Newton's Method

Consider

$$g(x) = x - \Lambda f(x)$$

for some $\Lambda \in \mathbb{R}^{n \times n}$.

Assuming Λ is invertible,

$$g(x_*) = x_* \iff$$

$$x_* - \Lambda f(x_*) = x_* \iff f(x_*) = 0$$

Choose Λ so that $J_g(x_*) = 0$, i.e.,

$$0 = J_g(x_*) = I_n - \Lambda J_f(x_*)$$

$$\iff (+)\ \Lambda = J_f(x_*)^{-1}$$

However, x_* is not known, so it is not practical to choose based on (+).

~~Instead let us consider~~

$$\cancel{g(x) = x - \Lambda(x) f(x)}$$

~~$\Lambda(x) \in \mathbb{R}^{n \times n}$ but a function of $x \in \mathbb{R}^n$.~~

~~Choose it so that $J_g(x_*) = 0$~~

Instead, let us consider

$$g(x) = x - J_f(x)^{-1} f(x)$$

Would like $J_g(x_*) = 0$.

Letting $\delta_{ii} = 1$, $\delta_{ij} = 0$ if $i \neq j$,
and $K(x) := J_f(x)^{-1}$

$$\frac{\partial g_i(x)}{\partial x_j} = \frac{\partial x_i}{\partial x_j} - \frac{\partial\left([K(x) f(x)]_i\right)}{\partial x_j}$$

$$= \delta_{ij} - \sum_{r=1}^{n} \frac{\partial\{[K(x)]_{ir} f_r(x)\}}{\partial x_j}$$

$$= \delta_{ij} - \sum_{r=1}^{n} \frac{\partial [K(x)]_{ir}}{\partial x_j} f_r(x)$$

$$- \sum_{r=1}^{n} [K(x)]_{ir} [J_f(x)]_{rj}$$

$$\Longrightarrow \frac{\partial g_i(x_*)}{\partial x_j} = \delta_{ij} - \underbrace{\sum_{r=1}^{n} [K(x_*)]_{ir} [J_f(x_*)]_{rj}}_{\text{ith row of } K(x_*) \times \text{jth } \cancel{\text{row}}\text{ column of } J_f(x_*) \,=\, \delta_{ij}}$$

$$= 0$$

for $j, i = 1, \ldots, n$. Hence, $J_g(x_*) = 0$.

$$\boxed{\begin{array}{l}\text{Newton sequence} \\ \quad \{x^{(k)}\} \text{ such that} \\ \qquad x^{(k+1)} = x^{(k)} - J_f(x^{(k)})^{-1} f(x^{(k)}) \\ \quad \text{for a given } x^{(0)} \in \mathbb{R}^n \end{array}}$$

In practice, we proceed from $x^{(k)}$ to $x^{(k+1)}$ as follows:

(1) Solve the linear system

$$J_f(x^{(k)})\, p^{(k)} = -f(x^{(k)})$$

for $p^{(k)}$. (That is $p^{(k)} = -J_f(x^{(k)})^{-1} f(x^{(k)})$)

(2) $x^{(k+1)} = x^{(k)} + p^{(k)}$

<u>Ex</u>

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \quad f(x) = \begin{bmatrix} x_1^2 + 3x_1x_2 + x_2^2 - 29/400 \\ x_1^2 - x_1x_2 + x_2^2 - 13/400 \end{bmatrix}$$

$x^{(k)} = (1,0)$

$$J_f(x) = \begin{bmatrix} 2x_1 + 3x_2 & 3x_1 + 2x_2 \\ 2x_1 - x_2 & 2x_2 - x_1 \end{bmatrix}$$

$$J_f(1,0) = \begin{bmatrix} 2 & 3 \\ 2 & -1 \end{bmatrix} \qquad f(1,0) = \begin{bmatrix} 371/400 \\ 387/400 \end{bmatrix}$$

$x^{(k+1)} = (1,0) + p^{(k)}$

where $p^{(k)}$ is the solution of

$$\begin{bmatrix} 2 & 3 \\ 2 & -1 \end{bmatrix} p^{(k)} = \begin{bmatrix} -371/400 \\ -387/400 \end{bmatrix}$$

THM

Suppose $f: \mathbb{R}^n \to \mathbb{R}^n$ is such that $\partial^2 f_i / \{\partial x_j \partial x_k\}$ are continuous for $k, i, j = 1, \dots, n$ and all x in an open neighborhood $N(x_*)$ of x_*, a root of f. Suppose also $J_f(x_*)$ is invertible.

