JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 28, No. 2, March–April 2005

Engineering Notes

ENGINEERING NOTES are short manuscripts describing new developments or important results of a preliminary nature. These Notes cannot exceed 6 manuscript pages and 3 figures; a page of text may be substituted for a figure and vice versa. After informal review by the editors, they may be published within a few months of the date of receipt. Style requirements are the same as for regular contributions (see inside back cover).

Convergence of Newton's Method via Lyapunov Analysis

John E. Hurtado*
Texas A&M University,
College Station, Texas 77843-3141
and
Rush D. Robinett III†
Sandia National Laboratories,
Albuquerque, New Mexico 87185-0741

Introduction

THE local convergence properties of Newton's method for unconstrained minimization of smooth functions are well known: Newton's method converges to a local minimum if the current point is in a local solution space that is approximated well with a quadratic, positive-definite function. This property is discussed in many classic texts and is generally demonstrated using vector and matrix norms and related inequalities.[1] In this Note, the convergence of Newton's method is demonstrated using Lyapunov's direct method, which involves defining a positive-definite function and showing that the function's value strictly decreases when the proposed update equation is used. Although this approach has been used to discuss the global convergence of general descent algorithms,[1] it has not been used to demonstrate explicitly the local convergence of Newton's method.

The motivation for demonstrating the convergence of Newton's method in this manner stems from some recently developed cooperative-control laws that enable a team of autonomous vehicles to localize an unknown source that emits a measurable scalar field $F(\boldsymbol{x})$ (Ref. 2). Some practical applications include the localization of chemical plumes and light and acoustic beacons. Demonstrating the stability of these cooperative-control laws is tantamount to demonstrating the convergence of Newton's method, and because the closed-loop stability of a system is often demonstrated using Lyapunov's direct method, it is natural to consider this method to demonstrate the convergence of Newton's method.

Brief Review of Newton's Method for Unconstrained Minimization

Second-order methods of function minimization are based on a quadratic model of the smooth function F that is to be minimized. These methods rely on curvature information, which can be obtained from the Hessian matrix of second partials. Newton's method is a particular second-order method, which uses the first- and second-derivative information of the function F to determine the updated solution $\boldsymbol{p}_k$ from the solution to the following linear equation:

$$\boldsymbol{H}_k \boldsymbol{p}_k = -\boldsymbol{g}k \tag{1}$$

Equation (1) is obtained by minimizing a quadratic model of the objective function F. In Eq. (1), $\boldsymbol{H}_k$ is the square, symmetric Hessian matrix of second partials of F evaluated at the current location $\boldsymbol{x}_k$, $\boldsymbol{g}_k$, is the gradient vector of F at $\boldsymbol{x}_k$, $\boldsymbol{p}_k$ is the step from the current location $\boldsymbol{x}_k$ to the minimum, and the subscript k is used to denote the kth iterate.

In practice, Newton's method is usually modified in two ways. First, a step-length procedure usually accompanies the updated solution because a step of unity along the Newton direction $\boldsymbol{p}_k$ may not reduce the function F, even though it is the step that leads to the minimum of the quadratic model function. Consequently, with this first modification, the update equation becomes the following:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{p}_k \tag{2}$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{H}_k^{-1} \boldsymbol{g}_k \tag{3}$$

The parameter α_k is the step-length parameter, where $\alpha_k \in (0, 1]$. This parameter is often chosen to minimize $F(\boldsymbol{x}_k - \alpha_k \boldsymbol{M}_k \boldsymbol{g}_k)$, where $\boldsymbol{M}_k$ is some positive-definite matrix. This leads to the second type of modification, which depends on the Hessian.

The Hessian of the objective function F must be positive definite near the solution to satisfy the sufficiency conditions for a minimum. However, in regions far from the solution, the Hessian may not meet this requirement because of nonquadratic terms in the objective function. Consequently, the purpose of the second modification of Newton's method is to somehow accommodate the possible nonpositive definiteness of the Hessian at regions remote from the solution. This can be done by replacing the inverse of the Hessian matrix with a positive definite matrix $\boldsymbol{M}_k$. With this change the update equation is

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{M}_k \boldsymbol{g}_k \tag{4}$$

The matrix $\boldsymbol{M}_k$ can be set equal to the identity matrix, which results in a steepest descent algorithm, or set equal to the inverse of the Hessian matrix if the Hessian is positive definite.

