Lecture 13

How to update the matrices B_k ?

Symmetric Rank One Updates

Dyadic Product of Two Vectors

$$ab^T := \begin{pmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n b_1 & a_n b_2 & \cdots & a_n b_n \end{pmatrix}$$

Note that the rank of ab^T is 1 if $a \neq 0$ and $b \neq 0$.

SR1 - Updates

The SR1-method makes the following update:

$$(1) \quad \boxed{B_{k+1} = B_k + \sigma \cdot \underbrace{v \cdot v^T}_{\text{dyadic product}},}$$

where $\sigma = \pm 1$ and $v \in \mathbb{R}^n$ s.t. B_{k+1} satisfies the secant equation

$$(2) \quad \boxed{y^k = B_{k+1} s^k}$$

Multiplying (1) with s^k yields:

$$B_{k+1} s^k \overset{(2)}{=} y^k = B_k s^k + \sigma v v^T s^k$$

$$(3) \qquad = \underline{\underline{B_k s^k}} + \big(\underbrace{\sigma v^T s^k}_{\text{scalar}}\big) v$$

$\Rightarrow$ v must be a multiple of $y^k - B_k s^k$:

$$v = \delta (y^k - B_k s^k).$$

Substituting v in (3) gives:

$$\begin{aligned} y^k - B_k s^k &= (\sigma v^T s^k) v \\ &= \sigma\, \delta (y^k - B_k s^k)\, \delta (y^k - B_k s \\ &= \underbrace{\sigma \delta^2 (s^{kT}(y^k - B_k s^k))}_{=1} \cdot (y^k - B_k \end{aligned}$$

$$\Rightarrow \delta^2 = \frac{1}{\sigma (s^{kT}(y^k - B_k s^k))}$$

$$\Rightarrow \sigma = \operatorname{sign}(s^{kT}(y^k - B_k s^k))$$

$$\Rightarrow \delta = \pm | s^{kT}(y^k - B_k s^k)|^{-1/2}$$

Thus, (1) becomes

$$(4) \quad \boxed{\begin{aligned} B_{k+1} &= B_k + \sigma \cdot \delta^2 \cdot (y^k - B_k s^k)(y^k - B_k s^k \\ &= B_k + \frac{(y^k - B_k s^k)(y^k - B_k s^k)^T}{(y^k - B_k s^k)^T s^k} \end{aligned}}$$

(4) is a generalization of the secant method to multidimensional problems.

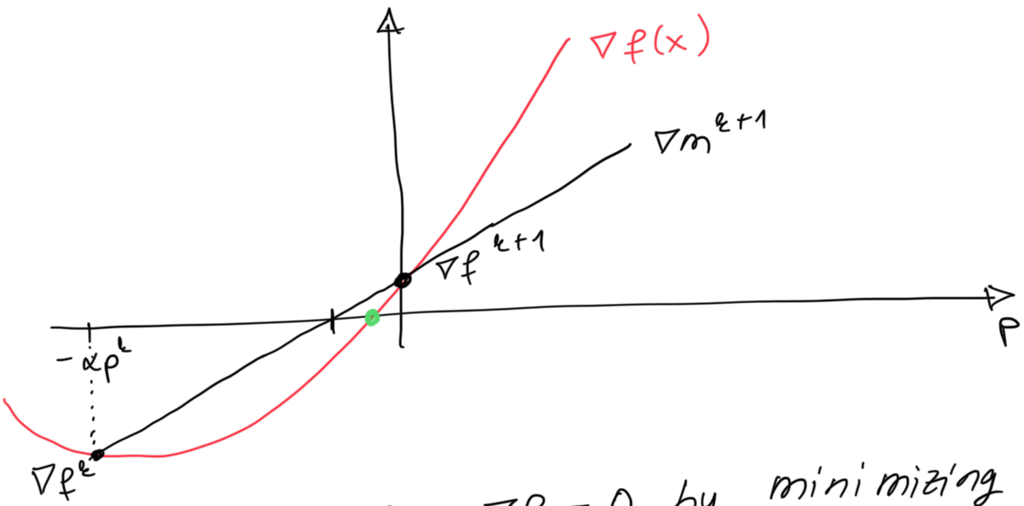
$$x^k \longrightarrow x^{k+1} = x^k + \alpha_k p^k$$

$$p^k = \operatorname{argmin} m^k(p) \qquad p^{k+1} = \operatorname{argmin} m^{k+1}(p$$

$$\boxed{m^k(p) = f^k + \nabla f^{k^T} p + \frac{1}{2} p^T B_k p}$$

$$\boxed{\begin{aligned} m^{k+1}(p) = f^{k+1} + \nabla f^{k+1^T} p \\ + \frac{1}{2} p^T B_{k+1} p \end{aligned}}$$

$$\left.\begin{aligned} &\cdot\ \nabla m^{k+1}(0) = \nabla f(x^{k+1}) \ \checkmark \\ &\cdot\ \nabla m^{k+1}(-\alpha_k p^k) = \nabla f(x^k) \end{aligned}\right\} \text{ these give the secant equation}$$



$\Rightarrow$ We search for $\nabla f = 0$ by minimizing a quadr. approximation, thus $\nabla m^{k+1} = 0$.

