بسم الله الرحمن الرحیم

١٤٣١٤٠-٢٠٢٠٥٠٧

۱۳۲۶
دانشگاه تبریز

دانشکده‌ی علوم ریاضی
گروه ریاضی کاربردی

پایاننامه

برای دریافت درجه‌ی کارشناسی ارشد در رشته‌ی
ریاضی کاربردی، گرایش تحقیق در عملیات

عنوان

روش‌های گاوس-نیوتن تقریبی برای حل مسئله‌ی کمترین مربعات غیرخطی

اساتید راهنما

دکتر میرکمال میرنیا

دکتر صداقت شهراد

پژوهشگر

حسین علی پور

شهریور ۱۳۸۹

تقدیم به

بهترین حامی زندگی ام، مادر مهربانم «خدیجه»،

بهترین ارمغان زندگی ام، «آرزو»

و تقدیم به

همه ی آنان که دوستشان دارم.

سپاس‌گزاری

در تهیه و تدوین این پایاننامه افراد زیادی نقش داشته‌اند که صمیمانه از همه‌ی آنان تشکر می‌کنم. لیک ذکر نام کسانی را که نقش بسزایی در کار به اتمام رسیدن این پایاننامه داشته و در این چند سال مرا یاری کرده‌اند، بر خود واجب می‌دانم.

از جناب آقای دکتر میرکمال میرنیا و جناب آقای دکتر صداقت شهمراد که اساتید راهنمای اینجانب بوده و با راهنمایی‌های خوبشان همواره مرا یاری نموده‌اند کمال تشکر را می‌نمایم. همچنین از جناب آقای دکتر جواد مهری که با راهنمایی‌های بی دریغشان بنده را مورد لطف خویش قرار داده‌اند تشکر می‌کنم.

از خانم پروفسور *Nichols* و آقای دکتر *Lawless* به خاطر همکاری که با بنده داشتند، ممنون هستم.

از کارمندان دانشکده‌ی ریاضی خانم‌ها ایزان و زحمتی مسئولین محترم کتابخانه و آقایان امراپور، معصومی، شریفی و خدایی به خاطر همکاری صمیمانه‌اشان سپاسگزارم.

از دوستان خوبم بویژه هم اتاقی‌هایم آقایان نظری، محمدی و ابدالی وهمچنین آقایان مروتی، رضا زاده و قادریان که در طول دوره‌ی کارشناسی ارشد به عنوان خانواده‌ی دوم من بوده‌اند تشکر می‌کنم.

از خانواده‌ی عزیزم که همیشه پشت و پناهم بوده‌اند و بویژه مادر رنج دیده‌ام که با رنج خویش راه را برای من هموار کرد بسیار سپاسگزارم.

زحمات و کمک‌های این عزیزان هیچ گاه از یاد من نخواهد رفت. از صمیم قلب از همه‌ی ایشان ممنونم. امید است که توانسته باشم از کمک‌های آنان بهره‌ی کافی برده و کار مفید و درخور توجهی ارائه کرده باشم.

