LECTURE 14

CONVERGENCE OF LINE-SEARCH METHODS

A line-search method is called <u>globally convergent</u> if it generates a sequence $\{x_k\}$ s.t.

$$\lim_{k \to \infty} \nabla f(x_k) = 0$$

for all $x_0 \in \mathbb{R}^n$.

The global convergence of a line-search method is typically guaranteed as long as the descent directions p_k don't become perpendicular to $\nabla f(x_k)$ as $k \to \infty$ and step-lengths α_k are chosen properly.

Angle between $\nabla f(x_k)$ and p_k θ_k satisfies

$$\cos \theta_k = \frac{\nabla f(x_k)^T p_k}{\|\nabla f(x_k)\| \, \|p_k\|}$$

THM (Zoutendijk)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be

* Lipschitz continuously differentiable, that is there exists a $\gamma > 0$ s.t.

$$\| \nabla f(x) - \nabla f(y) \| \leqslant \gamma \, \| x - y \|$$

for all $x, y \in \mathbb{R}^n$, and

* bounded below.

Suppose also that a line-search method generates

* descent directions p_k, and
* step-lengts α_k satisfying Wolfe conditions.

Then the following holds

$$\sum_{k=0}^{\infty} \| \nabla f(x_k) \|^2 \cos^2 \theta_k < \infty .$$

PROOF

By the sufficient curvature condition

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geqslant c_2 \nabla f(x_k)^T p_k$$

$$\Longrightarrow$$

$$(*) \quad \left(\nabla f(x_k + \alpha_k p_k) - \nabla f(x_k) \right)^T p_k \geqslant (c_2 - 1) \nabla f(x_k)^T p_k$$

Furthermore by Lipschitz continuity of derivatives

$$\left(\nabla f(x_k+\alpha_k p_k) - \nabla f(x_k)\right)^T p_k$$

$$\leqslant$$

$$\|\nabla f(x_k+\alpha_k p_k) - \nabla f(x_k)\| \, \|p_k\|$$

$$\leqslant$$

$$(**) \quad \gamma \|\alpha_k p_k\| \, \|p_k\| = \gamma \alpha_k \|p_k\|^2$$

Combine $(*)$ and $(**)$ to obtain

$$\alpha_k \geqslant \frac{(\mu_2 - 1)\, \nabla f(x_k)^T p_k}{\gamma \|p_k\|^2}$$

Plugging this inequality in sufficient decrease condition

$$f(x_k+\alpha_k p_k) \leqslant f(x_k) + \alpha_k \mu_1 \nabla f(x_k)^T p_k$$

$$\leqslant f(x_k) + \frac{\mu_1(\mu_2-1)}{\gamma \|p_k\|^2} \left(\nabla f(x_k)^T p \right.$$

$$= f(x_k) + \frac{\mu_1(\mu_2-1)}{\gamma} \cos^2\theta_k \|\nabla f$$

That is

$$\frac{\mu_1(1-\mu_2)}{\gamma}\cos^2\theta_k\,\|\nabla f(x_k)\|^2 \leqslant f(x_k)-f(x_{k+1})$$

$$\Longrightarrow$$

$$\sum_{k=0}^{m}\cos^2\theta_k\,\|\nabla f(x_k)\|^2 \leqslant \frac{\gamma}{\mu_1(1-\mu_2)}\underbrace{\left(f(x_0)-f(x_{m+1})\right)}_{\substack{\text{bounded}\\ \text{since } f \text{ is}\\ \text{bounded below}}}$$

Taking the limit as $m \to \infty$,

$$\sum_{k=0}^{\infty}\cos^2\theta_k\,\|\nabla f(x_k)\|^2 \leqslant \frac{\gamma}{\mu_1(1-\mu_2)}\lim_{m\to\infty}\left(f(x_0)-f(x_{m+1})\right)$$

$$< \infty.$$

□

~~CONVERGENCE OF STEEPEST DESCENT~~

~~For steepest descent~~

~~$\cos\theta_k$~~

Suppose that for a line-search method there exists an $M > 0$ s.t.

$$|\cos\theta_k| \geqslant M$$

for all large k.

Then by Zoutendijk's theorem

$$M^2 \sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2 \leqslant \sum_{k=0}^{\infty} \cos^2\theta_k \, \|\nabla f(x_k)\|^2$$

$$< \infty$$

$$\Longrightarrow$$

$$\lim_{k \to \infty} \|\nabla f(x_k)\| = 0,$$

that is the method is globally convergent

<u>CONVERGENCE OF STEEPEST DESCENT</u>

~~For~~ steepest descent

$$\cos\theta_k = \frac{\nabla f(x_k)^T p_k}{\|\nabla f(x_k)\| \, \|p_k\|}$$

$$= -\frac{\nabla f(x_k)^T \nabla f(x_k)}{\|\nabla f(x_k)\| \, \|-\nabla f(x_k)\|} = -1$$

Therefore the method is globally convergent.