(i) There exists $\overline{B}(x_*, \varepsilon) \subset N(x_*)$ such that Newton sequence $\{x^{(k)}\}$ converges to x_* for all $x^{(0)} \in \overline{B}(x_*, \varepsilon)$.

(ii) The ~~rate~~ order of convergence of $\{x^{(k)}\}$ to x_* is at least quadratic.

Proof

(i) As $J_f(x_*)$ is invertible, and due to continuity of $\partial f_i / \partial x_j$, the Jacobian $J_f(x)$ remains invertible in an open neighborhood $\tilde{N}(x_*) \subset N(x_*)$ of x_*. Indeed, as $\partial^2 f_i / \partial x_j \partial x_k$ are also continuous, the entries of $J_f(x)^{-1}$ have continuous first derivatives in $\tilde{N}(x_*)$.

The entries of the Jacobian of $g(x) = x - J_f(x)^{-1} f(x)$ also have continuous first derivatives in $\tilde{N}(x_*)$ with $J_g(x_*) = 0$. Hence, the result follows (see the thm on page 2 of $\boxed{\text{CH4 - P2}}$), ~~where~~ for some $\overline{B}(x_*, \varepsilon) \subset \tilde{N}(x_*)$.

(ii) Let $M := \max \{ \| J_f(x)^{-1} \|_\infty \mid x \in \overline{B}(x_*, \varepsilon) \}$, γ be such that

$$\| J_f(x) - J_f(y) \|_\infty \leq \gamma \| x - y \|_\infty \quad \forall x, y \in B(x_*, \varepsilon)$$

(The existence of such a γ is due to continuity of $\frac{\partial^2 f_i}{\partial x_j \partial x_k}$.)

By Taylor's thm with integral remainder,

$$0 = f(x_*) = f(x^{(k)}) + \int_0^1 J_f(x^{(k)} + t(x_* - x^{(k)}))(x_* - x^{(k)})\,dt$$

$$\Longrightarrow 0 = \underbrace{J_f(x^{(k)})^{-1} f(x^{(k)})}_{x^{(k)} - x^{(k+1)}} + (x_* - x^{(k)}) + J_f(x^{(k)})^{-1} \int_0^1 \{ J_f(x^{(k)} + t(x_* - x^{(k)})) - J_f(x^{(k)}) \} (x_* - x^{(k)})\,dt$$

$$\Longrightarrow x^{(k+1)} - x_* = J_f(x^{(k)})^{-1} \int_0^1 \{ J_f(x^{(k)} + t(x_* - x^{(k)})) - J_f(x^{(k)}) \} (x_* - x^{(k)})\,dt$$

where we assume $x^{(k)} \in \bar{B}(x_*, \epsilon)$. Indeed, we can assume

(*) $\varepsilon \leqslant \sqrt{2\varepsilon/\gamma M}$ as we can choose ε as small as we wish.

Taking the norms of both sides and employing the submultiplicative property

$$(**)\ \|x^{(k+1)} - x_*\|_\infty \leqslant \|J_f(x^{(k)})^{-1}\|_\infty \int_0^1 \underbrace{\| J_f(x^{(k)} + t(x_* - x^{(k)})) - J_f(x^{(k)}) \|_\infty}_{\leqslant t\gamma \|x^{(k)} - x_*\|_\infty^2} \|x^{(k)} - x_*\|_\infty \,dt$$

$$\leqslant \frac{M\gamma}{2} \|x^{(k)} - x_*\|_\infty^2$$

Now

$$\|x^{(k)} - x_*\|_\infty \leqslant \varepsilon \Longrightarrow \|x^{(k+1)} - x_*\|_\infty \leqslant \frac{M\gamma}{2}\varepsilon^2 \overset{\text{due to } (*)}{\leqslant} \varepsilon .$$

Hence, given $x^{(0)} \in \bar{B}(x_*, \varepsilon)$, we have $x^{(k)} \in \bar{B}(x_*, \varepsilon)$ for all k.

Assume also $\varepsilon \leqslant 1/\gamma M$. Let $\varepsilon^{(k)} := \frac{1}{M}\left(\frac{1}{2}\right)^{2^{-k}}$, $M := \frac{\gamma M}{2}$.