حسین علی پور

مرداد ۱۳۸۹

نام خانوادگی: علی پور	نام: حسین

عنوان پایاننامه: روش‌های گاوس-نیوتن تقریبی برای حل مسئله‌ی کمترین مربعات غیرخطی

اساتید راهنما: دکتر میر کمال میرنیا و دکتر صداقت شهمراد

مقطع تحصیلی: کارشناسی ارشد	رشته: ریاضی کاربردی	گرایش: تحقیق در عملیات

دانشگاه: تبریز	دانشکده: علوم ریاضی

تاریخ فارغ‌التحصیلی: شهریور ۱۳۸۹	تعداد صفحه: ۷۰

کلیدواژه‌ها: بهگزینی، روش‌های تقریبی، روش گاوس-نیوتن، مسئله‌ی کمترین مربعات

چکیده

الگوریتم گاوس - نیوتن، یک روش تکراری است که‌معمولاً برای حل مسئله‌ی کمترین مربعات غیرخطی مورد استفاده قرار می‌گیرد. این الگوریتم بویژه برای مسائل برازش داده‌ها در ابعاد بسیار بزرگ همانند تحلیل داده‌های تغییراتی که از پیش‌بینی وضع جوی و اقیانوسی به وجود می‌آیند، خیلی مناسب است. این روند شامل یک دنباله از تقریب‌های کمترین مربعات خطی برای مسائل غیرخطی است، که هرکدام از آنها بایک روند مستقیم یا تکراری حل می‌شود. در مقایسه با روش نیوتن و گونه‌های مختلف آن، این الگوریتم جالب توجه است زیرا احتیاجی به برآورد مشتق مرتبه‌ی دوم که در هسیان تابع هدف ظاهر می‌شود، ندارد. در عمل، استفاده از روش گاوس - نیوتن در پیش بینی‌های هواشناسی، از نظر زمانی پر هزینه است به همین خاطر تقریب‌های گوناگونی برای کاهش زمان محاسبات ایجاد شده است. در این پایاننامه همگرایی روش گاوس - نیوتن را برای دو نوع از تقریب‌ها بررسی می‌کنیم که عبارتند از روش‌های **گاوس - نیوتن قطع شده** و روشهای **گاوس - نیوتن پریشیده**. در این پایاننامه، شرایطی را بدست می‌آوریم که تضمین می‌کند روشهای گاوس - نیوتن قطع شده و پریشیده، تحت آن شرایط همگرا می‌شوند و همچنین سرعت همگرایی نقاط تکرار را بدست می‌آوریم. نتایج با یک مثال عددی ساده توضیح داده می‌شود.

فهرست مطالب

فصل ۱
تعاریف

شعاع طیفی: شعاع طیفی برای یک ماتریس مربعی $A_{n\times n}$ را با $\rho(A)$ نشان می‌دهند و به صورت

$$\rho(A) = \max_i |\lambda_i|$$

تعریف می‌شود که λ_i، $i = 1, ..., n$ مقادیر ویژه‌ی ماتریس A هستند.

نرم برداری: فرض کنید

$$x = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{pmatrix}$$

یک بردار در فضای $\mathbb{R}^n$ باشد. اگر p را عددی حقیقی و ناکمتر از ۱ در نظر بگیریم آنگاه نرم p بردار x عبارت است از

$$||x||_p = (|x_1|^p + ... + |x_n|^p)^{1/p}\,.$$

نرم ماتریسی: ماتریس $A_{m\times n}$، نرم برداری $||.||_p$ و $x \in \mathbb{R}^n$ را در نظر بگیرید. نرم p ماتریس A عبارت است از

$$||A||_p = \max_{x\neq 0} \frac{||Ax||_p}{||x||_P}.$$

می‌توان نشان داد که

$$||A||_2 = \sqrt{\rho(A^T A)}.$$

گرادیان: برای تابع مشتق پذیری مانند $f(x)$ در نقطه‌ی x بردار مشتقات نسبی اول یا بردار گرادیان به صورت

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{x_1} \\ \frac{\partial f}{x_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f}{x_n} \end{bmatrix}$$

تعریف می‌شود که ∇ نشان دهنده‌ی عملگر گرادیان به شکل زیر است

$$\nabla = \left(\frac{\partial x}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n}\right).$$

ماتریس هسی: اگر $f(x)$ دو بار به طور پیوسته مشتق پذیر باشد یعنی $f \in C^2$ آنگاه ماتریس مشتقات نسبی مرتبه‌ی دوم موجود است که به آن ماتریس هسی می‌گویند و با نماد $\nabla^2 f(x)$ نشان می‌دهند و برای هر i, j، مؤلفه‌ی (i, j)ام به صورت

$$\frac{\partial^2 f}{\partial x_i \partial xj}$$

تعریف می‌شود. ماتریس هسی یک ماتریس مربعی و متقارن است.

مقادیر منفرد: A را یک ماتریس $m \times n$، $m \geq n$، در نظر بگیرید. مقادیر ویژه‌ی ماتریس متقارن $A^T A$ حقیقی و نامنفی هستند. فرض کنید این مقادیر ویژه با σ_i^2، $i = 1, 2, ..., n$ نشان داده شوند که $\sigma_1^2 \geq \sigma_2^2 \geq .... \geq \sigma_n^2$ در این صورت σ_1، σ_2، . . . و σ_n را مقادیر منفرد ماتریس A می‌نامند.

