چکیده

روش‌های گرادیان یک خانواده مهم از الگوریتم‌ها برای حل مسائل بهینه سازی نامقید هستند که به دلیل رفتار مناسب دور از جواب، استفاده نکردن از مشتقات مرتبه دوم و هزینه محاسباتی کم همواره مورد توجه محققان بوده اند. خواص مناسب این روش‌ها از یک سو و سرعت همگرایی پایین آنها از سوی دیگر باعث شد که محققان زیادی تلاش خود را صرف پیدا کردن روش‌هایی مرتبط با حفظ خواص مناسب و سرعت همگرایی بالاتر کنند. یکی از روش‌های این رده، روش گرادیان نستروف است که اولین بار توسط نستروف ارائه شد. او نشان داد این روش برای توابع محدب دارای همگرایی مجذوری است. در این پایان نامه ضمن بررسی روش گرادیان نستروف یک توسیع از این الگوریتم با افزودن یک رابطه به‌هنگام همراه با یک رابطه ریستارت جدید ارائه می‌شود. نتایج عددی ارائه شده در این پایان نامه نشان دهنده کارایی بهتر الگوریتم جدید نسبت به الگوریتم گرادیان نستروف است.

کلمات کلیدی: بهینه سازی نامقید، روش گرادیان، روش گرادیان نستروف، همگرایی سراسری، رابطه سکانت

فهرست مطالب

عنوان **صفحه**

فصل چهارم: بررسی پایداری یک روش تندترین کاهش با اندازه حرکت برای توابع درجه دوم

فصل پنجم: نتایج عددی

فهرست اشکال

عنوان **صفحه**

فهرست جداول

عنوان **صفحه**

پیش گفتار

بهینه سازی غیرخطی نامقید یکی از مهمترین زیر شاخه‌های بهینه سازی است که بسیار مورد توجه بوده و تا کنون روش‌های زیادی برای حل این مسائل ارائه شده است. از جمله مهمترین این روش‌ها، روش گرادیان است که به دلیل رفتار مناسب دور از جواب، هزینه محاسباتی کم و استفاده نکردن از مشتقات مراتب بالاتر همواره مورد توجه بوده است این روش دارای معایبی نیز می‌باشد که از جمله آنها می‌توان به سرعت همگرایی پایین آنها اشاره کرد. این امر باعث گردیده محققان زیادی تلاش خود را صرف پیدا کردن روش‌هایی با حفظ خواص مناسب و سرعت همگرایی بالاتر کنند. از میان روش‌های به دست آمده می توان به روش گرادیان نستروف اشاره کرد که اولین بار توسط نستروف ارائه شد. او نشان داد این روش برای توابع محدب دارای همگرایی مجذوری است. در این پایان نامه ضمن بررسی روش گرادیان نستروف و اثبات همگرایی مجذوری آن یک توسیع از این الگوریتم ارائه می‌شود که با افزودن یک رابطه به‌هنگام همراه با یک رابطه ریستارت به الگوریتم گرادیان نستروف به دست می‌آید. در بخش انتهایی این پایان نامه یک توسیع از الگوریتم تندترین کاهش با افزودن یک پارامتر حرکت ارائه و پایداری آن برای توابع مجذوری مورد بررسی قرار می گیرد.

این پایان نامه شامل پنج فصل است.

در فصل اول مقدمات، تعاریف و برخی از روش‌های تکراری برای حل مسئله بهینه سازی نامقید مورد بررسی قرار می گیرد.

در فصل دوم پس از یک مرور مختصر بر برخی روش‌های خانواده گرادیان، به بررسی روش گرادیان نستروف پرداخته می‌شود. در این فصل پس از بیان اصل نستروف ، با استفاده از پارامترهای دنباله تخمین و شتاب دهنده یک دنباله تخمینی از تابع هدف ساخته می‌شود و نشان داده می‌شود که اگر این دنباله در اصل نستروف صادق باشد روش گرادیان نستروف دارای همگرایی سراسری و مجذوری برای توابع محدب است.

