LECTURE 13

QUASI-NEWTON METHODS (PART II)

The difficulty with SR-1 method (based on rank one modifications) is that there is no easy way to ensure $H_{k+1} \succ 0$.

DFP (Davidon-Fletcher-Powell) METHOD

Based on rank two modifications to inverse Hessian estimate.

$$(*) \quad H_{k+1} = H_k + \alpha u u^T + \beta v v^T$$

subject to

$$H_{k+1} y_k = s_k$$

$$\Longrightarrow$$

$$\underbrace{H_{k+1} y_k}_{s_k} = H_k y_k + \alpha u (u^T y_k) + \beta v (v^T y_k)$$

unknowns : $u, v \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$

- Choose u so that it captures s_k
  v so that it captures $H_k y_k$

Say

1. $\boxed{u = s_k}$ and $\alpha u^T y_k = 1$

$$\Longrightarrow \boxed{\alpha = \frac{1}{s_k^T y_k}}$$

2. $\boxed{v = H_k y_k}$ and $\beta v^T y_k = -1$

$$\Longrightarrow \boxed{\beta = \frac{-1}{y_k^T H_k y_k}}$$

Plug these values of u, v, α, β in $(*)$.

DFP update rule

$$\boxed{H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}}$$

BFGS (Broyden-Fletcher-Goldfarb-Shanno) METHOD

Based on rank two modifications to Hessian estimate.

$$B_{k+1} = B_k + \alpha u u^T + \beta v v^T$$

subject to

$$B_{k+1} s_k = y_k$$

$$\Longrightarrow$$

$$\underbrace{B_{k+1} s_k}_{y_k} = B_k s_k + \alpha u (u^T s_k) + \beta v (v^T s_k$$

Choose

① $\boxed{u = y_k}$ and $\alpha u^T s_k = 1$

$$\Longrightarrow \boxed{\alpha = \frac{1}{y_k^T s_k}}$$

② $\boxed{v = B_k s_k}$ and $\beta v^T s_k = -1$

$$\Longrightarrow \boxed{\beta = \frac{-1}{s_k^T B_k s_k}}$$

BFGS update rule

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}$$

By letting $H_k = B_k^{-1}$ and $H_{k+1} = B_{k+1}^{-1}$ it can be deduced that

Inverse BFGS update rule

$$H_{k+1} = H_k + \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k} - \left(\frac{s_k y_k^T H_k + H_k y_k s_k^T}{s_k^T y_k}\right.$$

REMARKS

* The DFP Hessian approximation $B_{k+1} = H_{k+1}^{-1}$ is the global solution of the following problem.

$$\text{① } \underset{B \in \mathbb{R}^{n\times n}}{\text{minimize}} \quad \| B - B_k \|$$

subject to

$$B^T = B \text{ and } B s_k = y_k$$

* The BFGS inverse Hessian approximatio H_{k+1} is the global solution of the following problem.

$$\text{② } \underset{H \in \mathbb{R}^{n\times n}}{\text{minimize}} \quad \| H - H_k \|$$

subject to

$$H y_k = s_k \text{ and } H^T = H$$

In problems ① and ② $\|\cdot\|$ is a weighted Frobenius norm of the form

$$\|A\| = \|W A W\|_F$$

for some $W \in \mathbb{R}^{n\times n}$ satisfying

* $W^2 y_k = s_k$ (in ①) OR
* $W^2 s_k = y_k$ (in ②)

PROPOSITION

The following holds for two consecutive BFGS or DFP estimates H_k and H_{k+1}.

$$H_k \succ 0 \quad \text{and} \quad s_k^T y_k > 0$$

$$\Longrightarrow$$

$$H_{k+1} \succ 0.$$

PROOF

Exercise

The condition $s_k^T y_k > 0$ can be ensured within line search.