هر ماتریس $m \times n$ مانند A را می‌توان به صورت

$$A = UDV^T$$

تجزیه کرد که ماتریس‌های $U_{m\times m}$ و $V_{n\times n}$ متعامد بوده و D یک ماتریس قطری $m \times n$ است. این تجزیه را تجزیه‌ی مقدار منفرد نامیده و با SVD [۱] نشان می‌دهند. مقادیر منفرد σ_i، $i = 1, 2, ..., n$ درایه‌های روی قطر ماتریس D هستند. تعداد مقادیر نامنفرد غیر صفر با رتبه‌ی ماتریس A برابر است [۵].

ستون‌های U را بردارهای منفرد چپ و ستون‌های V را بردارهای منفرد راست ماتریس A می‌نامند.

دستگاه فرا معین: یک دستگاه از معادلات را فرا معین گویند هرگاه تعداد معادلات در آن بیشتر از تعداد مجهولات باشد.

نگاشت انقباض: یک نگاشت $G : D \subset \mathbb{R}^n \to \mathbb{R}^n$ روی مجموعه‌ی $D_0 \subset D$ یک نگاشت انقباض است هرگاه یک عدد مثبت مانند $\alpha < 1$ وجود داشته باشد طوری که به ازای هر $x, y \in D_0$ داشته باشیم

$$||G(x) - G(y)|| \leq \alpha ||x - y||.$$

روند رانگ-کوتا[۲]: معادله‌ی دیفرانسیل با شرایط اولیه‌ی

$$\frac{dy}{dt} - f(y,t) = 0, \quad t \in [0,T], \quad y(0) = y_0 \tag{۱.۱}$$

را در نظر بگیرید. روند رانگ-کوتای مرتبه‌ی دو برای حل این مسئله به صورت

$$k_1 = f(y_n, t_n),$$

$$k_2 = f(y_n + \triangle t f(y_n, t_n), t_n + \triangle t),$$

$$y_{n+1} = y_n + \frac{\triangle t}{2}(k_1 + k_2)$$

است که در آن $t_n = n\triangle t$ و $y_n = y(t_n)$. برای مسائل غیرخطی، معمولاً تابع f به طور صریح به t بستگی ندارد و بنابراین روند رانگ-کوتای مرتبه‌ی دو به صورت

$$k_1 = f(y_n),$$

$$k_2 = f(y_n + \triangle t f(y_n)),$$

$$y_{n+1} = y_n + \frac{\triangle t}{2}(k_1 + k_2)$$

[۱] Singular Value Decomposition
[۲] Runge-Kutta

در می‌آید.

انواع نرخ همگرایی: دنباله‌ی $\{x_n\} \subset \mathbb{R}^n$ و نقطه‌ی $x^* \in \mathbb{R}^n$ را در نظر بگیرید. در این صورت برای nهای به قدر کافی بزرگ گوییم

- $\{x_n\}$ به طور **خطی** با عامل $\sigma \in (0,1)$ به x^* همگرا می‌شود هرگاه

$$\|x_{n+1} - x^*\| \leq \sigma \|x_n - x^*\|.$$

- نرخ همگرایی $\{x_n\}$ به x^* **مرتبه‌ی دو** است هرگاه یک عدد ثابت و مثبت مانند K وجود داشته باشد به طوری که

$$\|x_{n+1} - x^*\| \leq K \|x_n - x^*\|^2.$$

- نرخ همگرایی $\{x_n\}$ به x^* **فوق خطی با مرتبه‌ی** $\alpha > 1$ است هرگاه یک عدد ثابت و مثبت مانند K وجود داشته باشد به طوری که

$$\|x_{n+1} - x^*\| \leq K \|x_n - x^*\|^\alpha.$$

- نرخ همگرایی $\{x_n\}$ به x^* **فوق خطی** است هرگاه داشته باشیم

$$\lim_{n\to\infty} \frac{\|x_{n+1} - x^*\|}{\|x_n - x^*\|} = 0.$$

فصل ۲

مقدمه

روش کمترین مربعات در سال ۱۷۹۵ توسط گاوس معرفی شد. از آن زمان تا کنون از این روش به عنوان ابزار مهمی برای کاهش تأثیر خطاها در برازش قالب‌ها با داده‌های مشاهده شده استفاده می‌شود.

