quad (9.3)$$

به دلایلی روشن، این رابطه را می‌توان به صورت فرمول تقریب تابع‌های خودسازگار روی فضاهاى متریک ژئودزیک با «تابع‌های درجه دوم» تفسیر کرد؛ گویی، رابطه (۹.۳) سه جمله نخست بسط تیلر f را به دست می‌دهد.

۴. روش بهینه‌سازی نیوتن

مطابق معمول فرض کنید (X, d) یک فضای متریک ژئودزیک کامل، $D \subset X$ یک زیرمجموعه ناتهی، باز و به‌طور یکتا ژئودزیک، و $f : D \rightarrow \mathbb{R}$ یک تابع خودسازگار باشد. برای هر $x \in D$ کمیت زیر را تعریف می‌کنیم

$$\lambda_f(x) = \sup \left\{ \frac{|(f \circ \eta)'(\circ)|}{((f \circ \eta)''(\circ))^{\frac{1}{2}}} \mid \eta \in S_x(X) \right\}. \quad (1.4)$$

توجه داشته باشیم که مقدار کسرها در تعریف فوق، به سرعتِ خم‌های η بستگی ندارد، همچنین f در x مینیمم خود را می‌گیرد اگر و تنها اگر $\lambda_f(x) = 0$. در واقع، این عدد را می‌توان کاهش نیوتنی^۱ تلقی کرد.

قضیه ۱.۴. علاوه بر مفروضات بالا، اگر X موضعاً فشرده باشد و برای حداقل یک $x_0 \in D$ داشته باشیم $\lambda_f(x_0) < 1$ ، آنگاه تابع f دارای یک نقطهٔ مینیمم یکتاست.

اثبات. روشن است که زیرمجموعهٔ $D_0 = \{x \in D \mid f(x) \leq f(x_0)\}$ در D بسته، محدب، و ناتهی است. اگر نشان دهیم که D_0 کراندار نیز هست، بنابه قضیهٔ ۳.۲ فشردگی D_0 نتیجه می‌شود. چون f یک تابع پیوسته است و مقدار مینیمم آن روی D و D_0 یکسان است، وجود نقطهٔ مینیمم تابع f ثابت می‌شود. پس کافی است کرانداری D_0 را ثابت کنیم. برای این منظور فرض کنید $x \in D_0$. بنابر (۵.۳) و (۱.۴) داریم

$$\begin{aligned} f(x_0) &\geq f(x) \geq f(x_0) + (f \circ \gamma_{x,x})'(\circ) + r - \ln(1+r) \\ &\geq f(x_0) - r\lambda_f(x_0) + r - \ln(1+r) \end{aligned}$$

که در آن $r = ((f \circ \gamma_{x,x})''(\circ))^{\frac{1}{2}}$. بنابراین

$$\frac{\ln(1+r)}{r} \geq 1 - \lambda_f(x_0) > 0.$$

اما

$$\lim_{r \rightarrow +\infty} \frac{\ln(1+r)}{r} = 0,$$

^۱Newton decrement

پس ضرورتاً یک عدد ثابت و مثبتی مانند K وجود دارد به طوری که برای هر $x \in D_0$ داشته باشیم

$$((f \circ \gamma_{x,x})''(\circ))^\frac{1}{r} \leq K.$$

از آنجایی که D_0 محدب است، با فرض بی کران بودن آن، یک ژنودزی $\eta : [\circ, +\infty] \rightarrow D_0$ وجود دارد، که بدون کاستن از کلیت، می توان فرض کرد سرعت آن برابر واحد باشد. در این صورت برای هر $t > \circ$ داریم $\eta^t = \gamma_{\eta(\circ)\eta(t)}$ و بنابراین

$$t((f \circ \eta)''(\circ))^\frac{1}{r} = ((f \circ \eta^t)''(\circ))^\frac{1}{r} \leq K,$$

ولی، غیرممکن است این رابطه برای همه اعداد مثبت t برقرار باشد. پس D_0 کراندار است. یکتایی نقطهٔ مینیم تقریباً واضح است. اگر $x^*, x^{**} \in D$ دو نقطهٔ متمایز و هر دو نقطهٔ مینیم تابع f باشند، آنگاه با استفاده از قضیهٔ مقدار میانگین برای تابع حقیقی $t \mapsto (f \circ \gamma_{x^*x^{**}})'(t)$ وجود عددی مانند $t \in (\circ, 1)$ ثابت می شود که

$$(f \circ \gamma_{x^*x^{**}})''(t) = (f \circ \gamma_{x^*x^{**}})'(1) - (f \circ \gamma_{x^*x^{**}})'(\circ) = \circ.$$

ولی این مطلب با اکیداً محدب بودن f در تناقض است. □

مشابه با روش کلاسیک، در اینجا هم با کاهش عدد $\lambda_f(x)$ به نقطهٔ مینیم نزدیک می شویم. در واقع، به طور دقیق تر قضیهٔ زیر را داریم.