Then

$$\|x^{(k)} - x_*\|_\infty \leqslant \varepsilon^{(k)} \overset{\text{using } (**)}{\Longrightarrow} \|x^{(k+1)} - x_*\|_\infty \leqslant M\left\{\frac{1}{M}\left(\frac{1}{2}\right)^{2^{-k}}\right\}^2$$

$$= \frac{1}{M}\left(\frac{1}{2}\right)^{2^{-(k+1)}} = \varepsilon^{(k+1)}$$

as well as $\|x^{(0)} - x_*\|_\infty \leqslant \varepsilon \leqslant \frac{1}{\gamma M} = \frac{1}{2M} = \varepsilon^{(0)}$. By induction $\|x^{(k)} - x_*\|_\infty \leqslant \varepsilon^{(k)}$ $\forall k$, and

$$\lim_{k \to \infty} \frac{\varepsilon^{(k+1)}}{\{\varepsilon^{(k)}\}^2} = M.$$

This yields an order of convergence at least two. $\square$

Variant of Newton's Method that usually Converges Globally

$\{x^{(k)}\}$ – Newton sequence

Would like

$$\boxed{\text{(GB)} \quad \lim_{k\to\infty} x^{(k)} = x_* \ \text{ s.t. } f(x_*) = 0 \ \text{ for all } x^{(0)} \in \mathbb{R}^n.}$$

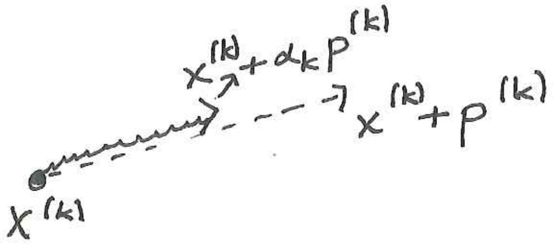
Line search techniques

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

$p^{(k)}$ s.t. $J_f(x^{(k)})\, p^{(k)} = -f(x^{(k)})$

$\alpha_k \in (0, 1]$ step-length

How shall we choose α_k?



Define $r(\alpha) := \|f(x^{(k)} + \alpha p^{(k)})\|_2 \quad (r : \mathbb{R} \to \mathbb{R}^*)$

$$r'(\alpha) = \frac{1}{2\,\|f(x^{(k)} + \alpha p^{(k)})\|_2}\, 2 f(x^{(k)} + \alpha p^{(k)})^T J_f(x^{(k)} + \alpha p^{(k)})\, p^{(k)}$$

$$r'(0) = \frac{f(x^{(k)})^T \left(J_f(x^{(k)})\, p^{(k)} \right)}{\| f(x^{(k)}) \|_2}$$

$-f(x^{(k)})$

$$= \frac{-\| f(x^{(k)}) \|_2^2}{\| f(x^{(k)}) \|_2} = -\| f(x^{(k)}) \|_2 < 0$$

Hence,

$r(\alpha) < r(0)$ for $\alpha > 0$ small enough

But would like more than a mere reduction, would like a sufficient reduction that guarantees (GB) usually.

Linear approximation of $f: \mathbb{R}^n \to \mathbb{R}^n$ at $x^{(k)}$

$$l(x) := \underbrace{f(x^{(k)})}_{n\times 1} + \underbrace{J_f(x^{(k)})}_{n\times n} \underbrace{(x - x^{(k)})}_{n\times 1}$$

Ex

$f: \mathbb{R}^2 \to \mathbb{R}^2$, $f(x) = \begin{bmatrix} x_1^2 - 5x_1x_2 + 2x_2^2 - 3 \\ 3x_1^2 + x_1x_2 + x_2^2 + 1 \end{bmatrix}$

linear approximation at $(0,1)$

$J_f(x) = \begin{bmatrix} 2x_1 - 5x_2 & -5x_1 + 4x_2 \\ 6x_1 + x_2 & x_1 + 2x_2 \end{bmatrix}$ $J_f(0,1) = \begin{bmatrix} -5 & 4 \\ 1 & 2 \end{bmatrix}$

$f(0,1) = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

$$l(x) = f(0,1) + J_f(0,1)\left(\begin{bmatrix} x_1 \\ x_2 - 1 \end{bmatrix} \right)$$

$$= \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} -5 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 - 1 \end{bmatrix} = \begin{bmatrix} -5x_1 + 4x_2 - 5 \\ x_1 + 2x_2 \end{bmatrix}$$