مسئله‌ی کمترین مربعات خطی در ابتدا با نیاز به تطبیق یک قالب ریاضی با یک دسته از داده‌های مشاهده شده بوجود آمد. برای کاهش تأثیر خطای مشاهدات، از روش تعداد بیشتر اندازه‌گیری نسبت به تعداد پارامترهای مجهول در قالب، استفاده می‌شود. اگر قالب خطی باشد آنگاه مسئله‌ی حاصل منجر به حل یک دستگاه خطی فرا معین می‌شود. شکل ماتریسی معادل این مسئله به این صورت است که با در دست داشتن یک بردار $b \in \mathbb{R}^m$ و یک ماتریس $A \in \mathbb{R}^{m\times n}$ و $m > n$، هدف تعیین یک بردار $x \in \mathbb{R}^n$ است به نحوی که Ax بهترین تقریب برای b باشد.

راه‌های زیادی برای تعریف بهترین جواب وجود دارد. انتخابی که با دلایل آماری توجیه و منجر به یک مسئله‌ی محاسباتی ساده می‌شود این است که x یک جواب مسئله‌ی

$$\min_x \|Ax - b\|_2 \qquad (۱.۲)$$

باشد که $\|.\|_2$ نشان دهنده‌ی نرم برداری اقلیدسی است. این مسئله را مسئله‌ی کمترین مربعات خطی نامیده و x را یک جواب کمترین مربعات خطی برای دستگاه $Ax = b$ گویند. همچنین از $r(x) = b - Ax$ به عنوان بردار

مانده اسم می‌بریم. بنابراین یک جواب کمترین مربعات، مقدار $||r||_2^2 = \sum_{i=1}^{m} r_i^2$ یا به عبارتی مجموع مربعات مانده‌ها را کمینه می‌کند که در آن r_i، iامین مؤلفه‌ی بردار $r = r(x)$ است.

اگر $rank(A) < n$، آنگاه جواب x برای (۱.۲) منحصر بفرد نیست. به هر حال در بین تمام جواب‌های کمترین مربعات، یک جواب منحصر بفرد وجود دارد به طوری که $||x||_2$ را نیز کمینه می‌کند (ر.ک. [۲]، قضیه‌ی ۱. ۲. ۱۰).

مسئله‌ی کمترین مربعات غیرخطی (NLSP)[1] نامقید عبارت است از یافتن کمینه ساز کلی برای مجموع مربعات m تابع غیرخطی. به عبارتی دیگر مسئله‌ی کمترین مربعات غیرخطی نامقید عبارت است از

$$\min_{x} \phi(x), \quad \phi(x) = \frac{1}{2} \parallel f(x) \parallel_2^2 \qquad (۲.۲)$$

که در آن، x یک بردار حقیقی n - بعدی و f یک تابع برداری حقیقی m - بعدی از x است [۲] و [۲۱]. این دسته از مسائل معمولاً در کنترل بهینه و پالایش و برازش داده‌ها بوجود می‌آیند. به عنوان مثال، اگر بخواهیم m داده‌ی مشاهده شده‌ی (t_i, y_i) را با یک قالب (الگو) $S(x,t)$ که با بردار n درایه‌ای x مشخص می‌شود، منطبق کنیم و f_i را نمایشگر خطا در قالب پیش بینی برای iامین مشاهده در نظر بگیریم آنگاه

$$f_i(x) = S(x, t_i) - y_i, \quad i = 1, ..., m,$$

و جواب بدست آمده به وسیله‌ی (۲.۲)، بهترین برازش داده‌ها یا به عبارتی دیگر ضابطه‌ای با کمترین مجموع مربعات خطا را به دست می‌دهد. در اینجا فرض می‌شود که تنها y_i، $i = 1, ..., m$ دارای خطا بوده ومقادیر t_i، $i = 1, ..., m$ از متغییر مستقل t دقیق هستند. حالتی که هر دو مقدار t_i و y_i، $i = 1, ..., m$ دارای خطا باشند در [۲] بررسی شده است و در این پایاننامه آن را بررسی نمی‌کنیم .