در فصل سوم با تحمیل شرط سکانت روی نقطه جستجو، پارامتر دنباله تخمین به‌هنگام می‌شود و با افزودن این رابطه به‌هنگام و یک رابطه ریستارت به الگوریتم گرادیان نستروف، یک الگوریتم نستروف اصلاح شده و مبتنی بر رابطه سکانت ارائه می‌شود. بررسی خواص همگرایی این الگوریتم اصلاح شده هدف بعدی این فصل است.

در فصل چهارم یک الگوریتم جدید که توسیعی از روش تندترین کاهش است و با افزودن یک پارامتر حرکت به آن به دست می‌آید ارائه شده و شرایط لازم و کافی برای پایداری این روش برای توابع مجذوری بررسی می‌شود.

فصل پنجم این پایان‌نامه به بررسی نتایج عددی الگوریتم گرادیان نستروف اصلاح شده در مقایسه با الگوریتم پایه اختصاص دارد.

فصل اول

مقدمه ای بر بهینه سازی غیر خطی

۱-۱ مقدمه

آدمی در طول دوران زندگی خود، هر روز و هرساعت با مسائلی مواجه می‌شود که با توجه به محدودیت-های زمانی و نیز شرایط زندگی خود مجبور به انتخاب می‌شود و در نهایت سعی می‌کند بهترین انتخاب را از میان گزینه‌های موجود داشته باشد. این میل به انتخاب بهترین گزینه به دنیای ریاضیات نیز وارد شده است و باعث پیدایش شاخه ای از ریاضیات به نام بهینه سازی شده است.

مدل سازی که شامل تعیین هدف، متغیرها و محدودیت‌های یک مسئله است اولین و موثرترین گام در بهینه سازی است. پس از آن باید به دنبال پیداکردن یک الگوریتم مناسب برای حل مسئله بود. اگرچه الگوریتم های بسیاری برای مسائل بهینه سازی ارائه شده است اما هیچ الگوریتم جامعی که از هر نظر مناسب باشد وجود ندارد و هر روش با وجود مزایایی که دارد به نوبه خود معایبی را هم دارد.

با توجه به نوع مسئله، تابع هدف و محدودیت‌ها مسئله بهینه سازی دسته بندی‌ها و ودر نتیجه روش‌های متفاوتی را شامل می‌شود. بهینه سازی نامقید یکی از شاخه‌های این گرایش می‌باشد هدف نهایی این پایان-نامه ارائه الگوریتم‌هایی برای حل این نوع مسائل می‌باشد.

۱-۲ مسائل بهینه سازی نامقید

شکل عمومی یک مسئله بهینه سازی نامقید به صورت زیر است:

$$\min_{x\in R^n} f(x) \tag{۱-۲-۱}$$

برای یافتن جواب بهینه مسئله فوق نیاز به تعاریف مقدماتی زیر داریم:

تعریف ۱-۲-۱. (گرادیان تابع) [۲۲] فرض کنید $f: R^n \to R$ به طور پیوسته مشتق پذیر باشد. در هر نقطه $x\epsilon R^n$ یک برداراز مشتقات جزیی مرتبه اول تابع f وجود دارد که بردار گرادیان نامیده می شود وبه صورت زیر تعریف می‌گردد:

$$g(x) = \nabla f(x) = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})^T. \tag{۱-۲-۲}$$

تعریف ۱-۲-۲. [۱۲] فرض کنید $f: R^n \to R$ مشتق پذیر باشد. گرادیان f را پیوسته لیپ شیتز با ثابت L گویند هرگاه

$$\exists L > 0 \ \ s.t \quad \|\nabla f(x) - \nabla f(y)\| \le L\|x - y\| \qquad \forall x , y \in R^n.$$

