X. Linearization and Newton's Method

**** linearization ****

X, Y nls's, $f : G \subseteq X \to Y$. Given $y \in Y$, find $z \in G$ s.t. $fz = y$. Since there is no assumption about f being linear, we might as well assume that $y = 0$.

Since the only equations we can solve numerically are *linear* equations, the solution of the equation $fz = 0$ is found by solving (the first few in) a sequence of linear equations. The typical step is this: With z_0 a guess for z, pick a linear map $A = A_{z_0}$ so that

$$fx \sim fz_0 + A(x - z_0) \qquad \text{for } x \sim z_0$$

and solve the *linear* equation

$$fz_0 + A(x - z_0) = 0$$

instead. Its solution, z_1, may be closer to z than z_0 is, and further improvement is possible by repetition of this process. This leads to the *iteration*

$$z_{n+1} = Tz_n, \qquad n = 0, 1, 2, \ldots$$

with T the (usually nonlinear) map given by the rule

$$Tx := x - (A_x)^{-1} fx.$$

The choice of A_x for given x is, of course, crucial for the convergence of the sequence (z_n) of iterates to z.

There is, in effect, only one technique for proving such convergence, and that is by contraction, i.e., by showing that T is a proper contraction on some nbhd of z (see (II.21)). We'll discuss variants of that argument below.

**** differentiation ****

The best known scheme and model for all others is to choose A_u in such a way that **the** affine **function**

$$x \mapsto fu + A_u(x - u)$$

touches f **at** u, i.e., so that

$$(1) \qquad \|fx - \big(fu + A_u(x - u)\big)\| = o(\|x - u\|).$$

Here, x is meant to vary over some open nbhd of u. Note that, if also the affine function $fu + C(\cdot - u)$ touches f at u, then

$$\|(C - A_u)(x - u)\| = o(\|x - u\|),$$

hence $\|C - A_u\| = 0$. This shows that A_u is uniquely defined by the touching condition (1). There is, of course, no guarantee that such a linear map A_u exists. But, if $fu + C(\cdot - u)$ touches f at u for some $C \in bL(X, Y)$, then we write

$$C = Df(u)$$

and call this map the **(Fréchet-)derivative of** f **at** u.

**** examples ****

If f is a bounded affine map, i.e., $f : x \mapsto y + Cx$ for some $C \in bL(X,Y)$, then $Df(u) = C$ for all $u \in X$.

If $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, then $Df(u) \in \mathbb{R}^{m\times n}$, i.e., $Df(u)$ is a matrix, called the **Jacobian** of f at u. If $f \in C^{(1)}(G, \mathbb{R}^m)$ for some open domain G, then $Df(u)$ exists for all $u \in G$ and depends continuously on u there.

In particular, if $Y = \mathbb{R}$, i.e., if f is a real-valued function of n variables, then $Df(u)$ (if it exists) is a linear functional, called the **gradient** of f at u and often denoted by $\nabla f(u)$. If $h \in X$ and $g : \mathbb{R} \to \mathbb{R} : t \mapsto f(u + th)$, then $g'(0) = \nabla f(u)h$.

More generally, if $Df(u)$ exists, then

$$(f(u+th) - fu)/t = \underbrace{(f(u+th) - fu)/t - Df(u)h}_{\|h\|o(t)/t} + Df(u)h,$$

hence

$$g'(0) = Df(u)h,$$

with

$$g : \mathbb{R} \to Y : t \mapsto f(u+th).$$

But $g'(0)$ may well exist even though $Df(u)$ does not. This leads to the weaker notion of the **directional** (or, **Gateaux**) derivative

$$D_h f(u) := \lim_{t\to 0+} (f(u+th) - fu)/t$$

and this equals $g'(0+)$. f is **Gateaux-differentiable at** u if $D_h f(u)$ exists for all $h \in X$. In any case, $h \mapsto D_h f(u)$ is positive homogeneous, and $f \mapsto D_h f(u)$ is linear.

If $Df(u)$ exists, then, as just remarked, $D_h f(u) = Df(u)h$. In particular, $h \mapsto D_h f(u)$ is then a bounded *linear* map. This makes it easy to come up with maps f that have all directional derivatives at a point, yet fail to be Fréchet-differentiable there. For *example*, the map $f : X \to \mathbb{R} : x \mapsto \|x\|$ has $D_h f(0) = \|h\|$, all h (since $\big(\|0 + th\| - \|0\|\big)/t = \|h\|$), but the resulting map $h \mapsto D_h f(0) = \|h\|$ obviously is not linear. On the other hand, if $h \mapsto D_h f(u)$ is a bounded linear map, then it provides the only possible candidate for the Fréchet-derivative, and so assists in the latter's construction.

