Lecture 3: Methods for local unconstrained optimization. Linesearch methods (continued)

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

A generic linesearch method

(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 descent 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

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

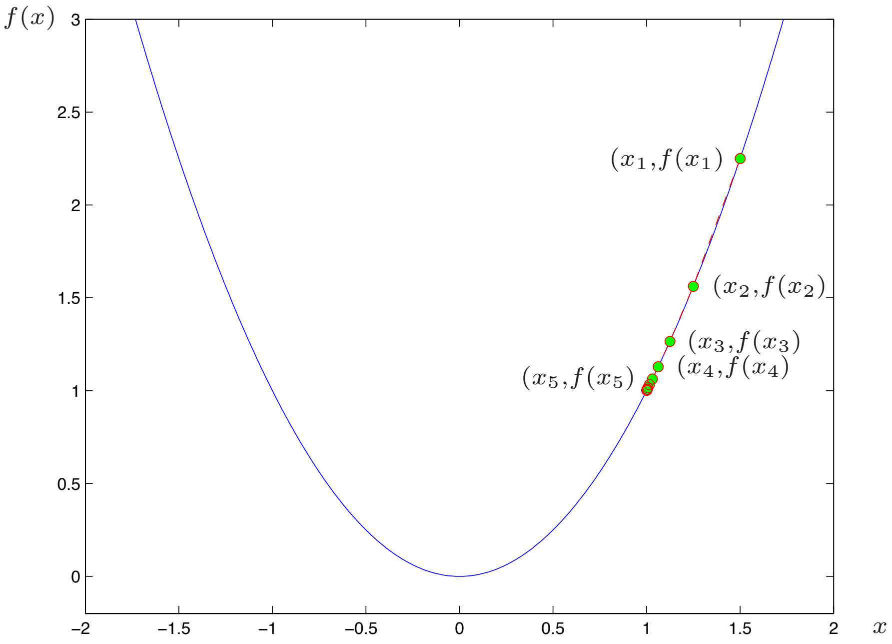
- set $x^{k+1} := x^k + \alpha^k s^k$ and $k := k + 1$. $\square$

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

Inexact linesearch

- want stepsize α^k not "too short".

Example: $f(x) = x^2$; $\quad x^0 = 2$; $s^k = -1$ and $\alpha^k = 1/(2^{k+1})$ for all k. Then GLM gives $x^k \longrightarrow 1$ as $k \longrightarrow \infty$. [see Pb Sheet 1]



Inexact linesearch ...

- want stepsize α^k not "too short".

A backtracking linesearch algorithm

Choose $\alpha_{(0)} > 0$ and $\tau \in (0,1)$.

While $f(x^k + \alpha_{(i)} s^k)'' \geq'' f(x^k)$, REPEAT:

- set $\alpha_{(i+1)} := \tau \alpha_{(i)}$ and $i := i + 1$.

END.

Set $\alpha^k := \alpha_{(i)}$. $\square$

Inexact linesearch ...

- want stepsize α^k not "too short".

A backtracking linesearch algorithm

`Choose` $\alpha_{(0)} > 0$ `and` $\tau \in (0,1)$.

`While` $f(x^k + \alpha_{(i)} s^k)'' \geq'' f(x^k)$, `REPEAT:`

- `set` $\alpha_{(i+1)} := \tau \alpha_{(i)}$ `and` $i := i + 1$.