تعریف ۱-۲-۳. (ماتریس هسی تابع) [۲۲] فرض کنید $f: R^n \to R$ در نقطه $x \in R^n$ دو بار به طور پیوسته مشتق پذیر باشد. آن گاه یک ماتریس شامل همه مشتقات جزیی مرتبه دوم تابع f وجود دارد که ماتریس هسیان نامیده می‌شود. این ماتریس، ماتریس مربعی و متقارن است که درایه (i , j) آن به صورت $\frac{\partial^2 f}{\partial x_i \partial x_j}$ تعریف می‌شود. در واقع ماتریس هسیان به صورت زیر است :

$$\nabla^2 f(x) = \nabla(\nabla f(x))^T = \nabla\left(\frac{\partial f}{\partial x_1}, \dots \frac{\partial f}{\partial x_n}\right) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

از این به بعد در جهت سهولت، اگر x_k عنصر kام دنباله و x^* جواب می‌نیمم (۱-۲-۱) باشد آن‌گاه $f(x_k)$ را با f_k، $\nabla f(x_k)$ را با g_k، $\nabla^2 f(x_k)$ را با G_k و $f(x^*)$ را با f^* نمایش می‌دهیم.

هدف از حل یک مسئله بهینه سازی نامقید یافتن می‌نیمم سراسری تابع f است اما این هدف معمولاً عملی نیست بنابراین به نقطه می نیمم نسبی اکتفا می‌شود. تحت شرایط تحدب تابع f می توان اثبات کرد که هرنقطه می‌نیمم موضعی یک می‌نیمم سراسری نیز هست.

در ادامه به برخی تعاریف تحدب می‌پردازیم.

تعریف ۱-۲- ۴ [۱۵]

الف) مجموعه محدب. مجموعه $S \subseteq R^n$ را یک مجموعه محدب گویند اگر برای هر $x , y \in S$ و $\alpha \in [0 , 1]$ داشته باشیم:

$$\alpha x + (1 - \alpha) y \in S .$$

ب) تابع محدب. تابع f تعریف شده روی مجموعه محدب و ناتهی $S \subseteq R^n$ را محدب گویند هرگاه

$$\forall x , y \in S , \alpha \in [0 , 1] \quad : \quad f(\alpha x + (1 - \alpha) y) \le \alpha f(x) + (1 - \alpha) f(y) ,$$

و f را اکیداً محدب گوییم هرگاه

$$\forall\, x , y \in S , \alpha \in [0 , 1] \ , x \neq y \ \ f(\alpha x + (1 - \alpha) y) < \alpha f(x) + (1 - \alpha) f(y).$$

ج) تابع به طور یکنواخت محدب. تابع f تعریف شده روی مجموعه محدب و ناتهی $S \subseteq R^n$ را به طور یکنواخت محدب است اگرثابت $\mu > 0$ وجود داشته باشد به طوریکه برای همه مقادیر $x \,, y \in S$ و $\alpha \in (0 \,, 1)$ داشته باشیم :

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{1}{2}\mu\alpha(1-\alpha)\|x-y\|^2.$$

لم ۱-۲-۱.[۲۲] فرض کنید $S \subseteq R^n$ یک مجموعه محدب و تابع $f\colon S \to R$ به طور پیوسته مشتق پذیر باشد

۱) تابع f محدب است اگروتنها اگر

$$\forall\, x \,, y \in S \;:\quad f(y) \geq f(x) + \nabla f(x)^T(y-x).$$

۲) تابع f اکیداً محدب است اگروتنها اگر

$$\forall\, x \,, y \in S \;\;, x \neq y \quad:\quad f(y) > f(x) + \nabla f(x)^T(y-x).$$

۳) تابع f به طور یکنواخت محدب است اگروتنها اگر

$$\forall\, x \,, y \in S \;:\; f(y) \geq f(x) + \nabla f(x)^T(y-x) + \tfrac{1}{2}\mu\|y-x\|^2.$$

که در آن $\mu > 0$ یک عدد ثابت است.