قضیه ۲.۴. فرض کنید تابع f در نقطهٔ $x^* \in D$ مینیم خود را بگیرد، $x \in W^\circ(x^*; 1)$ و $\lambda_f(x) < 1$ در این صورت داریم

$$((f \circ \gamma_{xx^*})''(\circ))^\frac{1}{r} \leq \frac{\lambda_f(x)}{1 - \lambda_f(x)}, \tag{۲.۴}$$

$$\circ \leq f(x) - f(x^*) \leq -\lambda_f(x) - \ln(1 - \lambda_f(x)). \tag{۳.۴}$$

اثبات. بنابه فرض $r = ((f \circ \gamma_{xx^*})''(\circ))^\frac{1}{r} \in (\circ, 1)$ بنابر (۴.۳) و (۷.۳) داریم

$$\circ \leq \frac{r^2}{1+r} \leq -(f \circ \gamma_{xx^*})'(\circ) \leq \frac{r^2}{1-r}$$

اما $|-(f \circ \gamma_{xx^*})'(\circ)| \leq r \lambda_f(x)$ ، پس $r/(1+r) \leq \lambda_f(x)$ که همان رابطهٔ (۲.۴) است.

اکنون از (۱.۴) و (۵.۳) نتیجه می‌شود که

$$\begin{aligned} f(x^*) - f(x) &\geq (f \circ \gamma_{xx^*})'(\circ) + r - \ln(1+r) \\ &\geq r - \ln(1+r) - r\lambda_f(x). \end{aligned}$$

از طرف دیگر، به سادگی می‌توان دید که برای هر دو عدد حقیقی $a \geq \circ$ و $\circ \leq b < 1$ داریم

$$a - \ln(1+a) - ab \geq b + \ln(1-b).$$

بنابراین

$$f(x^*) - f(x) \geq \lambda_f(x) + \ln(1 - \lambda_f(x))$$

□ که این نیز همان رابطه (۳.۴) است.

فرض کنید $x \in D$ و نیز $\eta \in S_x(X)$ راستایی با سرعت واحد باشد که مقدار ماکسیم

$$(1.4) \text{ در آن اختیار شود، یعنی } \lambda_f(x) = \frac{|(f \circ \eta)'(\circ)|}{((f \circ \eta)''(\circ))^{\frac{1}{2}}}. \text{ می‌نویسیم}$$

$$\tau = \frac{|(f \circ \eta)'(\circ)|}{(f \circ \eta)''(\circ)}$$

و تعریف می‌کنیم $\gamma_x = \eta^\tau$. توجه داشته باشید که خم γ_x لزوماً یکتا نیست.

لم ۳.۴. برای هر $x \in D$ رابطه زیر میان کمیت $\lambda_f(x)$ و راستای γ_x برقرار است

$$\lambda_f^2(x) = |(f \circ \gamma_x)'(\circ)| = (f \circ \gamma_x)''(\circ).$$

به علاوه، همیشه می‌توان γ_x را طوری انتخاب کرد که $\lambda_f^2(x) = -(f \circ \gamma_x)'(\circ)$.

اثبات. با استفاده از تعریف γ_x و رابطه (۲.۳) داریم

$$\lambda_f(x)^2 = \frac{((f \circ \eta)'(\circ))^2}{(f \circ \eta)''(\circ)} = \tau |(f \circ \eta)'(\circ)| = |(f \circ \gamma_x)'(\circ)|.$$

از طرف دیگر،

$$(f \circ \gamma_x)''(\circ) = \tau^2 (f \circ \eta)''(\circ) = \tau |(f \circ \eta)'(\circ)| = |(f \circ \gamma_x)'(\circ)|.$$

برای اثبات حکم آخر، توجه می‌کنیم که بنابر تعریف، خم γ_x روی یک بازه مانند $(-\delta, \delta)$ با شرط $\delta > \circ$ تعریف شده است. اکنون اگر خم $\gamma_x^-(t) = \gamma(-t)$ را وقتی $t \in (-\delta, \delta)$ در نظر بگیریم،

داریم $(f \circ \gamma_x)'(\circ) = -(f \circ \gamma_x)'(\circ)$ و $(f \circ \gamma_x)''(\circ) = (f \circ \gamma_x)''(\circ)$. پس یکی از دو خم γ_x یا γ_x^- ما را به مقصود می‌رساند.

خم γ_x در لم فوق دقیقاً همان نقش گام نیوتنی^۱ در روش کلاسیک را دارد. اکنون الگوریتم روش بهینه‌سازی نیوتن برای تابع‌های خودسازگار روی فضا‌های متریک ژئودزیکِ موضعاً فشرده در الگوریتم ۱ ارائه می‌شود.

