SW9

Speed of convergence:

- Linear Convergence: ← Gradient descent The distance to the optimal is getting better every step. The speed is depending c

$\| x^* - x^{i+1} \| \leq c \, \| x^* - x^i \|$

- Superlinear convergence: ← Newton's method

Repetition of the sequence $\{c_i\}_{i \in \mathbb{N}}$ with $\lim_{n \to \infty}$ times (n)

- Quadratic convergence ← Broyden's method

There is $c > 0$ and $i_0 \in \mathbb{N}$ such that for all $i \geq i_0$ we have

$\| x^* - x^{i+1} \| \leq c \, \| x^* - x^i \|^2$

Approximating partial derivatives:

Sometimes it's hard to find the derivatives. Then just get two points of the function and calculate the differents over the stepsize. To get the best appox. get the point you want the derivative and go one step right, one step left and divide by two step sizes:

$$\frac{\partial f(x)}{\partial x_i} \approx \frac{f(x_1, x_2, \dots x_i + \epsilon \dots x_n) - f(x_1, x_2, \dots x_i - \epsilon \dots x_n)}{2\epsilon}$$

Second derivative:

$$\frac{\partial^2 f(x)}{\partial x_i^2} \approx \frac{f(x_1 \dots x_i + \epsilon \dots x_n) - 2f(x_1, \dots x_n) + f(x_1 \dots x_i - \epsilon \dots x_n)}{\epsilon^2}$$

→ Derivative of the first expression.

Second partial derivative with two variables:

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \approx \frac{1}{4\epsilon^2} \cdot \Big(f(x_1 \dots x_i + \epsilon \dots x_j + \epsilon \dots x_n) - f(x_1 \dots x_i + \epsilon \dots x_j - \epsilon \dots x_n) - f(x_1 \dots x_i - \epsilon \dots x_j + \epsilon \dots x_n) + f(x_1 \dots x_i - \epsilon \dots x_j - \epsilon \dots x_n) \Big)$$

Broyden's method:

idea: Computing and inverting the Hessian-Matrix is expensive. in this method, we do just an approximation.

Hence it make sense to require that the approximation A^i satisfies:

$$A^i(x^i - x^{i-1}) = \nabla f(x^i) - \nabla f(x^{i-1})$$

Broyden's method fullfills this fact with following equation:

$$A^i = A^{i-1} + \frac{\left(\left(\nabla f(x^i) - \nabla f(x^{i-1})\right) - A^{i-1}(x^i - x^{i-1})\right)(x^i - x^{i-1})^T}{\| x^i - x^{i-1} \|^2}$$

And without getting the invert A^{-1} in each step, we can compute $(A^i)^{-1}$ directly from $(A^{i-1})^{-1}$:

$$(A^i)^{-1} = \left(A^{i-1} + \frac{(g^i - A^{i-1}d^i)\, d^{iT}}{\|d^i\|^2}\right)^{-1} \qquad \text{//Sherman-Morrison}$$

$$\overset{SM}{=} A^{-1} - \frac{A^{-1} v u^T A^{-1}}{1 + u^T A^{-1} v}$$

! Final Formula / procedure of the Broyden's method:

- Initialization: Start with x^0
  Compute $\nabla f(x^0)$ and the inverse of the Hessian matrix $H_f(x^0)$, and set $(A^0)^{-1} := (H_f(x^0))^{-1}$. Use this to compute
  $$x^1 = x^0 - (A^0)^{-1} \nabla f(x^0)$$
- Iteration step. To compute x^{i+1} from x^i, compute first $\nabla f(x^i)$, $g^i = \nabla f(x^i) - \nabla f(x^{i-1})$ and $d^i = x^i - x^{i-1}$, then
  $$(A^i)^{-1} = (A^{i-1})^{-1} - \frac{\left((A^{i-1})^{-1} g^i - d^i\right)(d^i)^T (A^{i-1})^{-1}}{(d^i)^T (A^{i-1})^{-1} g^i}$$
  and set $x^{i+1} = x^i - (A^i)^{-1} \nabla f(x^i)$

Aitken's acceleration method:

if the scalar sequence $\{x^i\}_{i\in\mathbb{N}}$ converges linearly toward x^*, then

$$\frac{x^* - x^{i-1}}{x^* - x^{i-2}} \approx \frac{x^* - x^i}{x^* - x^{i-1}}$$ // Konvergiert gleichmässig zu x^*

Assuming this, we get $$x^* = x^i - \frac{(x^i - x^{i-1})^2}{x^i - 2x^{i-1} + x^{i-2}}$$

Now create a new sequence $\{y^i\}_{i\in\mathbb{N}}$ from the previous sequence $\{x^i\}_{i\in\mathbb{N}}$.

$$y^i = x^i - \frac{(x^i - x^{i-1})^2}{x^i - 2x^{i-1} + x^{i-2}} = x^i - \frac{(\Delta x^i)^2}{\Delta^2 x^i}$$

// with this, $\{y^i\}_{i\in\mathbb{N},\, i\geq 2}$ converges quadratically towards the same limit x^*.