مسئله‌ی کمترین مربعات غیرخطی حالت خاصی از مسئله‌ی بهینه سازی نامقید کلی است. بنابراین روش‌هایی

[1] Nonlinear Least Squares Problem

که برای بهگزینی نامقید کلی بکار می‌رود، برای این مسئله نیز قابل بکارگیری است. اما به دلیل شکل خاص این مسئله، روش‌های مؤثرتری برای حل آن بکار گرفته می‌شود. عمده‌ترین این روش‌ها در [۲] و [۲۰] بررسی شده‌اند. برای مطالعه‌ی بیشتر در مورد حل مسئله‌ی (۲.۲) می‌توان به متونی که حاوی روش‌های عددی برای بهگزینی غیرخطی هستند مراجعه کرد.

یکی از مهمترین و قدیمی‌ترین روش‌هایی که برای حل (۲.۲) بکار می‌رود، **روش گاوس-نیوتن (GN)**[۲] است که در این پایاننامه بررسی می‌شود. روش GN شامل حل یک دنباله از تقریب‌های کمترین مربعات خطی شده برای مسئله‌ی کمترین مربعات غیر خطی (NLSP) است، که هرکدام از آنها بایک روند مستقیم یا تکراری حل می‌شود. در مقایسه با روش نیوتن و گونه‌های مختلف آن، روش GN برای حل NLSP جالب توجه است زیرا احتیاجی به برآورد یا تخمین مشتق مرتبه‌ی دوم تابع $f(x)$ ندارد و بنابراین از لحاظ عددی خیلی مؤثرتر است.

در عمل، بویژه برای مسائل با ابعاد خیلی بزرگ که در برازش داده‌ها بوجود می‌آیند، تقریب‌هایی مطابق روند GN برای کاهش محاسبات ساخته می‌شود. لازم است که تأثیر این تقریب‌ها روی همگرایی روشها مورد بررسی قرار گیرد. در این پایاننامه، تأثیر دو نوع از تقریب‌های مورد استفاده‌ی متداول در برازش داده‌ها را بررسی می‌کنیم: نخست روش‌های GN **قطع شده (TGN)**[۳] را بررسی می‌کنیم که در آن مسئله‌ی کمترین مربعات خطی درونی به طور دقیق حل نمی‌شود. سپس روش GN **پریشیده (PGN)**[۴] را بررسی می‌کنیم که‌در آن مسئله‌ی خطی شده‌ی واقعی درونی (داخلی) را با یک مسئله‌ی کمترین مربعات خطی قطع شده یا پریشیده‌ی ساده شده تقریب می‌کنیم. در این نوشته، شرایطی ارائه می‌شود که تضمین می‌کند روشهای GN قطع شده و پریشیده، تحت آن شرایط همگرا می‌شوند و همچنین نرخ همگرایی نقاط تکرار را بدست می‌آوریم. همچنین به بررسی بعضی از نظریه‌های اساسی برای روشهای دقیق می‌پردازیم. مطالب ارائه شده در این پایاننامه، در راستای مطالب موجود

[۲] Gauss-Newton
[۳] Truncated GN
[۴] Perturbed GN

در [۱۴] است.

در فصل ۳ به بررسی روش‌های عددی پایه برای حل مسئله‌ی (۲.۲) می‌پردازیم. در فصل ۴ الگوریتم GN را همراه با بعضی از نتایج همگرایی آن بررسی می‌کنیم. در فصل ۵ روش‌های GN تقریبی را معرفی کرده و سپس در فصل ۶ همگرایی این روش‌ها بررسی می‌شود. در فصل ۷ به بررسی انتخاب معیار توقف برای قطع تکرارها پرداخته و چگونگی استفاده از این معیار در عمل توضیح داده می‌شود. یک مثال عددی در فصل ۸ برای بررسی نظریه‌ها و نتایج ارائه شده از نظر عملی، داده می‌شود. نهایتاً در فصل ۹ نتایج و پیشنهادهای مربوطه ارائه خواهد شد.

فصل ۳

روش‌های عددی پایه

NLSP تعریف شده در (۲.۲) را در نظر بگیرید. ژاکوبی تابع f را با $J(x) \equiv f'(x)$ نشان می‌دهیم. گرادیان وهسی $\phi(x)$ به صورت

$$\nabla\phi(x) = J(x)^T f(x), \tag{۱.۳}$$

$$\nabla^2\phi(x) = J(x)^T J(x) + Q(x) \tag{۲.۳}$$

است که در آن $Q(x)$ نشان دهنده‌ی جملات مرتبه‌ی دوم به شکل

$$Q(x) = \sum_{i=1}^{n} f_i(x)\nabla^2 f_i(x) \tag{۳.۳}$$

است. در اینجا دو دیدگاه متفاوت برای مسئله‌ی (۲.۲) وجود دارد.

