Lecture 6: Second-order methods: Newton's method for unconstrained optimization (continued)

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Global convergence of linesearch-Newton's method

Theorem 9 Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ be bounded below on $\mathbb{R}^n$.

Let ∇f be Lipschitz continuous. Apply Newton's method to minimizing f with bArmijo linesearch and $\epsilon := 0$. For all $k \geq 0$, let the eigenvalues of $\nabla^2 f(x^k)$ at the iterates x^k be positive and uniformly bounded below, away from zero, independently of k. Then

either

there exists $l \geq 0$ such that $\nabla f(x^l) = 0$

or

$\|\nabla f(x^k)\| \to 0$ as $k \to \infty$. □

Global convergence of linesearch-Newton's method

Theorem 9   Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ be bounded below on $\mathbb{R}^n$. Let ∇f be Lipschitz continuous. Apply Newton's method to minimizing f with bArmijo linesearch and $\epsilon := 0$. For all $k \geq 0$, let the eigenvalues of $\nabla^2 f(x^k)$ at the iterates x^k be positive and uniformly bounded below, away from zero, independently of k. Then

either

$$\text{there exists } l \geq 0 \text{ such that } \nabla f(x^l) = 0$$

or

$$\|\nabla f(x^k)\| \to 0 \text{ as } k \to \infty. \qquad \square$$

• Theorem 9 is satisfied if $f \in \mathcal{C}^2$ with ∇f Lipschitz continuous is also strongly convex (i.e., the eigenvalues of $\nabla^2 f(x)$ for all x are positive, bounded below, away from zero). Then s^k is descent for all k. [Much stronger conditions than for SD methods.]

Global convergence of linesearch-Newton's method ...

Proof of Theorem 9. The conditions of Theorem 4 (Global convergence of GLM with bArmijo linesearch) are satisfied. Thus Th 4 gives that either $\exists\, l \geq 0$ such that $\nabla f(x^l) = 0$ or

$$E_k := \min\left\{\frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}, |\nabla f(x^k)^T s^k|\right\} \longrightarrow 0 \text{ as } k \to \infty. \ (\dagger)$$

Global convergence of linesearch-Newton's method ...

Proof of Theorem 9. The conditions of Theorem 4 (Global convergence of GLM with bArmijo linesearch) are satisfied. Thus Th 4 gives that either $\exists\, l \geq 0$ such that $\nabla f(x^l) = 0$ or

$$E_k := \min\left\{\frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}, |\nabla f(x^k)^T s^k|\right\} \longrightarrow 0 \text{ as } k \to \infty. \ (\dagger)$$

Assume now that $\nabla f(x^k) \neq 0$ for all $k \geq 0$. We are left with showing that ($\dagger$) implies that $\nabla f(x^k) \longrightarrow 0$ as $k \to \infty$. For this, we are going to express the terms in ($\dagger$) in terms of $\nabla f(x^k)$.

Global convergence of linesearch-Newton's method ...

Proof of Theorem 9. The conditions of Theorem 4 (Global convergence of GLM with bArmijo linesearch) are satisfied. Thus Th 4 gives that either $\exists\, l \geq 0$ such that $\nabla f(x^l) = 0$ or

$$E_k := \min \left\{ \frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}, |\nabla f(x^k)^T s^k| \right\} \longrightarrow 0 \text{ as } k \to \infty. \quad (\dagger)$$

Assume now that $\nabla f(x^k) \neq 0$ for all $k \geq 0$. We are left with showing that $(\dagger)$ implies that $\nabla f(x^k) \longrightarrow 0$ as $k \to \infty$. For this, we are going to express the terms in $(\dagger)$ in terms of $\nabla f(x^k)$.

Let $\nabla^2 f(x^k) := H_k$. The eigenvalues of $\nabla^2 f(x^k) = H_k$ are positive and uniformly bounded below, away from zero (for all k) $\Longrightarrow$ the smallest eigenvalue of H_k, $\lambda_{\min}(H_k) > 0$ and bounded away from zero, namely, there exists $\lambda_{\min} > 0$, independent of k, such that

$$\lambda_{\min}(H_k) \geq \lambda_{\min} \text{ for all } k \geq 0. \qquad (1)$$

Global convergence of linesearch-Newton's method ...