`END.`

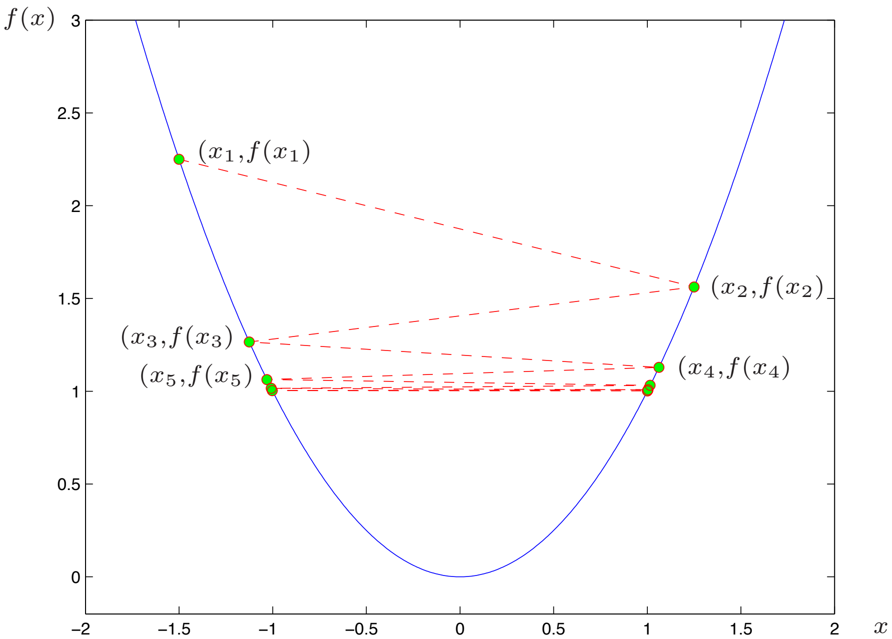
`Set` $\alpha^k := \alpha_{(i)}$. $\square$

- $\alpha_{(0)} := 1; \tau := 0.5 \implies \alpha_{(0)} := 1, \alpha_{(1)} := 0.5, \alpha_{(2)} := 0.25, \ldots$
- "<": simple or more sophisticated decrease in f at x^k.

Inexact linesearch ...

- want stepsize α^k not “too long” compared to the decrease in f.

  Example: $f(x) = x^2$; $\quad x^0 = 2$; $s^k = (-1)^{k+1}$ and $\alpha^k = 2 + 3/2^{k+1}$ for all k. Then GLM gives $x^k \longrightarrow \pm 1$ as $k \longrightarrow \infty$. [see Pb Sheet 1]



Inexact linesearch ...

- want stepsize α^k not "too long" compared to the decrease in f.

The Armijo condition

Choose $\beta \in (0,1)$.

Compute $\alpha^k > 0$ such that

$$f(x^k + \alpha^k s^k) \leq f(x^k) + \beta \alpha^k \nabla f(x^k)^T s^k \qquad (*)$$

is satisfied. $\square$

Inexact linesearch ...

- want stepsize α^k not "too long" compared to the decrease in f.

The Armijo condition

`Choose` $\beta \in (0,1)$`.`
`Compute` $\alpha^k > 0$ `such that`

$$f(x^k + \alpha^k s^k) \leq f(x^k) + \beta \alpha^k \nabla f(x^k)^T s^k \qquad (*)$$

`is satisfied.` $\Box$

- in practice, $\beta := 0.1$ or even $\beta := 0.001$.

- due to the descent condition, $\exists\, \overline{\alpha}^k > 0$ (unknown explicitly in general) such that $(*)$ holds for all $\alpha \in [0, \overline{\alpha}^k]$. [see Pb Sheet 2]
Choose α^k as large as possible in $(0, \overline{\alpha}^k]$ or in other (greater) intervals of positive α-values that may satisfy $(*)$.

Inexact linesearch ...

- $\Phi_k : \mathbb{R} \to \mathbb{R}, \quad \Phi_k(\alpha) := f(x^k + \alpha s^k), \quad \alpha \geq 0$. Then

$$\text{Armijo} \iff \Phi_k(\alpha^k) \leq \Phi_k(0) + \beta\alpha^k\Phi_k'(0).$$

Let $y_\beta(\alpha) := \Phi_k(0) + \beta\alpha\Phi_k'(0), \quad \alpha \geq 0$.