Newton sequence in terms of linear approximations

$$x^{(k+1)} \text{ is s.t. } \ell(x^{(k+1)}) = 0$$

$$\text{where } \ell(x) = f(x^{(k)}) + J_f(x^{(k)})(x - x^{(k)}).$$

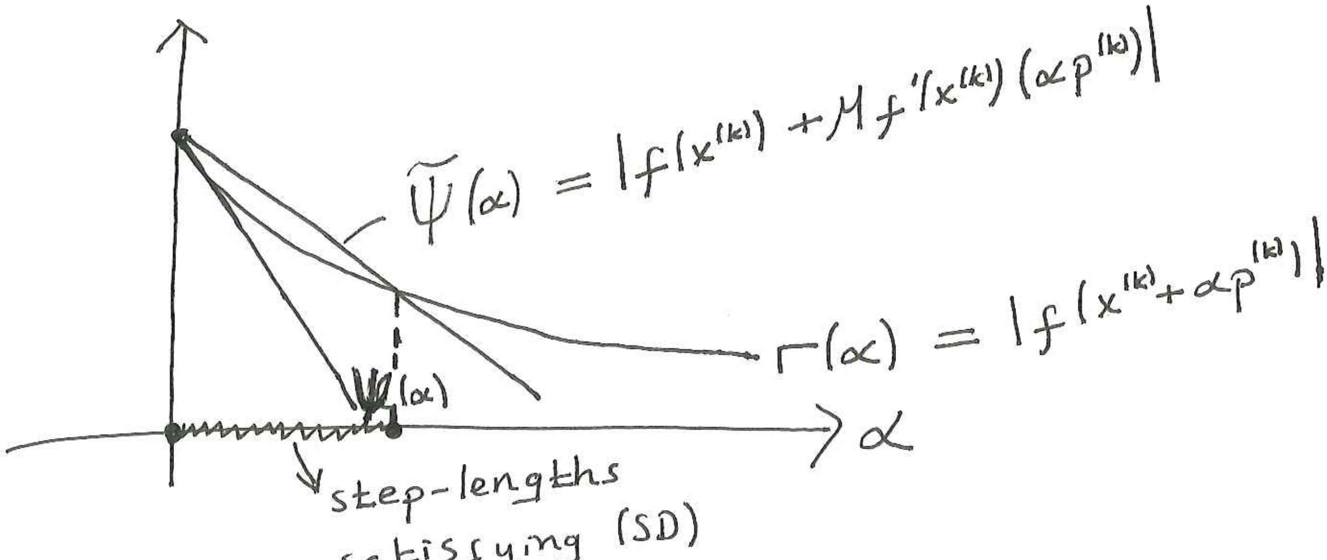
Sufficient reduction condition

$$\text{(SD)} \quad \begin{cases} -r(\alpha) + r(0) \\ \geqslant \\ \mu(-\psi(\alpha) + \psi(0)) \end{cases}$$

for a given $\mu \in (0,1)$, and

$$\psi(\alpha) := \| \ell(x^{(k)} + \alpha p^{(k)}) \|_2$$

Illustration in the univariate case



Armijo backtracking line search
(given $x^{(k)}$ and $p^{(k)}$)

$\alpha_k \leftarrow 1$

while $\{(r(0) - r(\alpha)) < \mu(\psi(0) - \psi(\alpha))\}$

$\qquad \alpha_k \leftarrow \alpha_k / 2$

end

return α_k

Remark

Guaranteed to terminate, because

$$\lim_{\alpha \to 0^+} \frac{r(0) - r(\alpha)}{\psi(0) - \psi(\alpha)} = 1$$

implying

$$\frac{r(0) - r(\alpha)}{\psi(0) - \psi(\alpha)} \geq \mu$$

for all $\alpha > 0$ sufficiently close to 0.

Simplification of (SD)

$$\|f(x^{(k)})\|_2 - \|f(x^{(k)} + \alpha p^{(k)})\|_2$$

$$\geq \mu \{ \|f(x^{(k)})\|_2 - \|f(x^{(k)}) + \underbrace{J_f(x^{(k)}) p^{(k)}}_{-f(x^{(k)})} \alpha \|_2 \}$$

$$= \alpha \mu \|f(x^{(k)})\|_2$$

$$\iff \underbrace{(1 - \alpha\mu)}_{\in (0,1)} \|f(x^{(k)})\|_2 \geq \|f(x^{(k)} + \alpha p^{(k)})\|_2$$

Suppose α_k are chosen to satisfy (SD) and $\alpha_k > \beta$ $\forall k$ for some $\beta > 0$.

$$\| f(x^{(k+1)}) \|_2 \leqslant \underbrace{(1-\alpha_k \mu)}_{\leqslant (1-\beta\mu) < 1} \| f(x^{(k)}) \|_2$$

$$\leqslant (1-\beta\mu)^{k+1} \| f(x^{(0)}) \|_2$$

so $\lim_{k\to\infty} \| f(x^{(k)}) \|_2 = 0.$