لم۱-۲-۲. [۲۳] تابع به طور پیوسته مشتق پذیر f روی مجموعه محدب و ناتهی $S \subseteq R^n$ به طور یکنواخت محدب است اگرو تنها اگر

$$\forall\, x \,, y \in S \;:\; (\nabla f(x) - \nabla f(y))^T(x-y) \geq \mu\|x-y\|^2.$$

مفهوم به طور یکنواخت محدب قویتر از مفهوم اکیداً محدب است در واقع هر تابع به طور یکنواخت محدب یک تابع اکیداً محدب نیز هست اما عکس آن درست نیست.

تعریف ۱-۲-۵.[۱۲] فرض کنید L ثابت لیپ‌شیتز و μ پارامتر تحدب باشد آن‌گاه $\mathcal{F}_L^{1,1}(R^n)$ دسته ای از توابع محدب روی R^n است که گرادیان آنها پیوسته لیپ‌شیتز می‌باشد و $\mathcal{S}_{\mu,L}^{1,1}(R^n)$ دسته ای از توابع به طور یکنواخت محدب روی R^n است که گرادیان آنها پیوسته لیپ شیتزمی‌باشد.

قضیه ۱-۲-۱. [۱۲] فرض کنید $f \epsilon \mathcal{F}_L^{1,1}(R^n)$ آن‌گاه برای هر $x , y \epsilon R^n$ داریم:

$$f(y) \leq f(x) + (y-x)\nabla f(x) + \frac{L}{2}\|y-x\|^2.$$

در ادامه به بیان قضیه تیلور می‌پردازیم که یکی از ابزارهای مهم در بهینه سازی محسوب می‌شود.

قضیه ۱-۲-۲. [۱۵] فرض کنید f یک تابع حقیقی بر $[a , b]$ و n یک عدد طبیعی باشد. اگر f روی $[a , b]$ ، n بار به طور پیوسته مشتق‌پذیر باشد و $f^{(n+1)}$ بر روی (a , b) موجود باشد آن‌گاه برای هر $x , c \epsilon [a , b]$ داریم:

$$f(x) = \sum_{k=0}^{n} \frac{f^k(c)}{k!}(x-c)^k + E_n(x)$$

که در آن

$$E_n(x) = \frac{1}{(n+1)!} f^{n+1}(\xi)(x-c)^{n+1} \quad , \quad \xi \epsilon (x,c).$$

تعریف ۱-۲-۶. [۲۲] فرض کنید A یک ماتریس متقارن باشد در این صورت A را

الف) نیمه معین مثبت گوییم هر گاه به ازای هر $y \epsilon R^n$ ، $y^T A y \geq 0$.

ب) معین مثبت گوییم هر گاه به ازای هر $y \epsilon R^n$ ، $y^T A y > 0$.

تعریف ۱-۲-۷ .[۲۲] فرض کنید A یک ماتریس مربعی، λ یک اسکالر و $x \epsilon R^n$ یک بردار ناصفرباشد معادله برداری

$$Ax = \lambda x$$

را در نظر بگیرید. مقداری از λ را که به ازای آن معادله فوق دارای جواب ناصفر باشد مقدار ویژه ماتریس A و بردار x بردار ویژه ماتریس A، متناظر با این مقدار ویژه است.

نکته: همه مقادیر ویژه یک ماتریس متقارن اعداد حقیقی هستند.

قضیه ۱-۲-۴. [۲۲] ماتریس متقارن حقیقی A را در نظر بگیرید این ماتریس

الف) نیمه معین مثبت است اگر وتنها اگر همه مقادیر ویژه آن نامنفی باشند.

ب) معین مثبت است اگر و تنها اگر همه مقادیر ویژه آن مثبت باشند.