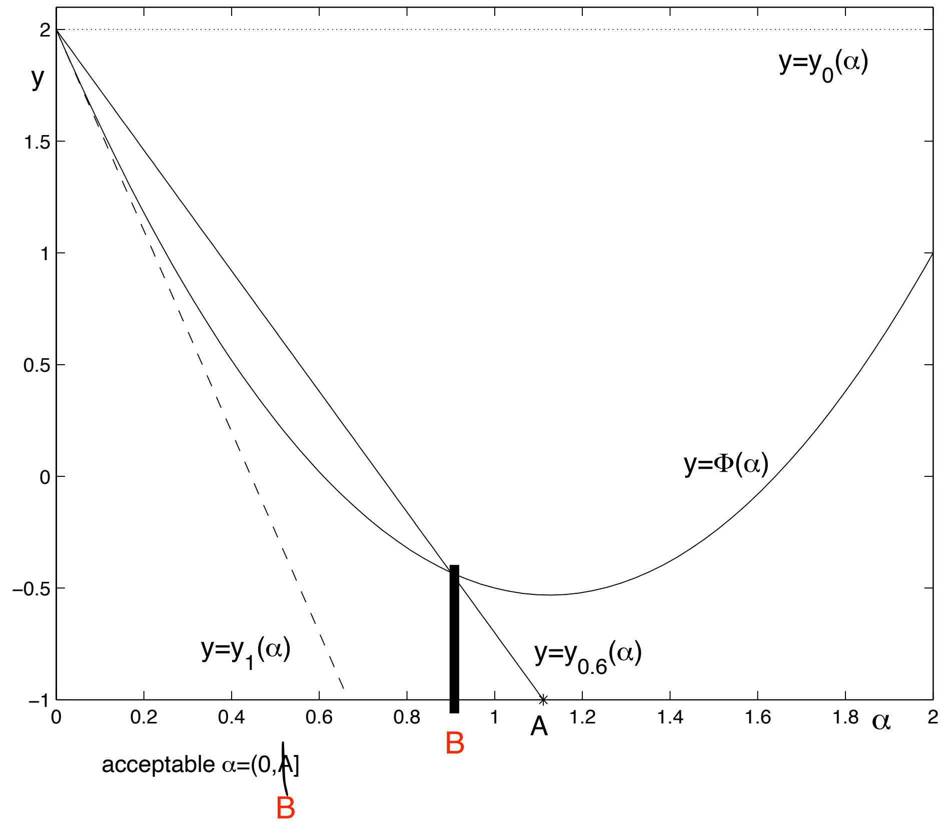