الگوریتم ۱ روش نیوتن

(۱) فرض کنید $x_0 \in D$ نقطه‌ای دلخواه و $0 < \epsilon$ مقداری داده‌شده باشد.

(۲) تا زمانی‌که $\epsilon \leq \lambda_f(x_k)$ ، قرار دهید

$$x_{k+1} = \gamma_{x_k} \left(\frac{1}{1 + \lambda_f(x_k)} \right)$$

در قضیه پایانی این مقاله، همگرایی دنباله تولیدشده توسط الگوریتم ۱ را ثابت می‌کنیم و نشان می‌دهیم که برای رسیدن به میزان دقتِ مطلوب، الگوریتم دست‌کم چه تعداد گام باید تکرار شود.

قضیه ۴.۴. برای دنباله‌ای که الگوریتم ۱ تولید می‌کند داریم

$$f(x_{k+1}) \leq f(x_k) - \lambda_f(x_k) + \ln(1 + \lambda_f(x_k)). \quad (۴.۴)$$

به علاوه، اگر تابع f در نقطه x^* مینیمم شود، آنگاه

(۱) این الگوریتم در حداکثر $\lfloor \frac{f(x_0) - f(x^*)}{\epsilon - \ln(1 + \epsilon)} \rfloor$ گام متوقف می‌شود.

(۲) اگر الگوریتم را متوقف نکنیم، برای دنباله نامتناهی ایجادشده داریم

$$\lim_{k \rightarrow +\infty} f(x_k) = f(x^*).$$

(۳) هر زیردنباله همگرا از دنباله $\{x_k\}$ به x^* همگراست. به‌ویژه، اگر X موضعاً فشرده و

$$D \text{ کراندار باشد آنگاه } \lim_{k \rightarrow +\infty} x_k = x^*$$

اثبات. تعریف کنید $t_k = \frac{1}{1 + \lambda_f(x_k)}$. در این صورت، با استفاده از لم ۳.۴، داریم

$$\left((f \circ \gamma_{x_k}^{t_k})''(\circ) \right)^{\frac{1}{2}} = t_k \lambda_f(x_k) < 1,$$

^۱Newton step

بنابراین $x_{k+1} \in W^\circ(x_k; 1)$. اکنون با استفاده مجدد از لم ۳.۴، رابطه (۸.۳) و تعریف t_k ، و رابطه (۲.۳) خواهیم داشت

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + (f \circ \gamma_{x_k}^{t_k})'(\circ) - ((f \circ \gamma_{x_k}^{t_k})''(\circ))^{\frac{1}{\bar{v}}} \\ &\quad - \ln \left(1 - ((f \circ \gamma_{x_k}^{t_k})''(\circ))^{\frac{1}{\bar{v}}} \right) \\ &= f(x_k) + t_k (f \circ \gamma_{x_k})'(\circ) - t_k ((f \circ \gamma_{x_k})''(\circ))^{\frac{1}{\bar{v}}} \\ &\quad - \ln \left(1 - t_k ((f \circ \gamma_{x_k})''(\circ))^{\frac{1}{\bar{v}}} \right) \\ &= f(x_k) - \lambda_f(x_k) + \ln(1 + \lambda_f(x_k)). \end{aligned}$$

برای اثبات (۱)، فرض کنید الگوریتم تا گام n ادامه داشته باشد. بنابر (۴.۴) و صعودی بودن تابع $\lambda \mapsto \lambda - \ln(1 + \lambda)$ روی مجموعه اعداد حقیقی مثبت، داریم

$$\begin{aligned} n(\epsilon - \ln(1 + \epsilon)) &\leq \sum_{k=1}^n \left(\lambda_f(x_{k-1}) - \ln(1 + \lambda_f(x_{k-1})) \right) \\ &\leq \sum_{k=1}^n (f(x_{k-1}) - f(x_k)) \\ &= f(x_\circ) - f(x_n) \leq f(x_\circ) - f(x^*). \end{aligned}$$

اگر در رابطه اخیر، عدد n بتواند به دلخواه بزرگ شود (یعنی، با فرض $\epsilon = \circ$)، سری فوق همگراست. بنابراین

$$\lim_{k \rightarrow +\infty} (\lambda_f(x_k) - \ln(1 + \lambda_f(x_k))) = \circ$$

و در نتیجه $\lim_{k \rightarrow +\infty} \lambda_f(x_k) = \circ$. اکنون حکم (۲) از رابطه (۳.۴) به دست می‌آید. سرانجام، از پیوستگی تابع f و یکتایی نقطهٔ مینیمم، گزارهٔ نخست (۳) و از قضیهٔ ۳.۲ گزارهٔ دوم آن نتیجه می‌شود. \square

تشکر و قدردانی

از آقای دکتر روح‌الله جهانی‌پور به دلیل تشویق نویسنده به نگارش این مقاله و پیشنهادهای ارزشمندی که برای بهبود نگارش آن کردند، بسیار سپاسگزارم.

