Newton's Method

Let X and Y be Banach spaces. Let $P: X \to Y$ be a C^1 map. Newton's algorithm

$$x_{n+1} = x_n - [DP(x_n)]^{-1} P(x_n)$$

is simply the successive approximation algorithm for the fixed point problem

$$x = T(x)$$

where $T(x) = x - [DP(x)]^{-1} P(x)$.

Clearly x such that $P(x) = 0$ is a fixed-point of T. We aim to show that under suitable hypotheses Newton's method converges to a solution of the equation $P(x) = 0$, i.e. it finds a 'root' of P.

Why do we care?

Consider the problem of optimizing

$$\phi : X \to \mathbb{R}$$

for a functional ϕ on a Banach space X.

The first order necessary condition for an extremum of ϕ is (assuming ϕ is C^1),

$$P(x) \triangleq (D\phi)(x) = 0$$

$$\longleftrightarrow \quad (D\phi)(x)\, h = 0 \quad \forall \; h \in X .$$

Here $P(x)$ is the Fréchet derivative at x, and hence it is a bounded linear functional $P(x) : X \to \mathbb{R}$. Denote $X^* =$ space of all bounded linear functionals on X, with $\|\cdot\|$ given by $\|f\| = \sup_{\|x\| \leq 1} |f(x)|$.

X^* is a Banach space.

Finding an extremal, i.e. a solution to $P(x) = 0$, is a root-finding problem for $P : X \to Y = X^*$. Newton's algorithm can be used here.

Set up in this way, the algorithm is applicable to a wide-variety of problems including problems in the calculus of variations.

Note: ϕ is C^k if $(D^k\phi)(x)$, the k^{th} Fréchet derivative exists and is continuous in x.

Two background results

Two results related to meanvalue theorems are needed. We quote:

<u>Proposition 1</u> (mean value inequality)

Let $T : D \subset X \to Y$ be a map defined on an open set D of a normed space X, with values in normed space Y. Assume T is Fréchet differentiable on D. Let $x \in D$ and suppose that $x + \alpha h \in D$ for all $\alpha \in [0,1]$.

Then

$$\| T(x+h) - T(x) \| \leq \left(\sup_{0<\alpha<1} \| T'(x+\alpha h) \| \right) \cdot \|h\|$$

(Here T' denotes the Fréchet derivative of T.)

<u>Proposition 2</u>

Let T be twice Fréchet differentiable on $D \subset X$. Let $x \in D$ and suppose that $x + \alpha h \in D$ for all $x \in [0,1]$. Then

$$\| T(x+\alpha h) - T(x) - T'(x)h \| \leq \frac{1}{2} \|h\|^2 \sup_{0<\alpha<1} \| T''(x+\alpha h) \|.$$

For a proof of Proposition 1 see separate <u>notes on MVT</u>. The proof of Proposition 2 follows the same pattern as that of Proposition 1.

If in proposition 1, $\| T'(x) \| \le \alpha < 1 \quad \forall x \in D$ and D is such that $x + \alpha h \in D$ whenever $x, h \in D$ and $0 < \alpha < 1$, then T is a <u>contraction</u>. If further, $T: S \to S$, S closed $S \subset D$ $D \overset{open}{\subset} X$ a Banach space, then the sequence $\{x_n\}$ defined by $x_{n+1} = T(x_n)$ converges to a unique fixed point of T. (See notes on <u>contraction mappings and fixed points</u>)

Thus the strategy for proving convergence of Newton's method <u>might</u> be check a (global) contraction property for $T(x) = x - [DP(x)]^{-1} P(x)$. We need a little more as in Theorem below. (Why? Because we need to make sure $[DP(x)]^{-1}$ is well defined)

<u>A Formula</u>

It is easy to use Leibnitz rule for Fréchet derivatives to show that, given

$$T(x) = x - [DP(x)]^{-1} P(x)$$

$$= x - [P'(x)]^{-1} P(x),$$

$$DT(x) = T'(x) = [P'(x)]^{-1} [P''(x)] [P'(x)]^{-1} P(x)$$

Theorem

Let $P: X \to Y$ be a C^2 map on Banach space X with values in Banach space Y. Assume further that:

1. $\| P''(x) \| \leq K$

2. There is a point $x_1 \in X$ such that $p_1 = P'(x_1)$ has a bounded inverse p_1^{-1} with $\| p_1^{-1} \| \leq \beta_1$ and

$$\| p_1^{-1} \cdot P(x_1) \| \leq \eta_1 .$$

3. The constant $h_1 = \beta_1 \eta_1 K$ satisfies

$$h_1 < 1/2 .$$

Then the sequence (where $p_n = P'(x_n)$)

$$x_{n+1} = x_n - p_n^{-1} P(x_n)$$

~~converges for n≥1~~ is well defined for all $n \geq 1$ and converges to a solution of $P(x) = 0$.