Proof of Theorem 9 (continued).

Problem 6, Sheet 2 $\Longrightarrow \forall s \in \mathbb{R}^n$, $s \neq 0$, and any symmetric $n \times n$ matrix M, $\lambda_{\min}(M) \leq \frac{s^T M s}{\|s\|^2} \leq \lambda_{\max}(M)$. (2)

Global convergence of linesearch-Newton's method ...

Proof of Theorem 9 (continued).

Problem 6, Sheet 2 $\Longrightarrow \forall s \in \mathbb{R}^n$, $s \neq 0$, and any symmetric $n \times n$ matrix M, $\lambda_{\min}(M) \leq \frac{s^T M s}{\|s\|^2} \leq \lambda_{\max}(M)$. (2)

Problem 6, Sheet 2: $f \in \mathcal{C}^2$ and ∇f Lipschitz continuous $\Leftrightarrow$ $\nabla^2 f$ is uniformly bounded above, which implies that all its eigenvalues are bounded above, and so there exists $\lambda_{\max} > 0$ independent of k such that

$$\lambda_{\max}(H_k) \leq \lambda_{\max}, \text{ for all } k \geq 0. \qquad (3)$$

Global convergence of linesearch-Newton's method ...

Proof of Theorem 9 (continued).

Problem 6, Sheet 2 $\Longrightarrow \forall s \in \mathbb{R}^n$, $s \neq 0$, and any symmetric $n \times n$ matrix M, $\lambda_{\min}(M) \leq \frac{s^T M s}{\|s\|^2} \leq \lambda_{\max}(M)$. (2)

Problem 6, Sheet 2: $f \in \mathcal{C}^2$ and ∇f Lipschitz continuous $\Leftrightarrow$ $\nabla^2 f$ is uniformly bounded above, which implies that all its eigenvalues are bounded above, and so there exists $\lambda_{\max} > 0$ independent of k such that

$$\lambda_{\max}(H_k) \leq \lambda_{\max}, \text{ for all } k \geq 0. \quad (3)$$

Returning to (†), we have from the definition of s^k (Newton direction), and $\nabla f(x^k) \neq 0$, for all k,

$$|\nabla f(x^k)^T s^k| = |\nabla f(x^k)^T H_k^{-1} \nabla f(x^k)| \overset{(2)}{\geq} \lambda_{\min}(H_k^{-1}) \|\nabla f(x^k)\|^2$$

$$= \frac{\|\nabla f(x^k)\|^2}{\lambda_{\max}(H_k)} \overset{(3)}{\geq} \frac{\|\nabla f(x^k)\|^2}{\lambda_{\max}}. \quad (4)$$

Global convergence of linesearch-Newton's method ...

Proof of Theorem 9 (continued). $s^k = -H_k^{-1}\nabla f(x^k)$ implies

$$\|s^k\|^2 = \nabla f(x^k)^T H_k^{-2} \nabla f(x^k) \overset{(2)}{\leq} \lambda_{\max}(H_k^{-2})\|\nabla f(x^k)\|^2$$
$$= \frac{\|\nabla f(x^k)\|^2}{[\lambda_{\min}(H_k)]^2} \overset{(1)}{\leq} \frac{\|\nabla f(x^k)\|^2}{\lambda_{\min}^2},$$

$$\Longrightarrow \frac{1}{\|s^k\|} \geq \frac{\lambda_{\min}}{\|\nabla f(x^k)\|} \text{ for all } k. \qquad (5)$$

Global convergence of linesearch-Newton's method ...