3 Possible cases:

1. If $(y^k - B_k s^k)^T s^k \neq 0$, then the argument above show that there is a unique rank-1 updating formula satisfying the secant equation. This formula is given by (4).

only updating

2. If $y^k = B_k s^k$, then the only updating formula satisfying the secant equation is simply $B_{k+1} = B_k$.

3. If $y^k \neq B_k s^k$ and $(y^k - B_k s^k)^T s^k = 0$, then there is no symmetric rank 1 updating formula satisfying the secant equation.

<u>Algorithm</u>

- Initialize B_0, x^0; $k = 0$
- Compute $p^k = -B_k^{-1} \nabla f(x^k)$; Find step length $\alpha_k > 0$ and set $x^{k+1} = x^k + \alpha_k p^k$.
- Compute B_{k+1} according to (4); set $k := k+1$ and repeat.

<u>SR1 - Update : Convergence</u>

For general nonlinear functions, the SR1-update generates good Hessian approximations under certain conditions:

<u>Theorem</u>: Suppose that f is twice cont. diff., and that the Hessian is bounded and Lipschitz-cont. in a neighborhood of x^*. Let $\{x^k\}$ be any sequence of iterates s.t. $x^k \to x^*$. Suppose in addition that

$$|(s^k)^T (y^k - B_k s^k)| \geq r \cdot \|s^k\| \cdot \|y^k - B_k s$$

$\|(s \quad (y \quad \quad _k)$

for all k, for some $r \in (0,1)$, and that the steps s^k are uniformly linearly independent. Then the matrices B_k generated by the SR1-updating formula satisfy

$$\boxed{\lim_{k \to \infty} \| B_k - \nabla^2 f(x^*) \| = 0 .}$$

Symmetric Rank 1 Updates of the Inverse

Lemma: (Sherman-Morrison-Woodbury Formula)

If $\overset{B_{k+1}}{\bar{A}} = \overset{B_k}{A} + \overset{\sigma v v^T}{a b^T}$, then

$$\boxed{\bar{A}^{-1} = A^{-1} - \frac{A^{-1} a b^T A^{-1}}{1 + b^T A^{-1} a}}.$$

Proof:

$$\bar{A} \cdot \bar{A}^{-1} = (A + ab^T) \cdot \left(A^{-1} - \frac{A^{-1} a b^T A^{-1}}{1 + b^T A^{-1} a} \right)$$

$$= \underbrace{A A^{-1}}_{= I} + ab^T A^{-1} - \frac{\overset{= I}{A A^{-1}}\, a b^T A^{-1}}{1 + b^T A^{-1} a} - \frac{a b^T A^{-1} a b^T A^{-1}}{1 + b^T A^{-1} a}$$

$$= I + \frac{ab^T A^{-1} \cdot (1 + b^T A^{-1} a)}{1 + b^T A^{-1} a} - \frac{a b^T A^{-1}}{1 + b^T A^{-1} a} - \frac{a b^T A^{-1} a b^T}{1 + b^T A^{-1}}$$

$$= I + \frac{a b^T A^{-1}}{1 + b^T A^{-1} a} + \frac{a b^T A^{-1} b^T A^{-1} a}{1 + b^T A^{-1} a} - '' - ''$$

$$= I \quad \blacksquare$$

Let $H_k := B_k^{-1}$. Then we get using the above formula:

$$(5) \quad \boxed{H_{k+1} = H_k + \frac{(s^k - H_k y^k)(s^k - H_k y^k)^T}{(s^k - H_k y^k)^T y^k}}$$

Algorithm:

- Initialize H_0, x^0; $k = 0$
- Compute $p^k = - H_k \nabla f(x^k)$; Find step length $\alpha_k > 0$ and set $x^{k+1} = x^k + \alpha_k p^k$.
- Compute H_{k+1} according to (5); set $k := k+1$ repeat.

