The Newton Method

Moritz Diehl

Overview

- The Newton Method
- Newton Type Methods
- Convergence Theory
- Globalization
- Application to Unconstrained Optimization

Nonlinear Root Finding Problem

Regard nonlinear continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^n$, $w \mapsto F(w)$. Aim is to solve **nonlinear root finding problem**

$$F(w) = 0.$$

Newton's idea: start with guess w_0, and recursively generate sequence of iterates $\{w_k\}_{k=0}^{\infty}$ by linearizing the nonlinear equation at the current iterate:

$$F(w_k) + \frac{\partial F}{\partial w}(w_k)(w - w_k) \quad = \quad 0.$$

Can explicitly compute next iterate:

$$w_{k+1} \quad = \quad w_k - \left(\frac{\partial F}{\partial w}(w_k) \right)^{-1} F(w_k)$$

Note: we have to assume that Jacobian $\frac{\partial F}{\partial w}(w)$ is invertible.

Example: fifth root of two

Regard $F(w) = w^5 - 2$, where $\frac{\partial F}{\partial w}(w) = 5w^4$.

Newton iteration:

$$w_{k+1} = w_k - (5w_k^4)^{-1}(w^5 - 2)$$

Iterates quickly converge to solution w^* with $F(w^*) = 0$.
In fact, the convergence rate of Newton's method is *quadratic*.

Convergence Rates

Sequence w_k converges to limit point $w^* = \lim_{k\to\infty} w_k$. Rate of convergence is

- **q-linear** if there exists $\alpha < 1$ and k_0 such that for all $k \geq k_0$ holds
$$\|w_{k+1} - w^*\| \leq \alpha \|w_k - w^*\|$$
- **q-superlinear** if there exists a sequence α_k with $\lim_{k\to\infty} \alpha_k = 0$ such that
$$\|w_{k+1} - w^*\| \leq \alpha_k \|w_k - w^*\|$$
- **q-quadratic** if there exists a β and k_0 such that for all $k \geq k_0$ holds
$$\|w_{k+1} - w^*\| \leq \beta \|w_k - w^*\|^2$$
Correct digits double in each iteration. E.g. $\|w_k - w^*\| = 10^{-2^k}$ with $\beta = 1$.

Newton Type Methods

More general, can use an approximation M_k of the Jacobian $J(w_k) := \frac{\partial F}{\partial w}(w_k)$. The Newton type iteration is

$$w_{k+1} = w_k - M_k^{-1} F(w_k)$$

Depending on how closely M_k approximates $J(w_k)$, the convergence rate can be faster or slower.

Local Contraction of Newton Type Methods

THEOREM: Sequence w_k converges to w^* with contraction rate

$$\|w_{k+1} - w^*\| \leq \left(\kappa + \frac{\omega}{2}\|w_k - w^*\|\right)\|w_k - w^*\|$$

if $\|w_0 - w^*\|$ is sufficiently small and there exist $\omega < \infty$ and $\kappa < 1$ such that for all w_k and w holds

$$\|M_k^{-1}(J(w_k) - J(w))\| \leq \omega\|w_k - w\| \qquad \text{(Lipschitz condition)}$$

$$\|M_k^{-1}(J(w_k) - M_k)\| \leq \kappa \qquad \text{(compatibility condition)}$$

Note: $\kappa = 0$ for exact Newton.

Proof (1):

$$
\begin{aligned}
w_{k+1} - w^* &= w_k - w^* - M_k^{-1} F(w_k) \\
&= w_k - w^* - M_k^{-1}(F(w_k) - F(w^*)) \\
&= M_k^{-1}(M_k(w_k - w^*)) \\
&\quad - M_k^{-1} \int_0^1 J(w^* + t(w_k - w^*))(w_k - w^*)\mathrm{d}t \\
&= M_k^{-1}(M_k - J(w_k))(w_k - w^*) \\
&\quad - M_k^{-1} \int_0^1 \Big[J(w^* + t(w_k - w^*)) - F(w_k) \Big] (w_k - w^*)\mathrm{d}t
\end{aligned}
$$

Proof (2):

Taking the norm of both sides:

$$
\begin{aligned}
\|w_{k+1} - w^*\| &\leq \kappa \|w_k - w^*\| \\
&\quad + \int_0^1 \omega \|w^* + t(w_k - w^*) - w_k\| \mathrm{d}t \, \|w_k - w^*\| \\
&= \left(\kappa + \omega \underbrace{\int_0^1 (1-t)\mathrm{d}t}_{=\frac{1}{2}} \|w_k - w^*\|\right) \|w_k - w^*\| \\
&= \left(\kappa + \frac{\omega}{2} \|w_k - w^*\|\right) \|w_k - w^*\|
\end{aligned}
$$

Globalization

The condition $\|w_0 - w^*\|$ is usually not satisfied, and the iteration needs to be *globalized* before it enters the area of local convergence. Different strategies:

- Homotopy: modify $F(w) = 0$ to an easier or solved problem (e.g. $F(w) - \lambda F(w_0) = 0$), slowly change homotopy parameter λ from 1 to 0, each time restarting Newton iteration at previous solution
- Line search: take smaller steps $\alpha_k \leq 1$

$$w_{k+1} = w_k - \alpha_k M_k^{-1} F(w_k)$$

  (equivalent to an increase of M_k to $\alpha_k^{-1} M_k$)
- Trust region/Levenberg Marquard: modify iteration to ensure that steps $\|w_{k+1} - w_k\|$ remain sufficiently small, e.g.

$$w_{k+1} = w_k - (\beta_k I + M_k)^{-1} F(w_k)$$

Newton for unconstrained optimization

Optimization problem $\min_x f(x)$ is solved by $\nabla f(x^*) = 0$. Identify $w \equiv x$ and $F(w) \equiv \nabla f(x)$. Exact Newton method iterates

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

General Newton type methods iterate

$$x_{k+1} = x_k - B_k^{-1} \nabla f(x_k)$$

with some **Hessian approximation** B_k, for example:

- gradient method: $B_k = I$
- quasi Newton updates such as BFGS
- Gauss-Newton, for least squares problems $f(x) = \|R(x)\|_2^2$ use $B_k = 2\frac{\partial R}{\partial x}(x_k)^\top \frac{\partial R}{\partial x}(x_k)$
- exact Hessian with Levenberg-Marquard modification: $B_k = \nabla^2 f(x_k) + \beta_k I$ to ensure positive definiteness and descent

Literature

- J. Nocedal and S. Wright: Numerical Optimization, Springer, 2006 (2nd edition)
- S. Boyd and L. Vandenberghe: Convex Optimization, Cambridge Univ. Press, 2004