یک دیدگاه این است که آن را به عنوان یک مسئله‌ی بوجود آمده از یک دستگاه فرا معین از معادلات غیرخطی $f(x)$ در نظر بگیریم. بنابراین طبیعی است که $f(x)$ با یک قالب خطی در همسایگی یک نقطه‌ی داده شده‌ی x_c به صورت

$$\tilde{f}(x_c) = f(x_c) + J(x_c)(x - x_c) \tag{۴.۳}$$

تقریب شود. در این صورت می‌توان از حل مسئله‌ی کمترین مربعات خطی

$$min_x \|f(x_c) + J(x_c)(x - x_c)\|_2 \quad (۵.۳)$$

برای بدست آوردن جواب تقریبی خوبی برای (۲.۲) استفاده کرد. این روش که تنها از اطلاعات مشتق مرتبه‌ی اول درباره‌ی $f(x)$ بهره می‌گیرد، روش‌های گاوس-نیوتن و لوئنبرگ-مارکوارت را نتیجه می‌دهد [۲].

دیدگاه دیگر این است که (۲.۲) را به عنوان حالت خاصی از یک مسئله‌ی بهینه سازی کلی در نظرگرفته و یک تقریب درجه‌ی دوم

$$\bar{\phi}_c(x_c + z) = \phi(x_c) + \nabla\phi(x_c)^T z + \frac{1}{2} z^T \nabla^2 \phi(x_c) z \quad (۶.۳)$$

برای $\phi(x)$ در نظر گرفته می‌شود. این روش از اطلاعات مشتق مرتبه‌ی دوم درباره‌ی $f(x)$ بهره می‌گیرد. کمینه ساز $\bar{\phi}_c(x)$ عبارت است از

$$x_N = x_c - \left(J(x_c)^T J(x_c) + Q(x_c)\right)^{-1} J(x_c)^T f(x_c) \quad (۷.۳)$$

که $Q(x)$ در (۳.۳) آمده است. این روش معادل است با روش نیوتن بکار گرفته شده برای مسئله‌ی (۲.۲)که در آن نرخ همگرایی محلی اغلب از مرتبه‌ی دو است [۲].

توجه کنید که روش گاوس-نیوتن (۵.۳) همان روش (۷.۳) است که از جمله‌ی $Q(x)$ صرفنظر شده است. با توجه به (۳.۳) نتیجه می‌شود که این جمله کوچک خواهد بود اگر کمیت‌های

$$\|f_i(x)\| \left\|\nabla^2 f_i(x)\right\|, \quad i = 1, ..., m$$

کوچک باشند. این حالت هنگامی اتفاق می‌افتد که $f_i(x)$ها در x_c رفتار غیر خطی ملایمی داشته باشند یا مانده‌های $f_i(x)$ کوچک باشند. در این حالت انتظار می‌رود که رفتار روش گاوس-نیوتن شبیه روش نیوتن باشد. به ویژه

برای یک مسئله‌ی سازگار که $f(x^*) = 0$، همگرایی محلی برای هر دو روش از مرتبه‌ی دو خواهد بود. برای مسائل با مانده‌ی نسبتاً بزرگ، نرخ همگرایی محلی برای روش گاوس-نیوتن می‌تواند خیلی کمتر از روش نیوتن باشد [۲]. به هر حال هزینه‌ی محاسبات تعداد mn^2 مشتق مرتبه‌ی دوم $\nabla^2 f_i(x)$ می‌تواند به طور بازدارنده‌ای سنگین باشد.

برای مسئله‌ی برازش منحنی، مقادیر تابعی $f_i(x) = S(x, t_i) - y_i$، $i = 1, ..., m$ و مشتقات را می‌توان از روی تابع $S(x,t)$ بدست آورد. اگر $S(x,t)$ ترکیبی از توابع ساده‌ی نمایی و مثلثاتی باشد آنگاه مشتقات مرتبه‌ی دوم بعضی اوقات با هزینه‌ی کم قابل محاسبه خواهد بود. همچنین اگر $J(x)$ تُنُک باشد آنگاه $Q(x)$ نیز اغلب تُنُک خواهد بود و هزینه‌ی محاسباتی $Q(x)$ که شامل مشتقات نسبی مرتبه‌ی دوم است، ممکن است در این حالت خیلی پرهزینه نباشد [۲].

