§ 4.4. Quasi-Newton Methods

We can avoid derivatives (even $f(x_k)$ can be difficult to evaluate, but $f'(x_k)$ may be way harder).

Idea: replace $f'(x_k)$ with some g_k:

Iterations $x_{k+1} = x_k - \dfrac{f(x_k)}{g_k}$ w/ $g_k \approx f'(x_k)$

are called quasi-Newton methods.

(§ 4.4.2) Constant Slope Method

Idea: $g_k = g = f'(x_0)$, i.e. f' is evaluated once!

$$\Rightarrow \boxed{x_{k+1} = x_k - \frac{f(x_k)}{g}, \quad k = 0, 1, 2, \ldots} \qquad (1)$$

(if $f'(x)$ does not change much, (1) can be close to Newton's method)

(1) is called the constant slope method.

What about convergence? At the kth iteration,

$$(2) \quad f(x_k) = \underbrace{f(x^*)}_{=0} + f'(x^*)\underbrace{(x_k - x^*)}_{\text{Error } e_k} + O\left((x_k - x^*)^2\right)$$

Also: $\underbrace{x_{k+1} - x^*}_{e_{k+1}} = \underbrace{(x_k - x^*)}_{e_k} - \dfrac{f(x_k)}{g}$ from (1)

(if we subtract x^* from (1))

$$\Rightarrow \quad e_{k+1} = e_k - \underbrace{\frac{f(x_k)}{g}}_{=\frac{f'(x^*)}{g} \cdot e_k + O(e_k^2)} = e_k \left(1 - \frac{f'(x^*)}{g}\right) + O(e_k^2)$$

from (2) $\rightarrow$ $\dfrac{f'(x^*)}{g} \cdot e_k + O(e_k^2)$

If $\left|1-\frac{f'(x^*)}{g}\right| < 1$ then for x_0 suff. close to x^*, $e_{k+1} < e_k$ $\left(\text{w/ } O(e_k^2) \xrightarrow[k\to\infty]{} 0\right)$ ②

$\Rightarrow$ therefore, convergence is linear, i.e. w/ error reduced by a factor <1 at each step.

<u>Modification</u>: we may update g occasionally.

Say, $g = f'(x_0)$ to start, and then, if the method slows down, change g to be $f'(x_k)$ at the kth step.

(§ 4.4.3) <u>Secant Method.</u>

$$x_{k+1} = x_k - \frac{f(x_k)}{g_k}, \quad \text{where } g_k = \frac{f(x_k)-f(x_{k-1})}{x_k - x_{k-1}}$$

$k = 1, 2, \ldots$

slope of the secant line through $(x_{k-1}, f(x_{k-1}))$ and $(x_k, f(x_k))$

for $x_{k-1} \xrightarrow[k\to\infty]{} x_k$,

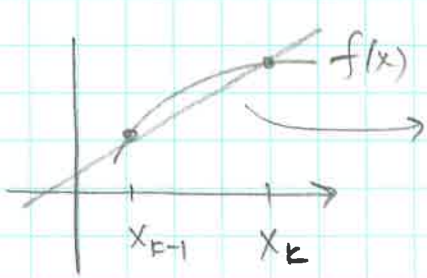
$g_k \to f'(x_k)$

$\Rightarrow$ g_k is a reasonable approximation of $f'(x_k)$.

So, altogether, the secant method is

$$(3) \quad \boxed{x_{k+1} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})}, \quad k = 1, 2, \ldots}$$

One need x_0 & x_1 to start the method!



<u>Example</u>: $f(x) = x^2 - 2$, $x^* = \sqrt{2}$ (one of 2 roots)

Start w/ $x_0 = 1$, $x_1 = 2$ then

$$x_2 = x_1 - \frac{f(x_1)(x_1 - x_0)}{f(x_1) - f(x_0)} = 2 - \frac{2(2-1)}{2-(-1)} = \frac{4}{3} \approx 1.3333$$

$$e_2 = x_2 - x^* = \frac{4}{3} - \sqrt{2} = -0.0809$$

$x_3 = 1.4$ w/ $e_3 = -0.0142$, etc. (see p. 91)

<u>Interesting</u>: The convergence rate is faster than linear ($\frac{|e_{k+1}|}{|e_k|}$ is decreasing w/ k) but slower than quadratic ($\frac{|e_{k+1}|}{e_k^2}$ is growing w/ k). The order, in fact, is $\frac{1+\sqrt{5}}{2} \approx 1.62$! How? There is

<u>Lemma (4.4.1)</u>

If $f \in C^2$, and if x_0 & x_1 are sufficiently close to a root x^*, and if $f'(x^*) \neq 0$, then the error e_k in the secant method (3) satisfies

$$\lim_{k \to \infty} \frac{e_{k+1}}{e_k e_{k-1}} = C^* = \frac{f''(x^*)}{2f'(x^*)}$$

(That is, $e_{k+1} \approx C^* e_k e_{k-1}$ for large k)

(See proof on p. 92)

<u>Note</u>: $|e_{k+1}| \approx O(|e_k|^\alpha)$. Why?

w/ $\alpha = \frac{1+\sqrt{5}}{2}$

If we assume that $|e_{k+1}| \approx a\,|e_k|^{\alpha}$

$$\Rightarrow \; |e_k| \approx a\,|e_{k-1}|^{\alpha} \; \Rightarrow$$

$$|e_{k-1}| \approx \left(\frac{|e_k|}{a}\right)^{1/\alpha}$$

From lemma, assuming $e_{k+1} \approx C^* e_k e_{k-1}$

gives: $$\underbrace{a\,|e_k|^{\alpha}}_{e_{k+1}} \approx C^*\,|e_k| \underbrace{\left(\frac{|e_k|}{a}\right)^{1/\alpha}}_{|e_{k-1}|}$$

$$= C^*\,|e_k|^{1+1/\alpha}\, a^{-1/\alpha} \Rightarrow$$

$$\underbrace{|e_k|^{1+1/\alpha-\alpha}}_{\substack{\text{must be} \\ \text{indep. of } k \\ \text{too}}} \approx \underbrace{\frac{a^{1+1/\alpha}}{C^*}}_{\text{independent of } k!}$$

as $k \to \infty$, this is $O(1)$

$$\Rightarrow \quad 1 + \frac{1}{\alpha} - \alpha = 0$$

$$\alpha + 1 - \alpha^2 = 0 \Rightarrow \alpha = \frac{1 \pm \sqrt{5}}{2}$$

For convergence, need $\alpha > 0 \Rightarrow \alpha = \frac{1+\sqrt{5}}{2}$.

- Secant method can be combined w/ bisection method if you can find x_{k-1}, x_k s.t. $f(x_{k-1})f(x_k) < 0$ to guarantee to be close to x^*. This "combo" has a name: the regula falsi algorithm ("false position") Converges <u>linearly</u>.