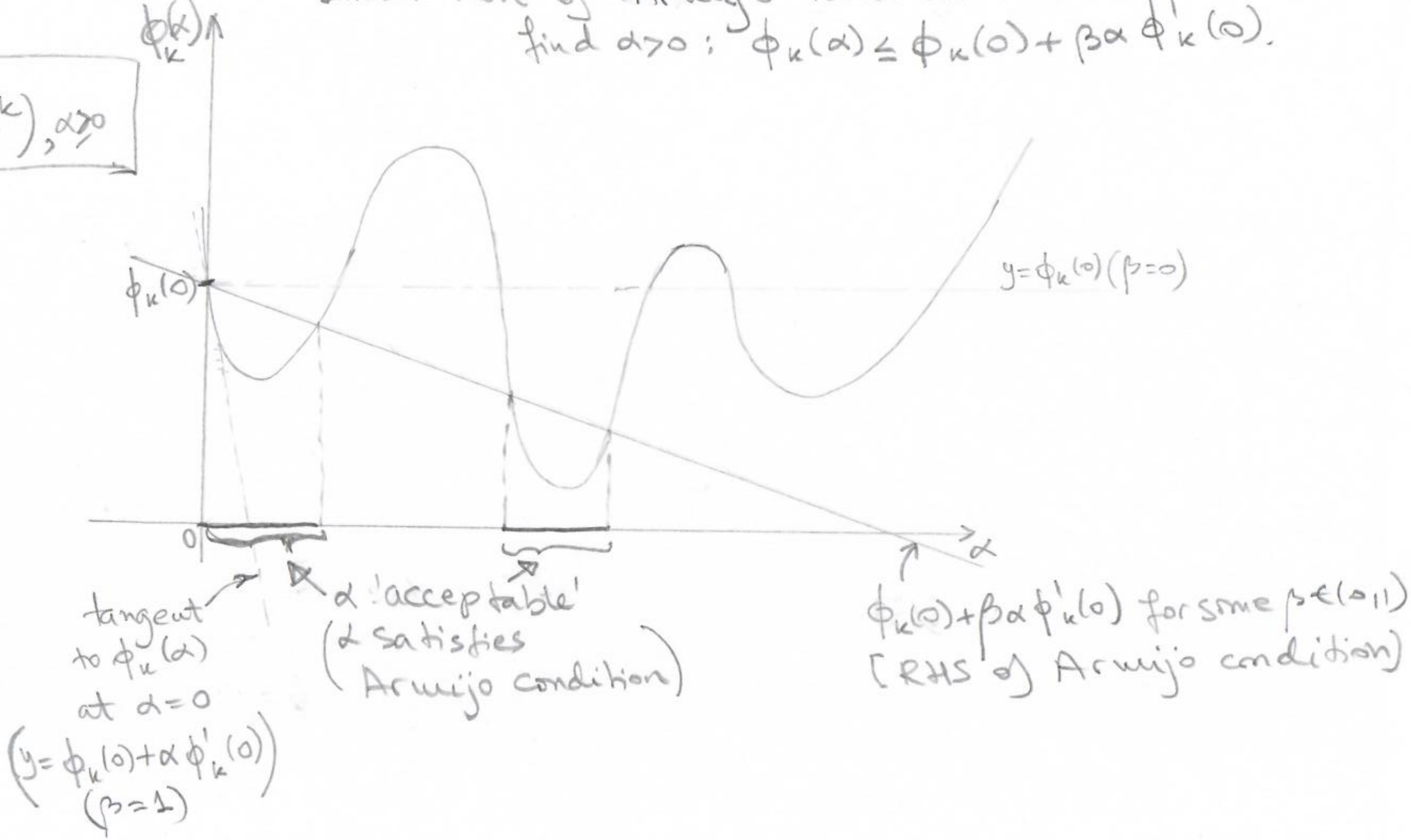