ASIDE

Lemma $H = \mathbb{1} - A \qquad \|A\| \leq h_1 < 1$

$$H^{-1} = (\mathbb{1} - A)^{-1} \quad \text{exists}$$

$$= \mathbb{1} + A + A^2 + \cdots$$

$$\|H^{-1}\| \leq \frac{1}{1-h_1}$$

proof:

$$S_n = \mathbb{1} + A + A^2 + \cdots + A^n$$

$$HS_n = (\mathbb{1} - A)(\mathbb{1} + A + \cdots + A^n)$$

$$= \mathbb{1} - A + A - A^2 + A^2 + \cdots - A^{n-1} + A^{n-1} - A^n + A^n - A^{n+1}$$

$$= \mathbb{1} - A^{n+1}$$

$$\|HS_n - \mathbb{1}\| = \|A^{n+1}\| < h_1^{n+1} \to 0 \quad \text{as } n \to \infty$$

$k > 0$

$$\|S_{n+k} - S_n\| \leq \|A\|^n \, \|A^k - \mathbb{1}\|$$

$$\Rightarrow 0 \quad \text{as } n \to \infty$$

$\Rightarrow$ $\{S_n\}$ is a Cauchy sequence

$\Rightarrow$ it is convergent, since $B(X,Y)$ is Banach <u>whenever Y is Banach</u>.

$$S = \lim_{n\to\infty} S_n$$

$$~~HS = \lim_{n\to\infty} HS_n = H \lim_{n\to\infty} S_n~~$$

$$HS = H \lim_{n\to\infty} S_n = \lim_{n\to\infty} HS_n = \mathbb{1}.$$

$$\|S\| = \|\mathbb{1} + A + A^2 + \cdots\| \leq \sum_{n=1}^{\infty} \|A\|^n = \frac{1}{1-\|A\|} < \frac{1}{1-h_1}$$

$\|A\| < h_1 \quad -\|A\| > -h_1$

$1 - \|A\| > 1 - h_1 ; \quad \frac{1}{1-\|A\|} < \frac{1}{1-h_1}$

Proof:

First we show that x_2 is well defined and has associated constants β_2, η_2, h_2. This means, by induction, all x_n are well-defined.

$$x_2 = x_1 - p_1^{-1} P(x_1)$$

is clearly well-defined by hypothesis 2.
Also $\| x_2 - x_1 \| \leq \eta_1$ by hypothesis 2.

Mean value inequality $\Rightarrow$

$$\| p_1^{-1} [p_1 - p_2] \| \leq \beta_1 \sup_{\bar{x}} \| P''(\bar{x}) \| \cdot \| x_2 - x_1 \|$$

where $\bar{x} = x_1 + \alpha(x_2 - x_1)$, $0 \leq \alpha \leq 1$.

Thus $\| p_1^{-1} [p_1 - p_2] \| \leq \beta_1 K \eta_1 = h_1$.

Since $h_1 < \frac{1}{2}$ (hypothesis 3), it follows that

as a linear operator $: X \to X$

$$H = \mathbb{1} - p_1^{-1} [p_1 - p_2]$$

$$= p_1^{-1} p_2$$

has a bounded inverse with $\| H^{-1} \| \leq \dfrac{1}{1 - h_1}$.

$$p_1 H = p_2 \; ; \quad (p_1 H)^{-1} = H^{-1} p_1^{-1}, \text{ so}$$

p_2^{-1} exists.

$$\| p_2 \|^{-1} \leq \| H^{-1} \| \cdot \| p_1^{-1} \| \leq \frac{\beta_1}{1 - h_1} = \beta_2 .$$

Let $T_1(x) = x - p_1^{-1} P(x)$.

Clearly $T_1(x_1) = x_2$ and $T_1'(x_1) = 0$.