مراجع

- [1] Absil, P. A., Mahony, R., Sepulchre, R., *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.
- [2] Ahmadi Kakavandi, B., Amini, M., Duality and subdifferential for convex functions on complete CAT(0) metric spaces, *Nonlinear Anal.*, **73** (2010), 3450–3455.
- [3] Ahmadi Kakavandi, B., Weak topologies in complete CAT(0) metric spaces, *Proc. Amer. Math. Soc.*, **141** (2013), 1029–1039.
- [4] Alexander, S., Kapovitch, V., Petrunin, A., *An Invitation to Alexandrov Geometry: CAT(0) spaces*, Springer, Cham, 2019.
- [5] Bačák, M., *Convex Analysis and Optimization in Hadamard Spaces*, De Gruyter, Berlin, 2014.
- [6] de A. Bortoloti, M. A., Fernandes, T. A., Ferreira, O. P. Yuan, J., Damped Newton's method on Riemannian manifolds, *J. Global Optim.*, **77** (2020), 643–660.
- [7] Boyd, S., Vandenberghe, L. , *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
- [8] Bridson, M., Haefliger, A., *Metric Spaces of Nonpositive Curvature*, Springer, Berlin, 1999.
- [9] Burago, D., Burago, Y., Ivanov, S., *A Course in Metric Geometry*, Amer. Math. Soc., Providence, RI, 2001.
- [10] Gromov, M., *Metric Structures for Riemannian and Non-Riemannian Spaces*; reprint of the 2001 English ed., Birkhäuser Boston, Boston, 2007.
- [11] Huang, W., Absil, P. A., Gallivan, K. A., A Riemannian BFGS method for nonconvex optimization problems, In: *Numerical Mathematics and Advanced Applications (ENUMATH 2015)*, Lecture Notes in Computational Science and Engineering, vol. 112, Springer, Cham, 2016, 627–634.
- [12] Fernandes, T. A., Ferreira, O. P., Yuan, J., On the superlinear convergence of Newton's method on Riemannian manifolds, *J. Optim. Theory Appl.*, **173** (2017), 828–843.
- [13] Ji, H., Optimization approaches on smooth manifolds, Ph.D. thesis, Australian National University, Canberra, 2007.
- [14] Jiang, D., Moore, J. B., Ji, H., Self-concordant functions for optimization on smooth manifolds, *J. Global Optim.*, **38** (2007), 437–457.
- [15] Kirk, W. A., Shahzad, N., *Fixed Point Theory in Distance Spaces*, Springer, Cham, 2014.

- [16] Lott, J., Villani, C., Ricci curvature for metric-measure spaces via optimal transport, *Ann. of Math.*, **169** (2009), 903–991.
- [17] Manton, J. H., A framework for generalising the Newton method and other iterative methods from Euclidean space to manifolds, *Numer. Math.*, **129** (2015), 91–125.
- [18] Nesterov, Y., Nemirovskii, A., *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, PA, 1994.
- [19] Nesterov, Y., Todd, M., On the Riemannian geometry defined by self-concordant barriers and interior-point methods, *Found. Comput. Math.*, **2** (2002), 333–361.
- [20] Papadopoulos, A., *Metric Spaces, Convexity and Nonpositive Curvature*, European Mathematical Society, Zürich, 2005.
- [21] Perelman, G., Spaces with curvature bounded below, In *Proceedings of the International Congress of Mathematicians (Zürich, 1994)*, Vol. 1, 2, Birkhäuser, Basel, 1995, 517–525.
- [22] Smith, S., Optimization techniques on Riemannian manifolds, *Fields Inst. Commun.*, **3** (1994), 113–146.
- [23] Sormani, C., Friedmann cosmology and almost isotropy, *Geom. Funct. Anal.*, **14** (2004), 853–912.
- [24] Shalev-Shwartz, S., Ben-David, S., *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, New York, 2014.
- [25] Udrişte, C., *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Academic Publishers Group, Dordrecht, 1994.

تاریخ ارسال: ۱۳۹۹/۳/۳۱؛ تاریخ بازنگری: ۱۳۹۹/۶/۲۳؛ تاریخ پذیرش: ۱۳۹۹/۶/۲۷

بیژن احمدی کاکاوندی: دانشگاه شهید بهشتی، دانشکده علوم ریاضی

تارنما: <http://facultymembers.sbu.ac.ir/bijanahmadi>

رایانامه: b_ahmadi@sbu.ac.ir