۱–۳ الگوریتم‌های بهینه سازی

اغلب الگوریتم‌هایی که برای مسائل بزرگ بهینه سازی مطرح شده اند از نوع تکراری هستند. در روند جستجو برای یافتن نقطه ای که مسئله بهینه سازی را حل کند معمولاً یک نقطه اولیه انتخاب می شود سپس در یک فرآیند تکراری به دنبال امکان بهبود جواب در هر مرحله هستیم. با ادامه این روند دنباله ای متوالی از نقاط به دست می‌آید که به نقطه جواب نزدیک می‌شود.

برای مسائل بهینه سازی خطی، دنباله تولید شده متناهی است و الگوریتم پس از تعداد متناهی مرحله به نقطه جواب می‌رسد. برای مسائل بهینه سازی غیر خطی، این دنباله در حالت کلی به یک نقطه جواب نمی-رسد اما به آن همگرا می‌شود بنابراین در عمل هنگامی‌که نقطه ایی به اندازه کافی نزدیک به نقطه جواب تولید گردد این فرآیند متوقف می‌شود.

مقایسه روش‌ها بر حسب هزینه محاسباتی جملات متوالی دنباله، نرخ همگرایی و تضمین همگرایی این دنباله می‌باشد.

در ادامه به تعریف مرتبه همگرایی دنباله که رابطه مستقیمی با سرعت همگرایی دارد می‌پردازیم.

تعریف ۱–۳–۱. [۱۵] فرض کنید $\{x_k\}_{k=1}^{\infty}$ دنباله ای در R^n و همگرا به x^* باشد

۱) همگرایی دنباله را Q – خطی گویند هرگاه یک ثابت $r \in (0,1)$ وجود داشته باشد به طوریکه برای هر k به اندازه کافی بزرگ داشته باشیم:

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq r$$

این بدان مفهوم است که در هر تکرار فاصله جواب از x^* توسط یک عامل ثابت کاهش می‌یابد.

۲) همگرایی دنباله را Q – زبرخطی گویند هرگاه

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0 \,.$$

۳) همگرایی را Q – مجذوری گویند هرگاه برای هر k به اندازه کافی بزرگ داشته باشیم:

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq M,$$

که در آن M یک ثابت مثبت است.

تعریف ۱–۳–۲. [۱۵] (تعریف کلی) فرض کنید $\{x_k\}$ دنباله ای در R^n و همگرا به x^* باشد. مرتبه همگرایی این دنباله را p^* گویند هرگاه

$$p^* = \sup\{\, p \,:\, \beta = \limsup \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} < \infty\}\,.$$

تعریف ۱–۳–۳. [۱۵] فرض کنید $\{x_k\}$ دنباله ای در R^n و همگرا به x^* باشد و یک دنباله از اسکالرهای نامنفی $\{\upsilon_k\}$ وجود داشته باشد به طوریکه

$$\forall\, k \epsilon N \;:\; \|x_k - x^*\| \leq \upsilon_k$$

آنگاه

۱) همگرایی $\{x_k\}$ را R– خطی گوییم هرگاه $\{\upsilon_k\}$ یک دنباله همگرای Q– خطی به صفر باشد.

۲) همگرایی $\{x_k\}$ را R– زبرخطی گوییم هرگاه $\{\upsilon_k\}$ یک دنباله همگرای Q– زبرخطی به صفر باشد.

۳) همگرایی $\{x_k\}$ را R– مجذوری گوییم هرگاه $\{\upsilon_k\}$ یک دنباله همگرای Q– مجذوری به صفر باشد.

مثال: دنباله زیر همگرای R– خطی به یک است.

$$x_k = \begin{cases} 1 + (\frac{1}{2})^k & \text{زوج } k \\ 1 & \text{فرد } k \end{cases}$$

واضح است که R– همگرایی نسبت به Q– همگرایی ضعیف تر است.

تعریف ۱–۳–۴. تابع $g(\alpha)$ را از مرتبه $o(\alpha)$ گوییم هرگاه رابطه زیر برقرار باشد :

$$\lim_{\alpha \to 0} \frac{g(\alpha)}{\alpha} = 0,$$