WOLFE CONDITIONS

$$0 < s_k^T y_k = (x_{k+1} - x_k)^T (\nabla f(x_{k+1}) - \nabla f(x$$

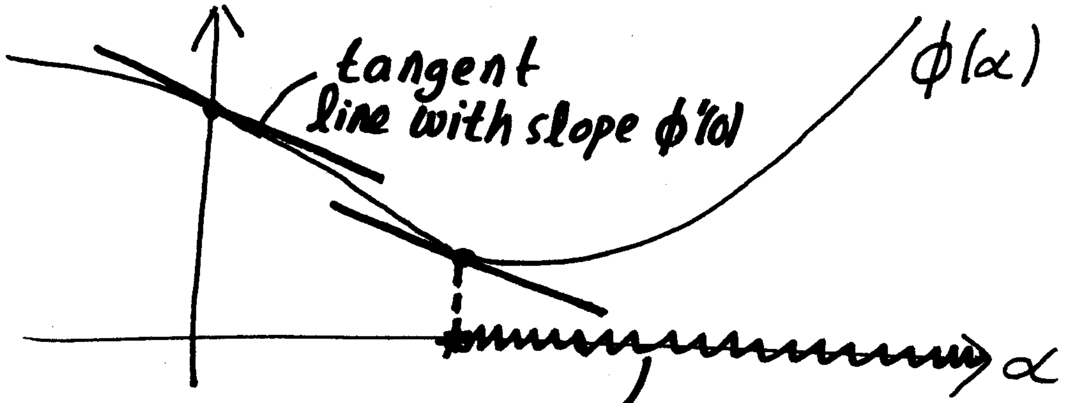
$$= (\alpha_k p_k)^T (\nabla f(x_k + \alpha_k p_k) - \nabla f(x_k$$

$$\Longleftrightarrow$$

$$(**)\ p_k^T \nabla f(x_k + \alpha_k p_k) > p_k^T \nabla f(x_k)$$

In terms of the line search function $\phi(\alpha) = f(x_k + \alpha p_k)$ eqn (**) can be expressed as

$$(***) \quad \phi'(\alpha_k) > \phi'(0)$$



Wolfe conditions require α_k to satisfy

① SUFFICIENT DECREASE (ARMIJO CONDITION)

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

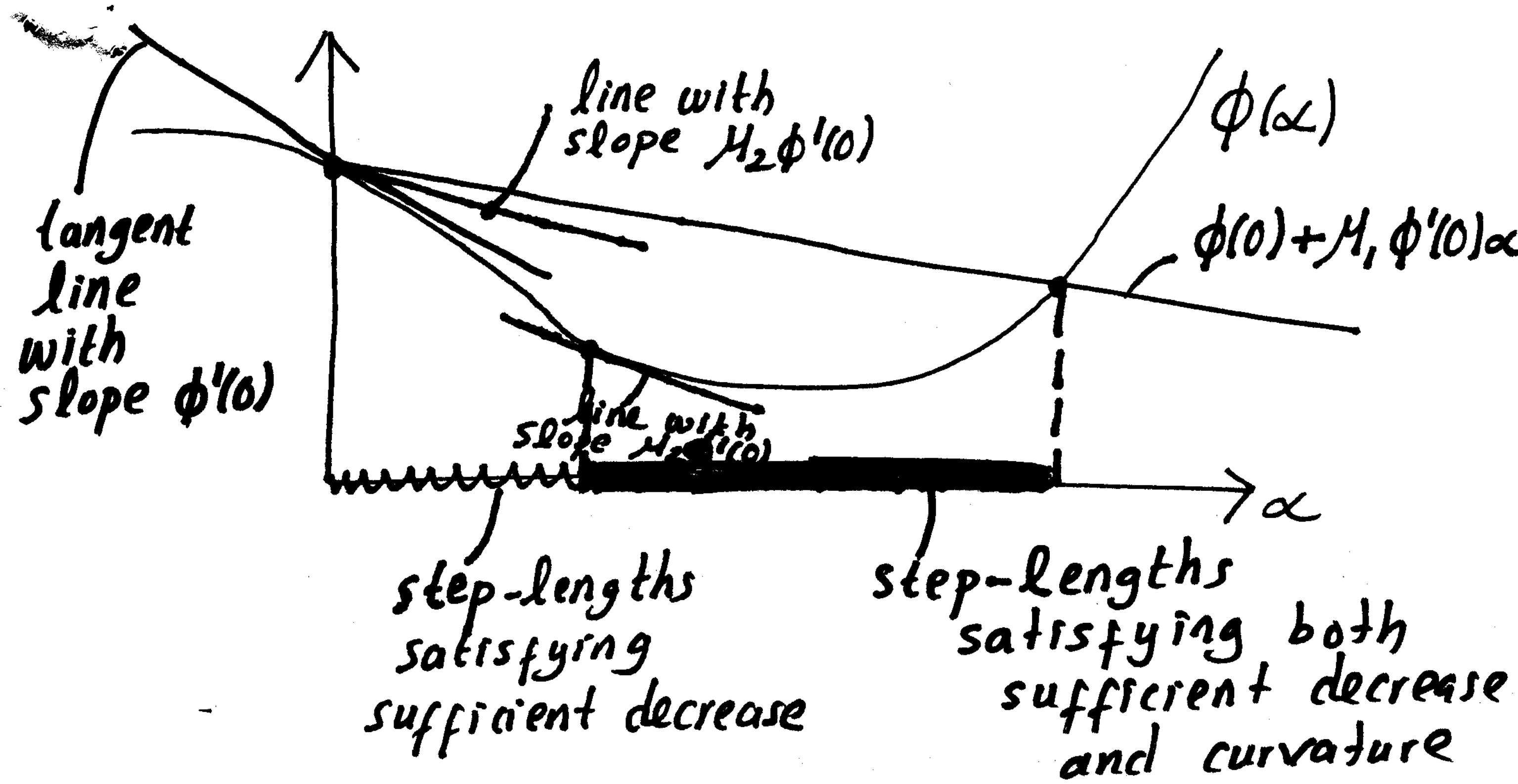
(equivalently $\phi(\alpha_k) \leqslant \phi(0) + \mu_1 \alpha_k \phi'(0)$)

