Lecture 4: Linesearch methods (continued)

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

A generic linesearch method (Lecture 2)

(UP): minimize $f(x)$ subject to $x \in \mathbb{R}^n$, where $f \in \mathcal{C}^1$ or $\mathcal{C}^2$.

A Generic Linesearch Method (GLM)

`Choose` $\epsilon > 0$ `and` $x^0 \in \mathbb{R}^n$. `For` $k \geq 0$, `do:`
`While` $\|\nabla f(x^k)\| > \epsilon$, `REPEAT:`

- `compute a` <u>`descent`</u> `search direction` $s^k \in \mathbb{R}^n$,

$$\nabla f(x^k)^T s^k < 0;$$

- `compute a stepsize` $\alpha^k > 0$ `along` s^k `such that`

$$f(x^k + \alpha^k s^k) < f(x^k);$$

- `set` $x^{k+1} := x^k + \alpha^k s^k$ `and` $k := k + 1$. $\square$

Recall property of descent directions (Lemma 1, Lecture 1).

Global convergence of GLM (Lecture 3)

- $f \in \mathcal{C}^1(\mathbb{R}^n)$; ∇f is Lipschitz continuous (on $\mathbb{R}^n$) iff $\exists L > 0$,
$$\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|, \quad \forall x,\, y \in \mathbb{R}^n.$$

Lemma 2. Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ with ∇f Lipschitz continuous with Lipschitz constant L. Assume a GLM is applied to minimizing f. Then at the kth iteration, the Armijo condition :

$$f(x^k + \alpha s^k) \leq f(x^k) + \beta\alpha\nabla f(x^k)^T s^k \qquad \text{(ac)}$$

is satisfied for all $\alpha \in [0, \alpha^k_{\max}]$, where $\alpha^k_{\max} = \dfrac{(\beta - 1)\nabla f(x^k)^T s^k}{L\|s^k\|^2}$.

Lemma 3. Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ with ∇f Lipschitz continuous with Lipschitz constant L. Assume a GLM is applied to minimizing f. Then at the kth iteration, the bArmijo stepsize α^k satisfies

$$\alpha^k \geq \min\{\alpha_{(0)}, \tau\alpha^k_{\max}\} \text{ for all } k \geq 0,$$

where $\alpha^k_{\max}$ is defined in Lemma 2.

Global convergence of GLM (continued)

Theorem 4. Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ be bounded below on $\mathbb{R}^n$ by f_{low}. Let ∇f Lipschitz continuous. Apply GLM with bArmijo linesearch to minimizing f with $\epsilon := 0$. Then

either

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

or

$$\lim_{k \to \infty} \min \left\{ \frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}, |\nabla f(x^k)^T s^k| \right\} = 0.$$

Global convergence of GLM (continued)

Theorem 4. Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ be bounded below on $\mathbb{R}^n$ by f_{low}. Let ∇f Lipschitz continuous. Apply GLM with bArmijo linesearch to minimizing f with $\epsilon := 0$. Then

either

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

or

$$\lim_{k \to \infty} \min \left\{ \frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}, |\nabla f(x^k)^T s^k| \right\} = 0.$$

Proof of Theorem 4. If there exists $l \geq 0$ such that $\nabla f(x^l) = 0$, we are done.
Assume now that $\nabla f(x^k) \neq 0$ for all $k \geq 0$. Then Armijo condition (ac) with $\alpha := \alpha^k$, and $x^{k+1} = x^k + \alpha^k s^k$, give

$$f(x^{k+1}) \leq f(x^k) + \beta \alpha^k \nabla f(x^k)^T s^k \text{ for all } k \geq 0,$$

or equivalently, for all $k \geq 0$,

Global convergence of GLM ...

Proof of Theorem 4. or equivalently, for all $k \geq 0$,

$$f(x^k) - f(x^{k+1}) \geq -\beta\alpha^k \nabla f(x^k)^T s^k = \beta\alpha^k |\nabla f(x^k))^T s^k|, \quad (1)$$

where in the last equality, we used

$$(-\nabla f(x^k))^T s^k = |\nabla f(x^k))^T s^k|$$

since $\nabla f(x^k)^T s^k < 0$ (s^k is descent).

Part of Proof of Theorem 4

Let $i \geq 0$. Then (1) implies

$$\boxed{f(x^0)} - \cancel{f(x^1)} \geq \beta\alpha^0 |\nabla f(x^0)^T s^0|$$

$$\cancel{f(x^1)} - \cancel{f(x^2)} \geq \beta\alpha^1 |\nabla f(x^1)^T s^1|$$

$$\vdots$$

$$\cancel{f(x^{i-2})} - \cancel{f(x^{i-1})} \geq \beta\alpha^{i-2} |\nabla f(x^{i-2})^T s^{i-2}|$$

$$\cancel{f(x^{i-1})} - \cancel{f(x^i)} \geq \beta\alpha^{i-1} |\nabla f(x^{i-1})^T s^{i-1}|$$

$$\cancel{f(x^i)} - \boxed{f(x^{i+1})} \geq \beta\alpha^i |\nabla f(x^i)^T s^i|$$

$\sum$: (telescopic sum)

$$f(x^0) - f(x^{i+1}) \geq \sum_{k=0}^{i} \beta\alpha^k |\nabla f(x^k)^T s^k|.$$

Global convergence of GLM ...