۱.۳ روش گاوس-نیوتن

روش گاوس-نیوتن برای مسئله‌ی (۲.۲) بر مبنای یک دنباله از تقریب‌ها برای $f(x)$ بنا می‌شود. اگر x_k نشان دهنده‌ی تقریب جاری باشد آنگاه یک اصلاحیه‌ی p_k به عنوان یک جواب از مسئله‌ی کمترین مربعات خطی

$$min_p \|f(x_k) + J(x_k)p\|_2 , \quad p \in \mathbb{R}^n \qquad (۸.۳)$$

محاسبه می‌شود و تقریب جدید عبارت خواهد بود از $x_{k+1} = x_k + p_k$.

روش گاوس-نیوتن اشاره شده در بالا دارای همگرایی محلی سریع برای مسائل غیرخطی ملایم و مسائل نزدیک به حالت سازگاری است [۲]. به هر حال ممکن است این روش برای برخی مسائل حتی به طور محلی هم همگرا نباشد. این حالت را با مسئله‌ی زیر که منسوب به پاول [1] است نشان می‌دهیم [۲۰]. همچنین رجوع کنید به [۱۱].

[1] Powell

مثال: مسئله‌ی زیر با $m = 2$ و $n = 1$ را در نظر بگیرید

$$\begin{aligned} r_1(x) &= x + 1 \\ r_2(x) &= \lambda x^2 + x - 1 \end{aligned}$$

که λ یک پارامتر است. کمینه ساز $r_1^2(x) + r_2^2(x)$ عبارت است از $x^* = 0$. در روش گاوس-نیوتن داریم

$$x_{k+1} = \lambda x_k + O(x_k^2)$$

و بنابراین این روش برای $|\lambda| > 1$ به طور محلی همگرا نمی‌شود.

۲.۳ روش‌هایی از نوع نیوتن

در مسائل با مانده‌ی بزرگ یا غیرخطی قوی، قالب درجه‌ی دوم (۵.۳) برای نمایش تابع $\phi(x)$ مناسب نیست زیرا قسمت مرتبه‌ی دوم هسی ($\nabla^2\phi(x)$) بزرگتر از آن است که بتوان آنرا نادیده گرفت. بعضی از آماردانها اهمیت الگوریتم‌ها را برای حل این دسته از مسائل کم ارزش می‌دانند. آنها بر این باورند که اگر مانده‌ها بزرگ باشند آنگاه نمی‌توان قالب را با مشاهده‌ها تطبیق داد و بنابراین به جای حل مسئله‌ی کمترین مربعات، ترجیح می‌دهند که یک قالب جدید و بهتری بسازند. البته اغلب مانده‌های بزرگ به دلیل اشتباه در مشاهده‌ها است که بوسیله‌ی آدمی صورت می‌گیرد. در این حالت ممکن است هنوز بخواهیم مسئله‌ی کمترین مربعات را حل کنیم زیرا جواب مسئله می‌تواند در تعیین مشاهده‌های دارای خطا مفید باشد. این مشاهده‌ها را می‌توان از مجموعه‌ی داده‌های مشاهده شده حذف کرد و یا وزن کمی را به آنها اختصاص داد (کمترین مربعات وزن‌دار) [۲].

عملکرد الگوریتم‌های گاوس-نیوتن و لوئنبرگ-مارکوارت در حالت مانده‌ی بزرگ معمولاً ضعیف است. همچنین این روش‌ها در نقاطی که رتبه‌ی ژاکوبی کامل نیست، دچار مشکل می‌شوند. همگرایی مجانبی این روش‌ها در این حالت تنها خطی و کندتر از نرخ همگرایی فوق خطی الگوریتم‌هایی مانند نیوتن و یا شبه نیوتن است که برای مسائل بهگزینی نامقید کلی بکار می‌روند. روش‌های نیوتن با ناحیه‌ی قابل اطمینان یا با جستجوی