Below, we perform the above algorithm with $\alpha_k = 1 \,\forall$

Theorem: Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is the function $f(x) = b^T x + \frac{1}{2} x^T A x$, where A is symm. pos. def. Then for any starting point x^0 and any symmetric starting matrix H_0, the iterates $\{x^k\}$ generated by the SR1-method converge to the minimizer in at most n steps, provided that $\boxed{(s^k - H_k y^k)^T y^k \neq 0}$ for all k.

Moreover, if n steps are performed, and if the search directions p^i are linearly independent, then $H_n = A^{-1}$.

Proof: Let us first note that the update is well-defined, as $(s^k - H_k y^k)^T y^k \neq 0$ ... that

always ... We now show that
for all k. We now show that

(6) $\quad H_k y^j = s^j \, , \quad j = 0, \dots, k-1,$

which means that the secant equation is fulfilled along <u>all</u> previous directions.

Induction:

- $k = 1$: We have $H_1 y^0 = s^0$ holds true because of the definition of the SR1-update.

- Assume that (6) holds for some $k \geq 1$.
- We show that (6) holds for $k+1$.

We have:

$$(s^k - H_k y^k)^T y^j = s^{k^T} y^j - y^{k^T} \underbrace{H_k y^j}_{\substack{= s^j \\ \text{because of (6)} \\ \text{for } k}}$$

$$= s^{k^T} y^j - y^{k^T} s^j .$$

It hold:

$$y^j = \nabla f^{j+1} - \nabla f^j$$
$$= \cancel{b} + A^T x^{j+1} - (\cancel{b} + A^T x^j)$$
$$= A^T (\underbrace{- x^j + x^{j+1}}_{= + s^j})$$
$$= + A^T s^j .$$

Therefore, we obtain:

$$s^{k^T} (+ A^T s^j) - (+ A^T s^k) s^j = \underline{\underline{0}} .$$

Hence,

$$\boxed{\forall j<k: \quad (s^k - H_k y^k)^T y^j = 0.}$$

Furthermore,

$$\forall j<k: \quad H_{k+1} y^j = H_k y^j + \frac{(s^k - H_k y^k)(s^k - H_k y^k)^T y^j}{\underbrace{(s^k - H_k y^k)^T y^k}_{=0}}$$

$$= H_k y^j$$

$$\overset{(6)}{=} s^j$$

Since $H_{k+1} y^k = s^k$ (secant equation), (6) holds if k is replaced by $k+1$. By induction, (6) holds for all k.

If the algorithm performs n steps and if these steps s^i ($s^i = x^{i+1} - x^i$) are linearly independent, we have:

$$(7) \quad \forall j = 0, 1, \dots, n-1: \quad s^j \overset{(6)}{=} H_n A s^j,$$

as

$$y^j = \nabla f^{j+1} - \nabla f^j$$
$$= \cancel{b} + A x^{j+1} - (\cancel{b} + A x^j)$$
$$= A(x^{j+1} - x^j)$$
$$= A s^j.$$

By (7), we have $H_n A = I$, or, $H_n = A^{-1}$. This means that the step taken at x^n is the Newton step, and the next iterate x^{n+1} will be the

so, the . This means that the algorithm
solution. This means that the algorithm terminates after n steps.

In case that the steps become linearly dependent, let us assume that s^k is a linear combination of the previous steps, that is,

$$(8) \qquad \boxed{s^k = \varsigma_0 s^0 + \dots + \varsigma_{k-1} s^{k-1}}$$

for $\varsigma_i \in \mathbb{R}$, $i = 0, \dots, k-1$.

It follows that

$$\begin{aligned}
\underline{\underline{H_k y^k}} &\overset{y^k = As^k}{=} H_k A s^k \\
&\overset{(8)}{=} \varsigma_0 H_k A s^0 + \dots + \varsigma_{k-1} H_k A s^{k-1} \\
&\overset{y^i = As^i}{=} \varsigma_0 H_k y^0 + \dots + \varsigma_{k-1} H_k y^{k-1} \\
&\overset{(6)}{=} \varsigma_0 s^0 + \dots + \varsigma_{k-1} s^{k-1} \\
(9) \qquad &\overset{(8)}{=} \underline{\underline{s^k}}
\end{aligned}$$

Further:

$$\begin{aligned}
H_k y^k &= \underline{H_k (\nabla f^{k+1} - \nabla f^k)} \\
&\overset{(9)}{=} s^k
\end{aligned}$$

$$\begin{aligned}
p^k = s^k = x^{k+1} - x^k \quad &\text{as } \alpha_k = 1 \\
= p^k \\
= \underline{-H_k \nabla f^k}
\end{aligned}$$

Since H_k is nonsingular, we conclude with $\nabla f^{k+1} = 0 \implies x^{k+1}$ is ■

a minimizer.