Numerical Optimization

Exercise II

Jonathan Balzer

Geometric Modeling and Scientific Visualization Center
King Abdullah University of Science and Technology

Wolfe Condition

Assume: $\boldsymbol{p}_k$ is descent direction, i.e. $\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{p}_k < 0$.

- Sufficient decrease (Armijo) condition:

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) \leq f(\boldsymbol{x}_k) + c_1 \alpha_k \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{p}_k$$

$c_1 \in (0, 1)$

- Sufficient step size (curvature) condition:

$$\nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k)^\top \boldsymbol{p}_k \geq c_2 \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{p}_k$$

$c_2 \in (c_1, 1)$

Other Notions of "Good Step Size"

- Goldstein condition:

$$f(\boldsymbol{x}_k) + (1-c)\alpha_k \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{p}_k \le f(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k) \le f(\boldsymbol{x}_k) + c\alpha_k \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{p}_k$$

  $c \in (0, \frac{1}{2})$

- Sufficient decrease condition while backtracking

Fundamental Convergence Result

Theorem (Zoutendijk)

Let $\boldsymbol{x}_k$ be a sequence iteratively defined by $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$, where $\boldsymbol{p}_k$ is a descent direction, and α_k satisfies the Wolfe condition for every step k. Assume f is bounded from below, continuously differentiable and has Lipschitz-continuous first derivative. Then,

$$\lim_{n\to\infty} \sum_{k=0}^{n} \cos^2 \theta_k \|\nabla f(\boldsymbol{x}_k)\|^2 < \infty$$

where

$$\cos \theta_k = \frac{-\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{p}_k}{\|\nabla f(\boldsymbol{x}_k)\| \|\boldsymbol{p}_k\|}.$$

Rates I (Steepest Descent)

Theorem (Gradient Descent with Exact Line Search)

When the steepest descent method with exact line search is applied to the strongly convex quadratic function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + \boldsymbol{c},$$

the error norm satisfies

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_{\mathbf{A}} \leq \frac{\kappa - 1}{\kappa + 1}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_{\mathbf{A}}$$

for all k greater than some k_0.

Rates II (Quasi-Newton)

Theorem (Dennis and Moré)

Suppose $f \in C^3(\mathbb{R}^n)$*. Consider the iteration* $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}_k$ *with descent direction* $\boldsymbol{p}_k = -\mathbf{B}_k \nabla f(\boldsymbol{x}_k)$*. Under the assumption that* f *is strictly convex in* $\boldsymbol{x}^*$*, the convergence of* $\boldsymbol{x}_k \to \boldsymbol{x}^*$ *is superlinear if and only if*

$$\lim_{k\to\infty} \frac{\|(\mathbf{B}_k - \nabla^2 f(\boldsymbol{x}^*))\boldsymbol{p}_k\|}{\|\boldsymbol{p}_k\|} = 0.$$

Rates III (Newton)

Theorem (Local result)

Assume $f \in C^2(\mathbb{R}^n)$ with Lipschitz-continuous Hessian around a minimizer $\boldsymbol{x}^*$. The iteration $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}_k$ with descent direction $\boldsymbol{p}_k = -\nabla^2 f(\boldsymbol{x}_k)^{-1} \nabla f(\boldsymbol{x}_k)$

(i) *converges quadratically if $\boldsymbol{x}_0$ is sufficiently close to $\boldsymbol{x}^*$.*

(ii) *Moreover, the gradient $\nabla f(\boldsymbol{x}_k)$ converges quadratically to $\mathbf{0}$ in norm.*

Step Size Selection

- exact line search
- backtracking
- interpolation

Exercises

1. Compute the minimizer of $f : \mathbb{R}^n \mapsto \mathbb{R}$,

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + \boldsymbol{c},$$

in the set

$$D = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \boldsymbol{x}_0 + \alpha \boldsymbol{p}, \boldsymbol{x}_0, \boldsymbol{p} \in \mathbb{R}^n, \alpha \in \mathbb{R}_{\geq 0}\}.$$

Suppose $\boldsymbol{x}_0 = \mathbf{0}$ and $\boldsymbol{p}$ lies in an eigenspace of $\mathbf{A}$. Show that steepest descent with an exact line search will find the minimum of f in a single step.

Exercises

2. Show by construction that if $0 < c_2 < c_1 < 1$, there may be no step lengths satisfying the Wolfe condition.

Exercises

3. Reason mathematically that if $\phi(\alpha_0)$ is large, α_1 can be quite small in the step size selection by interpolation. Give an example for which this situation arises.