Thus,
$$p_1^{-1} P(x_2) = T_1(x_1) - T_1(x_2)$$
$$= -\{T_1(x_2) - T_1(x_1)\}$$
$$= -\{T_1(x_2) - T_1(x_1) - T_1'(x_1)(x_2 - x_1)\}$$

By Mean Value Inequality (<u>Proposition 2</u>, page 3 of these notes),
$$\| p_1^{-1} P(x_2) \| \le \tfrac{1}{2} \sup \| T_1''(x) \| \cdot \| x_2 - x_1 \|^2$$
$$= \tfrac{1}{2} \sup \| p_1^{-1} P''(x) \| \cdot \| x_2 - x_1 \|^2$$
$$\le \tfrac{1}{2} \beta_1 K \eta_1^2 = \tfrac{1}{2} h_1 \eta_1 .$$

Then
$$\| p_2^{-1} P(x_2) \| = \| H^{-1} p_1^{-1} P(x_2) \|$$
$$< \frac{1}{2} \frac{h_1 \eta_1}{1 - h_1} = \eta_2 < \frac{1}{2} \eta_1$$

(<u>aside</u>: $0 < h_1 < 1/2 \Rightarrow -h_1 > -\frac{1}{2} \Rightarrow 1 - h_1 > 1 - 1/2 = 1/2$
$\Rightarrow \frac{1}{1-h_1} < 2 \Rightarrow \frac{h_1}{1-h_1} < \frac{1}{2} \cdot 2 = 1$)

Finally, set $h_2 = \beta_2 \eta_2 K$ to get,

$$h_2 \leq \frac{\beta_1}{(1-h_1)} \cdot \frac{1}{2} \frac{h_1 \eta_1}{(1-h_1)} K$$

$$= \frac{1}{2} \frac{h_1^2}{(1-h_1)^2}$$

$$< \frac{1}{2} \qquad \text{since } h_1 \leq \frac{1}{2} .$$

Hence hypotheses 1, 2, 3 hold at x_2 with constants β_2, η_2, and h_2. Newton's process defines x_n for all n.

Since $\eta_{n+1} < \frac{1}{2} \eta_n$

it follows that

$$\eta_{n+1} < \left(\frac{1}{2}\right)^{n-1} \eta_1 \qquad \text{(goes to zero geometrically)}$$

Also, since $\|x_{n+1} - x_n\| < \eta_n$ it follows that

$$\|x_{n+k} - x_n\| < 2\eta_n .$$

Hence the sequence $\{x_n\}$ is Cauchy. X is Banach $\Rightarrow$ $\{x_n\}$ converges to $x^* \in X$.

The sequence $\|P_n\|$ is bounded since

$$\|P_n\| \leq \|P_1\| + \|P_n - P_1\| \leq \|P_1\| + K \|x_n - x_1\|$$

(by Proposition 1)

and the sequence $\| x_n - x_1 \|$ is bounded since it is convergent.

For each n,

$$p_n \cdot (x_{n+1} - x_n) + P(x_n) = 0$$

Since $\| x_{n+1} - x_n \| \to 0$ as $n \to \infty$ and $\| p_n \|$ is bounded, it follows that $\| P(x_n) \| \to 0$ as $n \to \infty$.

By continuity $P(x^*) = 0$. ▨

Remark.

Suppose Newton's method converges to x^* such that $P'(x^*)^{-1}$ exists. Further, suppose that within an open region R containing x^* and the sequence $\{x_n\}$, the quantities $\| P'(x^*)^{-1} \|$, $\| P''(x^*) \|$ and $\| P'''(x) \|$ are bounded.

Then, from

$$x_{n+1} - x^* = T(x_n) - T(x^*)$$

and from

$$T'(x^*) = 0,$$

we get, $\| x_{n+1} - x^* \| \leq \frac{1}{2} \sup_{\bar{x}} \| T''(\bar{x}) \| \cdot \| x_n - x_0 \|^2$

where $\bar{x} = x_n + \alpha (x_n - x^*)$, $\quad 0 \le \alpha \le 1$.

Hence,

$$\| x_{n+1} - x^* \| \le c \, \| x_n - x^* \|^2$$

where,

$$c = \frac{1}{2} \sup_{x \in \mathbb{R}} \| T''(x) \|$$

a bound depending on $\| P'''(x) \|$.

This is quadratic convergence.