One of two common methods may be used to construct the matrix $\boldsymbol{M}_k$:

1) Let $\boldsymbol{M}_k = (\epsilon_k \boldsymbol{I} + \boldsymbol{H}_k)^{-1}$, where ϵ_k is nonnegative. There is always an ϵ_k that makes $\boldsymbol{M}_k$ positive definite. One way to choose ϵ_k is to calculate the eigenvalues of the current Hessian matrix $\boldsymbol{H}_k$ and let ϵ_k be the smallest nonnegative constant for which the matrix $\epsilon_k \boldsymbol{I} + \boldsymbol{H}_k$ has eigenvalues greater than or equal to a nonnegative constant δ (Ref. 1).

2) Compute the eigensystem associated with the current Hessian matrix $\boldsymbol{H}_k = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$. Then let $\boldsymbol{M}_k = \boldsymbol{U}\bar{\boldsymbol{\Lambda}}\boldsymbol{U}^T$ where $\bar{\boldsymbol{\Lambda}}$ is a diagonal matrix with $\bar{\lambda}_i = |\lambda_i|$. This causes the effect of $\boldsymbol{M}_k$ to be opposite to that of $\boldsymbol{H}_k$ in the portion of eigensubspace associated with a negative eigenvalue.[3]

Received 28 October 2002; revision received 19 January 2004; accepted for publication 20 February 2004.

*Assistant Professor, Department of Aerospace Engineering, 3141 TAMU; jehurtado@tamu.edu.

†Deputy Director, Energy and Transportation Security Center, P.O. Box 5800, Mail Stop 0741; rdrobin@sandia.gov.

Proof of Convergence via Lyapunov Analysis

Let $\boldsymbol{x}^*$ denote a minimum point of the function F, and suppose this solution is contained in a (closed) subspace Ω of the entire space X. Using Taylor's theorem (see Ref. 3) F may be expressed as follows:

$$F(\boldsymbol{x}_k + h\boldsymbol{q}) = F(\boldsymbol{x}_k) + h\boldsymbol{g}_k^T\boldsymbol{q} + \tfrac{1}{2}h^2\boldsymbol{q}^T\boldsymbol{H}_k\boldsymbol{q} + \mathcal{O}(h^3) \tag{5}$$

Here, $\boldsymbol{x}_k$ and $\boldsymbol{q}$ are vectors in Ω, and h is a scalar. When $\boldsymbol{x} = \boldsymbol{x}_k + h\boldsymbol{q}$, this equation may be rewritten

$$\begin{aligned} F(\boldsymbol{x}) &= b_k + (\boldsymbol{g}_k - \boldsymbol{H}_k\boldsymbol{x}_k)^T\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{H}_k\boldsymbol{x} + \mathcal{O}(h^3) \\ &= b_k + (\boldsymbol{g}_k - \boldsymbol{H}_k\boldsymbol{x}_k)^T\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{H}_k\boldsymbol{x} + e(\boldsymbol{x}; \boldsymbol{x}_k) \end{aligned} \tag{6}$$

Here, $e(\boldsymbol{x}; \boldsymbol{x}_k)$ contains cubic terms in the vector difference $(\boldsymbol{x} - \boldsymbol{x}_k)$. A Lyapunov function is now introduced:

$$V(\boldsymbol{x}) = F(\boldsymbol{x}) - F(\boldsymbol{x}^*) \tag{7}$$

This function is constructed so that $V(\boldsymbol{x}^*) = 0$ and $V(\boldsymbol{x}) > 0\,\forall\,\boldsymbol{x} \neq \boldsymbol{x}^*$, where $\boldsymbol{x} \in \Omega$ (Ref. 4). The differential of the Lyapunov function is

$$\begin{aligned} \mathrm{d}V(\boldsymbol{x}, \Delta\boldsymbol{x}) &= \frac{\partial V^T}{\partial \boldsymbol{x}}\Delta\boldsymbol{x} \\ &= \left(\boldsymbol{g}_k - \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{H}_k\boldsymbol{x} + \frac{\partial e}{\partial \boldsymbol{x}}\right)^T \Delta\boldsymbol{x} \end{aligned} \tag{8}$$

By the evaluation of the differential at the current location $\boldsymbol{x}_k$ and the use of the update equation $\Delta\boldsymbol{x}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k = -\alpha_k\boldsymbol{M}_k\boldsymbol{g}_k$, Eq. (8) may be rewritten as

$$\mathrm{d}V(\boldsymbol{x}_k, \Delta\boldsymbol{x}_k) = -\alpha_k\boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k - \alpha_k\left(\left.\frac{\partial e}{\partial \boldsymbol{x}}\right|_{x_k}\right)^T\boldsymbol{M}_k\boldsymbol{g}_k \tag{9}$$