Proof of Theorem 4. or equivalently, for all $k \geq 0$,

$$f(x^k) - f(x^{k+1}) \geq -\beta\alpha^k \nabla f(x^k)^T s^k = \beta\alpha^k |\nabla f(x^k))^T s^k|, \quad (1)$$

where in the last equality, we used

$$(-\nabla f(x^k))^T s^k = |\nabla f(x^k))^T s^k|$$

since $\nabla f(x^k)^T s^k < 0$ (s^k is descent).

Let $i \geq 0$. Summing up (1) from $k = 0$ to $k = i$ (see next slide), we find that consecutive terms on the left-hand side cancel to give

$$f(x^0) - f(x^{i+1}) \geq \beta \sum_{k=0}^{i} \alpha^k |\nabla f(x^k))^T s^k|. \quad (2)$$

Global convergence of GLM ...

Proof of Theorem 4. or equivalently, for all $k \geq 0$,

$$f(x^k) - f(x^{k+1}) \geq -\beta\alpha^k \nabla f(x^k)^T s^k = \beta\alpha^k |\nabla f(x^k))^T s^k|, \quad (1)$$

where in the last equality, we used

$$(-\nabla f(x^k))^T s^k = |\nabla f(x^k))^T s^k|$$

since $\nabla f(x^k)^T s^k < 0$ (s^k is descent).

Let $i \geq 0$. Summing up (1) from $k = 0$ to $k = i$ (see next slide), we find that consecutive terms on the left-hand side cancel to give

$$f(x^0) - f(x^{i+1}) \geq \beta \sum_{k=0}^{i} \alpha^k |\nabla f(x^k))^T s^k|. \qquad (2)$$

As f is bounded below by f_{low}, $f(x^{i+1}) \geq f_{\text{low}}$ for all $i \geq 0$. Thus, letting $i \longrightarrow \infty$ in (2), we deduce that

$$\infty > f(x^0) - f_{\text{low}} \geq \beta \sum_{k=0}^{\infty} \alpha^k |\nabla f(x^k))^T s^k|, \quad (3)$$

We deduce from the convergence of the series in (3) that

$$\lim_{k \longrightarrow \infty} \alpha^k |\nabla f(x^k))^T s^k| = 0. \quad (4)$$

Global convergence of GLM ...

Proof of Theorem 4. Recall Lemma 3 (i.e., $\alpha^k \geq \min\{\alpha_{(0)}, \tau\alpha^k_{\max}\}$). Define the following index sets

$$\mathcal{K}_1 = \{k : \alpha_{(0)} \geq \tau\alpha^k_{\max}\} \text{ and } \mathcal{K}_2 = \{k : \alpha_{(0)} < \tau\alpha^k_{\max}\},$$

where $\alpha^k_{\max}$ is defined in Lemma 2. Note that every iteration k belongs either to $\mathcal{K}_1$ or $\mathcal{K}_2$.

Global convergence of GLM ...

Proof of Theorem 4. Recall Lemma 3 (i.e., $\alpha^k \geq \min\{\alpha_{(0)}, \tau\alpha^k_{\max}\}$). Define the following index sets

$$\mathcal{K}_1 = \{k : \alpha_{(0)} \geq \tau\alpha^k_{\max}\} \text{ and } \mathcal{K}_2 = \{k : \alpha_{(0)} < \tau\alpha^k_{\max}\},$$

where $\alpha^k_{\max}$ is defined in Lemma 2. Note that every iteration k belongs either to $\mathcal{K}_1$ or $\mathcal{K}_2$.

For all $k \in \mathcal{K}_1$, we have from Lemmas 2 & 3 that

$$\alpha^k |\nabla f(x^k))^T s^k| \geq \frac{(1-\beta)\tau}{L} \cdot \left(\frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}\right)^2 \geq 0$$

and so (4) implies $\lim_{k\to\infty, k\in\mathcal{K}_1} |\nabla f(x^k)^T s^k|/\|s^k\| = 0$.

Global convergence of GLM ...