② SUFFICIENT CURVATURE

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geqslant \mu_2 \nabla f(x_k)^T p_k$$

(equivalently $\phi'(\alpha_k) \geqslant \mu_2 \phi'(0)$)

where μ_1, μ_2 are parameters s.t. $0 < \mu_1 < \mu_2 < 1$.



REMARKS

* Sufficient curvature implies $s_k^T y_k > 0$ ensuring that $H_{k+1} \succ 0$ for DFP and BFGS, i.e.,

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq \mu_2 \nabla f(x_k)^T p_k$$
$$> \nabla f(x_k)^T p_k$$
$$\Longleftrightarrow$$
$$\underbrace{\left(\nabla f(x_k + \alpha_k p_k) - \nabla f(x_k)\right)^T}_{y_k^T} \underbrace{(\alpha_k p_k)}_{s_k} > 0$$

* Provided p_k is a descent direction and f is a continuously differentiable function bounded below there exist step-lengths satisfying Wolfe conditions.

THM (STEP-LENGTHS SATISFYING WOLFE CONDITION

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function that is bounded below. Suppose also that $p_k \in \mathbb{R}^n$ is a descent direction. There exist step-lengths $\alpha_k > 0$ satisfying Wolfe conditions.

PROOF

For all small $\alpha > 0$ (as shown in class) the sufficient decrease condition

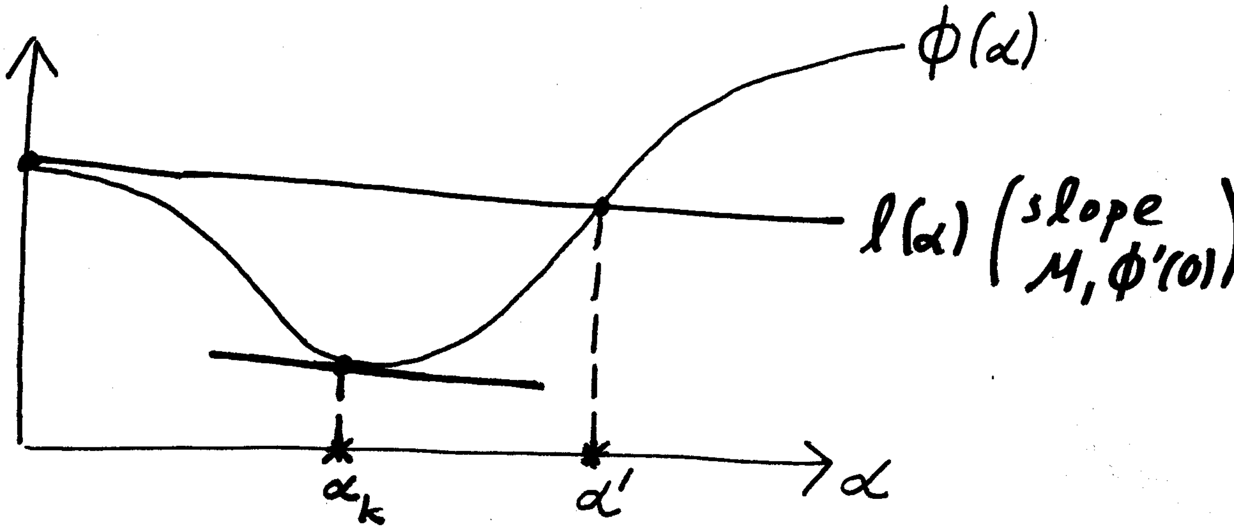
$$\phi(\alpha) \leqslant \underbrace{\phi(0) + \alpha \mu_1 \phi'(0)}_{l(\alpha)}$$