Illustration of Armijo condition

find $\alpha > 0$: $\phi_k(\alpha) \leq \phi_k(0) + \beta\alpha\phi'_k(0)$.

$\boxed{\phi_k(\alpha) = f(x^k + \alpha s^k), \alpha \geq 0}$



Inexact linesearch ...

The backtracking-Armijo (bArmijo) linesearch algorithm

`Choose` $\alpha_{(0)} > 0$, $\tau \in (0,1)$ `and` $\beta \in (0,1)$.

`While` $f(x^k + \alpha_{(i)} s^k) > f(x^k) + \beta\alpha_{(i)} \nabla f(x^k)^T s^k$, `REPEAT:`

- `set` $\alpha_{(i+1)} := \tau\alpha_{(i)}$ `and` $i := i + 1$.

`END.`

`Set` $\alpha^k := \alpha_{(i)}$. $\square$

- $\alpha_{(0)}$, β and τ chosen as before.

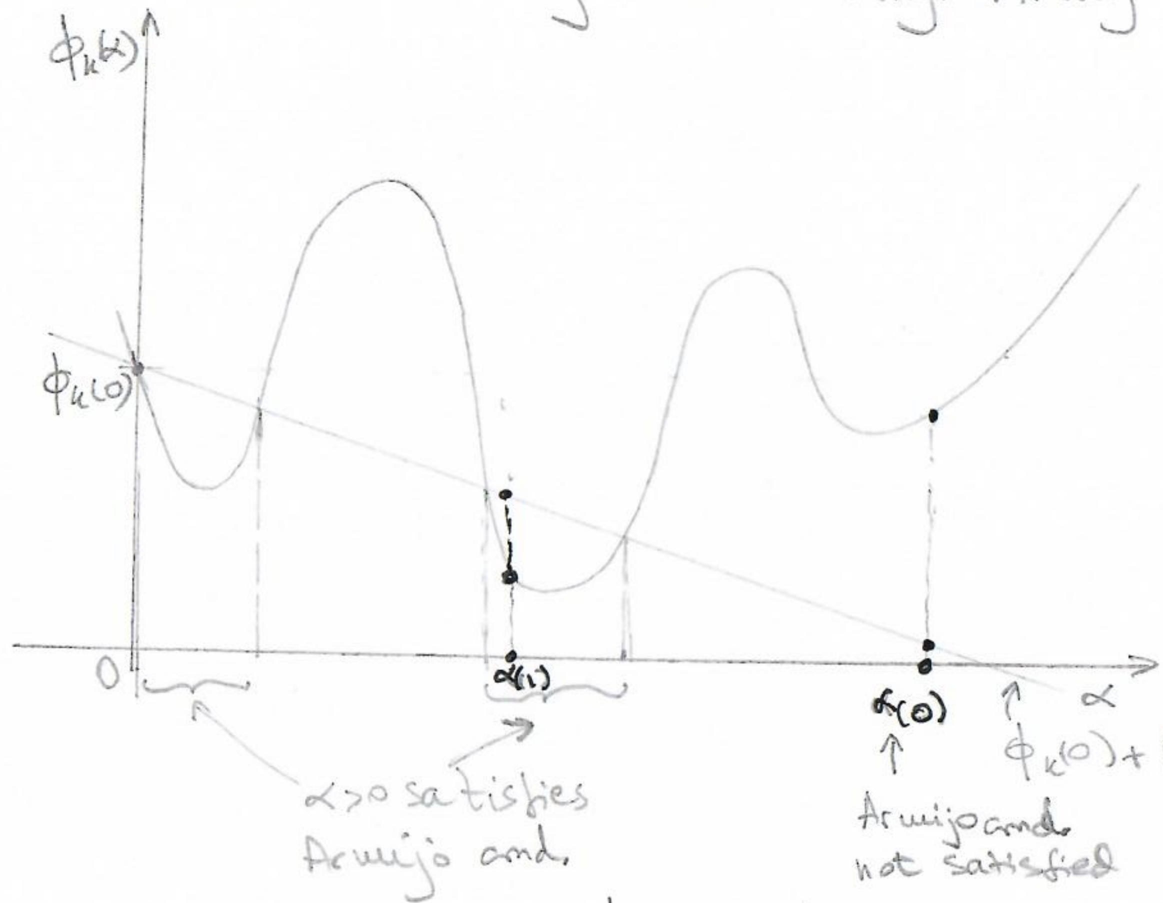
- on each GLM iteration k, the bArmijo linesearch algorithm terminates in a finite number of steps with $\alpha^k > 0$, due to the descent condition. [see Pb Sheet 2]

  [without any additional assumptions on $f \in \mathcal{C}^1$]

- other popular/useful inexact linesearch techniques: Wolfe, Goldstein-Armijo, etc.

Illustration of backtracking-Armijo linesearch

$\phi_k(\alpha) = f(x^k + \alpha s^k)$
$\alpha \geq 0$



$\underbrace{\alpha_{(1)} = \frac{1}{2}\alpha_{(0)}}$ ← backtracking linesearch.

Global convergence of GLM

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

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 :

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

is satisfied for all $\alpha \in [0, \alpha^k_{\max}]$, where

$$\alpha^k_{\max} = \frac{(\beta - 1)\nabla f(x^k)^T s^k}{L\|s^k\|^2}.$$

Global convergence of GLM

- $f \in \mathcal{C}^1(\mathbb{R}^n)$; ∇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

$$\alpha^k_{\max} = \frac{(\beta - 1)\nabla f(x^k)^T s^k}{L\|s^k\|^2}.$$

Proof. Firstly note that $\alpha^k_{\max} > 0$ since $\beta \in (0, 1)$ and s^k is descent direction.