To demonstrate asymptotic convergence to the solution $\boldsymbol{x}^*$ using a Lyapunov analysis, one must show that the differential $\mathrm{d}V$ is negative definite and vanishes only at $\boldsymbol{x}^*$. In the following paragraphs, the negative-definite property of $\mathrm{d}V$ is presented first, followed by a demonstration that $\mathrm{d}V$ vanishes only at $\boldsymbol{x}^*$.

Notice that the first term of Eq. (9) is quadratic and is always negative, whereas the second term is sign indefinite. Consequently, to guarantee that $\mathrm{d}V(\boldsymbol{x}_k, \Delta\boldsymbol{x}_k)$ is negative definite, one must show the following:

$$\boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k > -\left(\left.\frac{\partial e}{\partial \boldsymbol{x}}\right|_{x_k}\right)^T\boldsymbol{M}_k\boldsymbol{g}_k \tag{10}$$

To demonstrate that this inequality is always met, attention is turned to the step-length parameter $\alpha_k \in (0, 1]$. Recall that α_k is often chosen so that $F(\boldsymbol{x}_k - \alpha_k\boldsymbol{M}_k\boldsymbol{g}_k)$ is minimized. Consequently, the following inequality is considered:

$$F(\boldsymbol{x}_{k+1}) < F(\boldsymbol{x}_k) \tag{11}$$

Equations (4) and (6) may be used in this expression, which becomes the following after canceling and gathering terms:

$$-\alpha_k(1 - \alpha_k/2)\boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k < -e(\boldsymbol{x}_{k+1}; \boldsymbol{x}_k) \tag{12}$$

One may next assume that the right-hand side of Eq. (12) can be approximated with a differential:

$$\begin{aligned} -e(\boldsymbol{x}_{k+1}; \boldsymbol{x}_k) &\cong -\left(\left.\frac{\partial e}{\partial \boldsymbol{x}}\right|_{x_k}\right)^T\Delta\boldsymbol{x}_k \\ &\cong \alpha_k\left(\left.\frac{\partial e}{\partial \boldsymbol{x}}\right|_{x_k}\right)^T\boldsymbol{M}_k\boldsymbol{g}_k \end{aligned} \tag{13}$$

This result may be used in Eq. (12):

$$\left(1 - \frac{\alpha_k}{2}\right)\boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k > -\left(\left.\frac{\partial e}{\partial \boldsymbol{x}}\right|_{x_k}\right)^T\boldsymbol{M}_k\boldsymbol{g}_k \tag{14}$$

However, because α_k is positive, one obtains the final inequality:

$$\boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k > \left(1 - \frac{\alpha_k}{2}\right)\boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k > -\left(\left.\frac{\partial e}{\partial \boldsymbol{x}}\right|_{x_k}\right)^T\boldsymbol{M}_k\boldsymbol{g}_k \tag{15}$$

Inequality (15) shows that inequality (10) is always met, and hence, $\mathrm{d}V$ is negative definite. However, it is interesting to investigate inequality (15) in regions near the solution, where α_k is expected to approach unity. When α_k equals one, inequality (15) becomes the following:

$$\frac{1}{2}\boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k > -\left(\left.\frac{\partial e}{\partial \boldsymbol{x}}\right|_{x_k}\right)^T\boldsymbol{M}_k\boldsymbol{g}_k \tag{16}$$

Inequality (16) demonstrates that, for points that begin near the solution $\boldsymbol{x}^*$, the Lyapunov function that led to the inequality condition given by Eq. (10) is actually conservative. Conservative Lyapunov functions that lead to conservative conditions for stability (or convergence) are not an uncommon occurrence.[4]

For asymptotic convergence to the solution x^*, the following remains to be shown:

$$\mathrm{d}V(\boldsymbol{x}_k, \Delta\boldsymbol{x}_k) = 0 \Leftrightarrow \boldsymbol{x}_k = \boldsymbol{x}^* \tag{17}$$