· The rate of convergence is linear.
For a quadratic function

$$q(x) = \frac{1}{2} x^T A x + b^T x + c$$

with exact line searches the steepest descent iterates $\{x_k\}$ satisfy

$$\lim_{k \to \infty} \frac{\| x_{k+1} - x_* \|}{\| x_k - x_* \|} = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}$$

where

* x_* is the local minimizer of $q(x)$
* $\lambda_{min}, \lambda_{max}$ are the smallest and largest eigenvalues of A.

~~CONVERGENCE OF NEWTON'S METHOD~~

~~Suppose that $\|\nabla^2 f(x_k)\|$ and $\|\nabla^2 f(x_k)^{-1}\|$ remain bounded, that is there exists an $M \geq 0$ s.t.~~

$$\cancel{\|\nabla^2 f(x_k)\| \, \|\nabla^2 f(x_k)^{-1}\| \geq M}$$

CONVERGENCE OF MODIFIED NEWTON'S METHOD

Let $\widetilde{\nabla^2 f(x_k)}$ denote positive definite modified Hessian matrices.

Suppose $\| \widetilde{\nabla^2 f(x_k)} \|$ and $\| \widetilde{\nabla^2 f(x_k)}^{-1} \|$ remain bounded, that is there exists an $M > 0$ such that

$$(+) \quad \| \widetilde{\nabla^2 f(x_k)} \| \, \| \widetilde{\nabla^2 f(x_k)}^{-1} \| \leq M.$$

(+) implies

$$\cos \theta_k = - \frac{\nabla f(x_k)^T \widetilde{\nabla^2 f(x_k)}^{-1} \nabla f(x_k)}{\| \nabla f(x_k) \| \, \| \widetilde{\nabla^2 f(x_k)}^{-1} \nabla f(x_k)}$$

Noting that

$$* \; \nabla f(x_k)^T \widetilde{\nabla^2 f(x_k)}^{-1} \nabla f(x_k)$$

$$\geq$$

$$\lambda_{min} (\widetilde{\nabla^2 f(x_k)}^{-1}) \nabla f(x_k)^T \nabla f(x_k)$$

$$=$$

$$\lambda_{min} (\widetilde{\nabla^2 f(x_k)}^{-1}) \| \nabla f(x_k) \|^2$$

· and

$$* \ \lambda_{max}\left(\widetilde{\nabla^2 f(x_k)}^{-1}\right) = \left\| \widetilde{\nabla^2 f(x_k)}^{-1} \right\|$$

$$\geq \frac{\left\| \widetilde{\nabla^2 f(x_k)}^{-1} \nabla f(x_k) \right\|}{\| \nabla f(x_k) \|}$$

we have

$$|\cos \theta_k| \geq \frac{\lambda_{min}\left(\widetilde{\nabla^2 f(x_k)}^{-1}\right) \| \nabla f(x_k) \|^2}{\lambda_{max}\left(\widetilde{\nabla^2 f(x_k)}^{-1}\right) \| \nabla f(x_k) \|^2}$$

$$= \frac{1}{\lambda_{max}\left(\widetilde{\nabla^2 f(x_k)}\right) \lambda_{max}\left(\widetilde{\nabla^2 f(x_k)}^{-1}\right)}$$

$$= \frac{1}{\left\| \widetilde{\nabla^2 f(x_k)} \right\| \left\| \widetilde{\nabla^2 f(x_k)}^{-1} \right\|}$$

$$\geq \frac{1}{M}$$

Therefore ~~the~~ a modified Newton's method ensuring Wolfe conditions is globally convergent.

CONVERGENCE OF QUASI-NEWTON METHODS

Assuming the Hessian approximation B_k satisfy

$$(+) \quad \|B_k\| \, \|B_k^{-1}\| \leq M$$

it can be similarly shown that

$$|\cos \theta_k| \geq \frac{1}{M}.$$

Quasi-Newton methods with Wolfe conditions and satisfying (+) are globally convergent.

COMPARISON OF METHODS

	Rate of Convergence	# FLOPS per iteration
Steepest Descent	Linear	$O(n)$
Modified Newton	Quadratic	$O(n^3)$
Quasi-Newton	Superlinear	$O(n^2)$