Proof of Theorem 4. Recall Lemma 3 (i.e., $\alpha^k \geq \min\{\alpha_{(0)}, \tau\alpha^k_{\max}\}$). Define the following index sets

$$\mathcal{K}_1 = \{k : \alpha_{(0)} \geq \tau\alpha^k_{\max}\} \text{ and } \mathcal{K}_2 = \{k : \alpha_{(0)} < \tau\alpha^k_{\max}\},$$

where $\alpha^k_{\max}$ is defined in Lemma 2. Note that every iteration k belongs either to $\mathcal{K}_1$ or $\mathcal{K}_2$.

For all $k \in \mathcal{K}_1$, we have from Lemmas 2 & 3 that

$$\alpha^k |\nabla f(x^k))^T s^k| \geq \frac{(1-\beta)\tau}{L} \cdot \left(\frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}\right)^2 \geq 0$$

and so (4) implies $\lim_{k\to\infty, k\in\mathcal{K}_1} |\nabla f(x^k)^T s^k|/\|s^k\| = 0$.

Lemma 3 gives that $\alpha^k \geq \alpha_{(0)}$ for all $k \in \mathcal{K}_2$ and so (4) provides $\lim_{k\to\infty, k\in\mathcal{K}_2} |\nabla f(x^k)^T s^k| = 0$.

Global convergence of GLM ...

Proof of Theorem 4. Recall Lemma 3 (i.e., $\alpha^k \geq \min\{\alpha_{(0)}, \tau\alpha^k_{\max}\}$). Define the following index sets

$$\mathcal{K}_1 = \{k : \alpha_{(0)} \geq \tau\alpha^k_{\max}\} \text{ and } \mathcal{K}_2 = \{k : \alpha_{(0)} < \tau\alpha^k_{\max}\},$$

where $\alpha^k_{\max}$ is defined in Lemma 2. Note that every iteration k belongs either to $\mathcal{K}_1$ or $\mathcal{K}_2$.

For all $k \in \mathcal{K}_1$, we have from Lemmas 2 & 3 that

$$\alpha^k |\nabla f(x^k))^T s^k| \geq \frac{(1-\beta)\tau}{L} \cdot \left(\frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}\right)^2 \geq 0$$

and so (4) implies $\lim_{k\to\infty, k\in\mathcal{K}_1} |\nabla f(x^k)^T s^k|/\|s^k\| = 0$.

Lemma 3 gives that $\alpha^k \geq \alpha_{(0)}$ for all $k \in \mathcal{K}_2$ and so (4) provides $\lim_{k\to\infty, k\in\mathcal{K}_2} |\nabla f(x^k)^T s^k| = 0$.

These two limits for the $\mathcal{K}_1$ and $\mathcal{K}_2$ subsequences, and the property $\min\{a_k, b_k\} \leq a_k,\ b_k,\ \forall k$, give the required limit, namely, $\lim_{k\to\infty} \min\left\{\frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}, |\nabla f(x^k)^T s^k|\right\} = 0.$ $\Box$

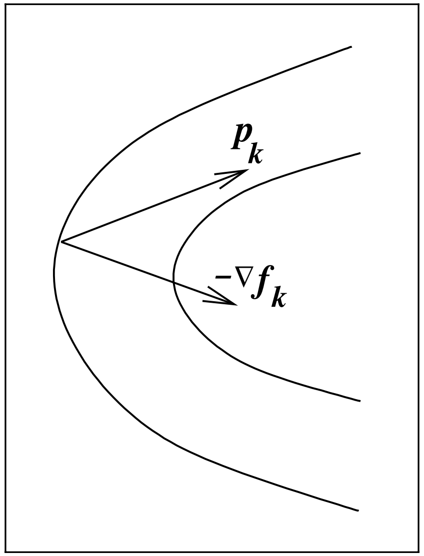
Global convergence of GLM ...

Interpretation of Theorem 4: Recall

$$\cos\theta_k = \frac{(-\nabla f(x^k))^T s^k}{\|\nabla f(x^k)\| \cdot \|s^k\|} = \frac{|\nabla f(x^k)^T s^k|}{\|\nabla f(x^k)\| \cdot \|s^k\|}.$$

Then Th 4 gives, if $\nabla f(x^k) \neq 0$ for all k,

$$\lim_{k\to\infty} \|\nabla f(x^k)\| \cdot \cos\theta_k \cdot \min\{1, \|s^k\|\} = 0.$$



A descent direction p_k.

Thus to ensure global convergence of GLM, namely, $\|\nabla f(x^k)\| \longrightarrow 0$ as $k \to \infty$, it is not sufficient to have s^k be descent for each k; we need $\cos\theta_k \geq \delta > 0$ for all k, so that s^k is prevented from becoming orthogonal to the gradient as k increases.

Summary and a look ahead

Linesearch methods:

- Linesearch: how to choose the stepsize α^k, from any x^k and along any descent direction s^k.
- How to choose a descent direction s^k? What are the important such choices of s^k?
  - Steepest descent direction (next).
  - Newton direction.