holds where $\phi(\alpha) = f(x_k + \alpha p_k)$.

But since $\mu_1 \phi'(0) < 0$, as $\alpha \to \infty$ $l(\alpha) \to -\infty$ whereas $\phi(\alpha) \geqslant L$ for some constant L (i.e. f is bounded below).

Therefore for small $\alpha > 0$
$$\phi(\alpha) \leqslant l(\alpha)$$
while for large $\alpha > 0$
$$\phi(\alpha) \geqslant l(\alpha).$$
By continuity of ϕ and l there exists an α' ~~$\in (0, \alpha)$~~ such that

$$\phi(\alpha') = l(\alpha')$$



By the mean value thm there exists an $\alpha_k \in (0, \alpha')$ such that

$$\frac{\phi(\alpha') - \phi(0)}{\alpha' - 0} = \phi'(\alpha_k)$$

$$\Longrightarrow$$

$$\frac{\cancel{\phi(0)} + \cancel{\alpha}\mu_1 \phi'(0) - \cancel{\phi(0)}}{\cancel{\alpha}} = \phi'(\alpha_k)$$

$$\Longrightarrow \quad (\text{since } \mu_2 < \mu_1)$$

$$\mu_2 \phi'(0) < \phi'(\alpha_k)$$

Both sufficient decrease and curvature conditions hold at α_k. $\square$

ALGORITHM (BFGS)

Given $x_0 \in \mathbb{R}^n$, $k=0$, $H_0 = I_n$

While $\|\nabla f(x_k)\| > \epsilon$

$$p_k = -H_k \nabla f(x_k)$$

Choose a step-length $\alpha_k > 0$ satisfying Wolfe conditions along p_k

$$x_{k+1} = x_k + \alpha_k p_k$$

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

$$H_{k+1} = H_k + \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right)\frac{s_k s_k^T}{s_k^T y_k} - \left(\frac{s_k (y_k^T H_k) + (H_k y_k) s}{s_k^T y_k}\right.$$

$k = k+1$

end

return x_k

EXAMPLE

Apply one iteration of BFGS to

$$f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$$

starting with $x_k = (0,0)$ and $H_k = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$

Use exact line searches.

① Search direction

$$\nabla f(x) = \begin{bmatrix} 2(x_1 - 2) \\ 2(x_2 - 1) \end{bmatrix} \Longrightarrow \nabla f(x_k) = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$$

$$p_k = -H_k \nabla f(x_k)$$

$$= \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$$

② Step-lengths

Recall that when f is a quadratic function exact step-length is given by

$$\alpha_k = \frac{-\phi'(0)}{p_k^T \nabla^2 f(x_k) p_k}$$

$$= \frac{-\nabla f(x_k)^T p_k}{p_k^T \nabla^2 f(x_k) p_k}$$

Noting $\nabla^2 f(x_k) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

$$\alpha_k = \frac{44}{200} = \frac{11}{50}$$

Consequently

$$x_{k+1} = x_k + \alpha_k p_k = \begin{bmatrix} 44/25 \\ 33/25 \end{bmatrix}$$

3 Update inverse Hessian approximation

$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

$$= \begin{bmatrix} 88/25 \\ 34/25 \end{bmatrix}$$

$$s_k = x_{k+1} - x_k = \begin{bmatrix} 44/25 \\ 33/25 \end{bmatrix}$$

Since $s_k^T y_k > 0$, $H_{k+1} \succ 0$

$$H_{k+1} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} + \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k} - \left(\frac{s_k (y_k^T H_k) + (H_k y_k) s_k^T}{s_k^T y_k}\right)$$