Proof of Theorem 9 (continued). $s^k = -H_k^{-1}\nabla f(x^k)$ implies

$$\|s^k\|^2 = \nabla f(x^k)^T H_k^{-2} \nabla f(x^k) \overset{(2)}{\leq} \lambda_{\max}(H_k^{-2})\|\nabla f(x^k)\|^2$$
$$= \frac{\|\nabla f(x^k)\|^2}{[\lambda_{\min}(H_k)]^2} \overset{(1)}{\leq} \frac{\|\nabla f(x^k)\|^2}{\lambda_{\min}^2},$$

$$\Longrightarrow \frac{1}{\|s^k\|} \geq \frac{\lambda_{\min}}{\|\nabla f(x^k)\|} \text{ for all } k. \qquad (5)$$

(4), (5) $\Longrightarrow E_k \geq \min\left\{\frac{\lambda_{\min}}{\lambda_{\max}}\|\nabla f(x^k)\|, \frac{1}{\lambda_{\max}}\|\nabla f(x^k)\|^2\right\} > 0$ for all k

This and (†) $\Longrightarrow \nabla f(x^k) \longrightarrow 0$ as $k \rightarrow \infty$. □

Global convergence for general second-order GLMs

In GLM, for all k, let B^k be symmetric, positive definite matrix and s^k given by $B^k s^k = -\nabla f(x^k)$. (*)

Theorem 10 Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ be bounded below on $\mathbb{R}^n$. Let ∇f be Lipschitz continuous. Apply GLM to minimizing f with s^k in (*), bArmijo linesearch and $\epsilon := 0$. For all k, let the eigenvalues of B^k be uniformly bounded above and below, away from zero, independently of k. Then

either

there exists $l \geq 0$ such that $\nabla f(x^l) = 0$

or

$\|\nabla f(x^k)\| \to 0$ as $k \to \infty$. $\square$

- Theorem requires locally strongly convex quadratic models of f for all k (but the Hessian of f may not be pos. def.).

Modified damped Newton methods

If $\nabla^2 f(x^k)$ is not positive definite, it is usual to solve instead

$$\left(\nabla^2 f(x^k) + M^k\right) s^k = -\nabla f(x^k),$$

where

- M^k chosen such that $\nabla^2 f(x^k) + M^k$ is “sufficiently” positive definite.
- $M^k := 0$ when $\nabla^2 f(x^k)$ is “sufficiently” positive definite.

Modified damped Newton methods

If $\nabla^2 f(x^k)$ is not positive definite, it is usual to solve instead

$$\left(\nabla^2 f(x^k) + M^k\right) s^k = -\nabla f(x^k),$$

where

- M^k chosen such that $\nabla^2 f(x^k) + M^k$ is "sufficiently" positive definite.
- $M^k := 0$ when $\nabla^2 f(x^k)$ is "sufficiently" positive definite.

Options:

$1.$ As $\nabla^2 f(x^k)$ is symmetric, we can factor $\nabla^2 f(x^k) = Q^k D^k (Q^k)^\top$, where Q^k is orthogonal and D^k is diagonal, and set

$$\nabla^2 f(x^k) + M^k := Q^k \max(\epsilon I, |D^k|)(Q^k)^\top,$$

for some "small" $\epsilon > 0$. Expensive approach for large problems.

Modified damped Newton methods

2. Estimate $\lambda_{\min}(\nabla^2 f(x^k))$ and set

$$M^k := \max(0, \epsilon - \lambda_{\min}(\nabla^2 f(x^k)))I.$$

Cheaper. Often tried in practice but “biased” (may overemphasize a large negative eigval at the expense of small, positive ones).

Modified damped Newton methods

2. Estimate $\lambda_{\min}(\nabla^2 f(x^k))$ and set

$$M^k := \max(0, \epsilon - \lambda_{\min}(\nabla^2 f(x^k)))I.$$

Cheaper. Often tried in practice but "biased" (may overemphasize a large negative eigval at the expense of small, positive ones).

3. Modified Cholesky: compute Cholesky factorization

$$\nabla^2 f(x^k) = L^k (L^k)^\top,$$

where L^k is lower triangular matrix. Modify the generated L^k if the factorization is in danger of failing (modify small or negative diagonal pivots, etc.).

Popular in computations.