Global convergence of GLM

- $f \in \mathcal{C}^1(\mathbb{R}^n)$; ∇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

$$\alpha^k_{\max} = \frac{(\beta - 1)\nabla f(x^k)^T s^k}{L\|s^k\|^2}.$$

Proof. Firstly note that $\alpha^k_{\max} > 0$ since $\beta \in (0, 1)$ and s^k is descent direction.
First-order Taylor (see Video 2) gives the first equality below: for any $\alpha > 0$ and some $\tilde{\alpha} \in (0, \alpha)$, we have

Global convergence of GLM...

Proof (continued).

$$f(x^k + \alpha s^k) = f(x^k) + \alpha \nabla f(x^k + \tilde{\alpha} s^k)^T s^k$$

$$= f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha [\nabla f(x^k + \tilde{\alpha} s^k) - \nabla f(x^k)]^T s^k$$

$$\leq f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha \|\nabla f(x^k + \tilde{\alpha} s^k) - \nabla f(x^k)\| \cdot \|s^k\|$$

by Cauchy-Schwarz inequality

$$\leq f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha L \|x^k + \tilde{\alpha} s^k - x^k\| \cdot \|s^k\|$$

by Lipschitz continuity of the gradient

$$\leq f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha^2 L \|s^k\|^2,$$

where we used $\tilde{\alpha} \leq \alpha$.

Global convergence of GLM...

Proof (continued).

$$f(x^k + \alpha s^k) = f(x^k) + \alpha \nabla f(x^k + \tilde{\alpha} s^k)^T s^k$$

$$= f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha [\nabla f(x^k + \tilde{\alpha} s^k) - \nabla f(x^k)]^T s^k$$

$$\leq f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha \| \nabla f(x^k + \tilde{\alpha} s^k) - \nabla f(x^k) \| \cdot \| s^k \|$$

by Cauchy-Schwarz inequality

$$\leq f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha L \| x^k + \tilde{\alpha} s^k - x^k \| \cdot \| s^k \|$$

by Lipschitz continuity of the gradient

$$\leq f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha^2 L \| s^k \|^2,$$

where we used $\tilde{\alpha} \leq \alpha$.