First, note that the necessary conditions for $\boldsymbol{x}^*$ to be a local minimum include the condition that the vector $\boldsymbol{g}^*$ equals zero.[3] Consequently, Eq. (9) gives the following:

$$\boldsymbol{x}_k = \boldsymbol{x}^* \Rightarrow \mathrm{d}V(\boldsymbol{x}^*, \Delta\boldsymbol{x}^*) = 0 \tag{18}$$

However, carefully notice from Eq. (9) that $\mathrm{d}V$ can also vanish should the following be true:

$$\boldsymbol{g}_k + \left.\frac{\partial e}{\partial \boldsymbol{x}}\right|_{x_k} = 0 \tag{19}$$

However, on examining inequality (14), one notes that the occurrence of Eq. (19) leads to a contradiction. That is, using Eq. (19) in inequality (14) leads to the following:

$$(1 - \alpha_k/2)\boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k > \boldsymbol{g}_k^T\boldsymbol{M}_k\boldsymbol{g}_k \tag{20}$$

The inequally cannot be true because $\boldsymbol{M}_k$ is positive definite and α_k is positive and nonzero.

Discussion

The usefulness of the result in this Note is in demonstrating the stability of control laws that are based on Newton's method for distributed, autonomous systems (see Ref. 2). Examples of robot trajectories are shown in Fig. 1, wherein nine robots are tasked with localizing an unknown source. The robot initial positions are marked with the ×, and their final positions are marked with the ○. It is assumed that each robot can measure the source intensity as its current position $F(x^i)$ and can transmit its information to neighboring robots. Each robot uses its own measurement and the received information to compute a quadratic approximation to the source field. Based on this approximation, each robot computes its position update according to $x_{k+1}^i = x_k^i - \alpha_k^i M_k^i \boldsymbol{g}_k^i$. The stability of the equilibrium point (i.e., the source location) along trajectories commensurate with the position update law for each robot, follows from the Lyapunov analysis convergence proof presented herein.

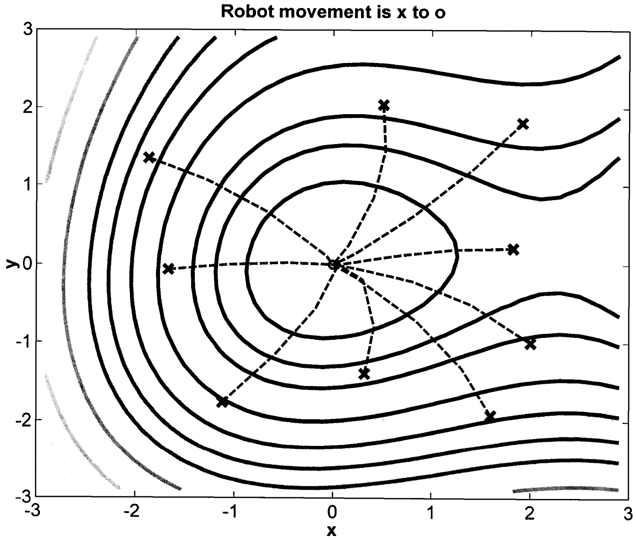

Fig. 1 Example of a robot swarm cooperatively localizing a source: - - -, robot position updates based on a Newton's method for unconstrained minimization of smooth functions and ——, level surfaces for unknown source modeled by a cubic polynomial.

Summary

As mentioned in the Introduction, the goal of this Note is to demonstrate the convergence of Newton's method using Lyapunov's direct method because of its bearing on some cooperative-control laws; the results do not impact Newton solver development nor Newton solver implementation. It is noteworthy that choosing the actual function that is to be minimized as the Lyapunov function leads to a conservative condition for convergence for points that begin near the solution.

Acknowledgment

Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy under Contract DE-AC04-94AL85000.

References

[1]Luenberger, D. G., *Linear and Nonlinear Programming*, 2nd ed., Addison Wesley Longman, Reading, MA, 1989, pp. 187, 225 ff.

[2]Robinett, R. D., III, and Hurtado, J. E., "Stability and Control of Collective Systems," *Journal of Intelligent and Robotic Systems, Theory and Applications*, Vol. 39, 2004, pp. 43–55.

[3]Gill, P. E., Murray, W., and Wright, M. H., *Practical Optimization*, Academic Press, Harcourt Brace and Co., New York, 1981, pp. 52, 63, 107.

[4]Junkins, J. L., and Kim, Y., *Introduction to Dynamic and Control of Flexible Structures,* AIAA Education Series, AIAA, Washington, DC, 1993, Chap. 3, passim, p. 15.