Thus Armijo condition (ac) is satisfied for all $\alpha \geq 0$ such that

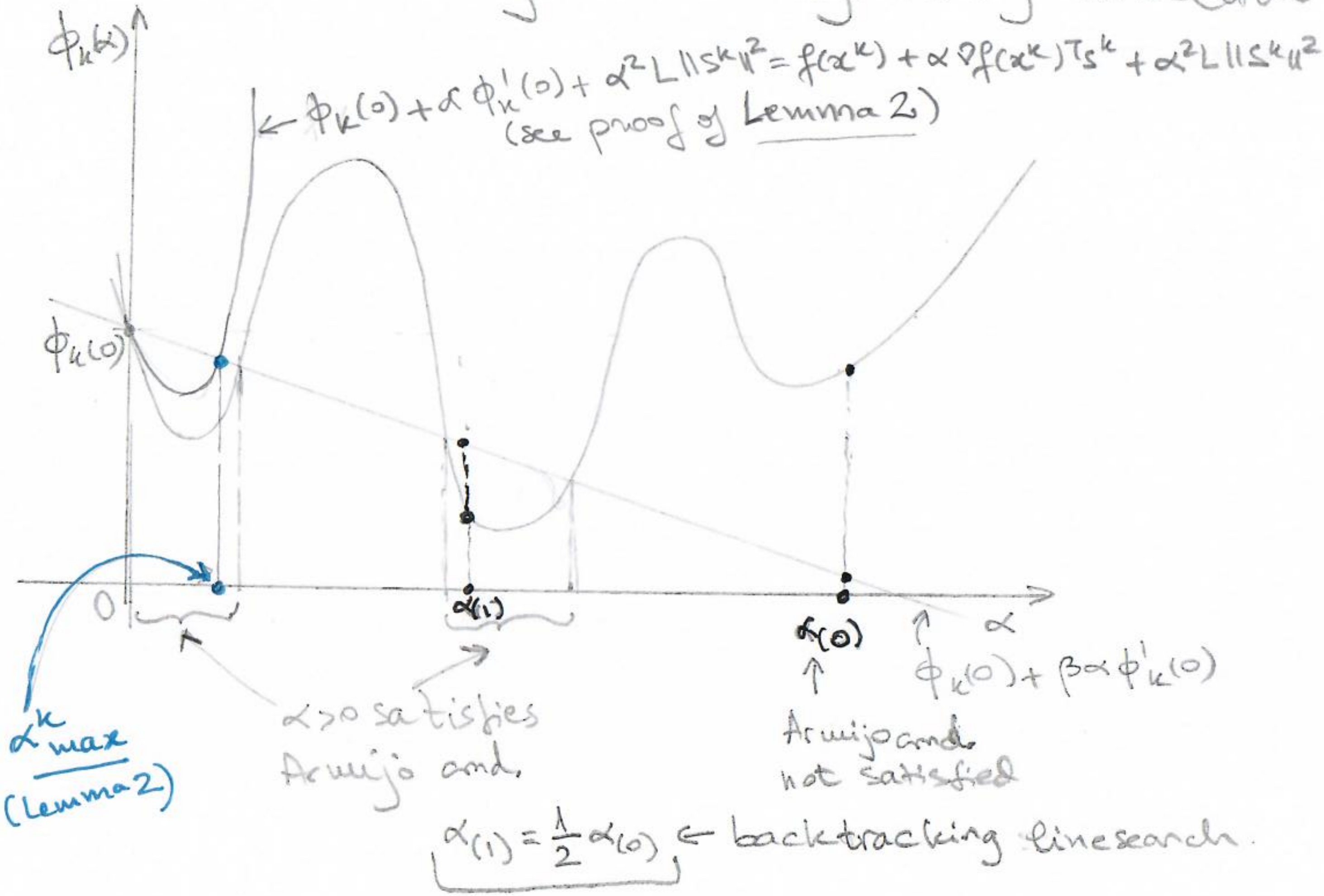
$$f(x^k) + \alpha \nabla f(x^k)^T s^k + \alpha^2 L \| s^k \|^2 \leq f(x^k) + \beta \alpha \nabla f(x^k)^T s^k,$$

which is equivalent to $\alpha \in [0, \alpha^k_{\max}]$. $\Box$

Illustration of backtracking-Armijo linesearch

$\phi_k(\alpha) = f(x^k + \alpha s^k)$

$\alpha \geq 0$



Global convergence of GLM ...

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 ...

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.

Proof of Lemma 3. If $\alpha_{(0)}$ satisfies the Armijo condition (ac), bArmijo terminates with $i = 0$ and $\alpha^k = \alpha_{(0)}$.
Else, bArmijo is guaranteed to terminate as soon as
(by Lemma 2) $\alpha^k \leq \alpha^k_{\max}$. Let $(i-1)$ be the last iteration such that

$$\alpha_{(i-1)} > \alpha^k_{\max} \text{ and } \alpha_{(i)} \leq \alpha^k_{\max}.$$

It follows that

$$\alpha^k = \alpha_{(i)} = \tau\alpha_{(i-1)} > \tau\alpha^k_{\max}.$$

Note that if $\alpha_{(0)} > \alpha^k_{\max}$, then $\alpha_{(i)} = \tau^i\alpha_{(0)} \leq \alpha^k_{\max}$ for any $i \geq \log(\alpha_{(0)}/\alpha^k_{\max})/|\log\tau|$. □

(global convergence of GLM to be continued ...)