For *example*, consider the map

$$f : C^{(m)}[a\,.\,.\,b] \to C[a\,.\,.\,b] : x \mapsto \big(t \mapsto F(t, x(t), \ldots, (D^m x)(t))\big)$$

with $F \in C^{(1)}(\mathbb{R}^{m+2})$. Then

$$\begin{aligned}\frac{f(u+sh) - fu}{s}(t) &= \frac{F\big(t, u(t) + sh(t), \ldots, D^m u(t) + sD^m h(t)\big) - F\big(t, u(t), \ldots, D^m u(t)\big)}{s}\\ &= \big(sh(t)D_2F + \cdots + sD^m h(t)D_{m+2}F + O(s\|h\|\omega_{DF}(s\|h\|))\big)/s\\ &= \sum_{j=2}^{m+2} D_jF\big(t, u(t), \ldots, D^m u(t)\big)D^{j-2}h(t) + o(s\|h\|^2).\end{aligned}$$

Hence, $(f(u+sh)-fu)/s$ approaches

$$D_h f(u) = \sum_{j=2}^{m+2} D_j F(\cdot, u(\cdot), \ldots, D^m u(\cdot)) D^{j-2} h,$$

as $s \to 0$, and this convergence is uniform in $\|h\|$. Also, $D_h f(u)$ is linear in h, and bounded with respect to $\|h\|$. This implies that

$$Df(u) = \sum_{j=2}^{m+2} D_j F(\cdot, u(\cdot), \ldots, D^m u(\cdot)) D^{j-2},$$

a linear m-th order OD operator.

**** basic rules for Fréchet and Gateaux derivative ****

The Fréchet-derivative shares all the basic properties of a derivative familiar from elementary Calculus. In particular, $Df(u)$ is linear in f and satisfies the **chain rule**:

$$D(gf)(u) = Dg(fu)Df(u).$$

Further, if $Df(u)$ exists, then f is continuous at u, since

$$\|fx - fu\| \le \underbrace{\|fx - fu - Df(u)(x-u)\|}_{o(\|x-u\|)} + \|Df(u)(x-u)\| = O(\|x-u\|).$$

This shows that f is even Lipschitz continuous, with (local) Lipschitz constant $\sim \|Df(u)\|$.

H.P.(1) Prove: *(i) (If f is Gateaux-differentiable at u, then) $h \mapsto D_h f(u)$ is positive homogeneous. (ii) $f \mapsto D_h f(u)$ is linear (as a map on the linear space of all maps on some nls X into the same nls Y and Gateaux-differentiable at u). (iii) chainrule: (If f is Fréchet-differentiable at u and g is Fréchet-differentiable at $f(u)$, then) $D(g \circ f)(u) = Dg(fu)Df(u)$. (iv) product rule: (If f and g are scalar-valued and Gateaux-differentiable at u and $fg : u \mapsto f(u)g(u)$, then) $D_h(fg)(u) = D_h f(u)g(u) + f(u)D_h g(u)$.*

**** meanvalue estimates ****

On the other hand, already for functions into $\mathbb{R}^2$, we no longer have the customary mean value theorem, i.e., $fy - fx$ usually does not equal $Df(\xi)(y-x)$ no matter how we choose $\xi \in [x \,.\,.\, y]$. For *example*, for $f : \mathbb{R} \to \mathbb{R}^2 : t \mapsto (t^2, t^3)$, we get $Df(t) = [2t\ 3t^2]'$, hence $(1,1) = f1 - f0 \overset{!}{=} Df(t)(1-0)$ would imply the contradictory statements $t = 1/2$ and $t = (1/3)^{1/2}$.

Nevertheless, one obtains even for $f : X \to Y$ with X, Y nls's, the usual

(2) Meanvalue Estimate.

$$\|fx - fy\| \le \sup \|Df([x \,.\,.\, y])\| \|x - y\| \tag{3}$$

with the aid of HB: By (IV.27)HB, one can find $\lambda \in S_{Y^*}$ so that

$$\|fy - fx\| = \lambda(fy - fx) = g(1) - g(0) = Dg(\theta), \quad \text{for some } \theta \in [0 \,.\,.\, 1],$$

with $g : [0\,..\,1] \to \mathbb{R} : t \mapsto \lambda f(x + t(y-x))$, hence

$$Dg(\theta) = \lambda Df\big(x + \theta(y-x)\big)(y-x) \le \|\lambda\| \|Df\big(x+\theta(y-x)\big)\| \|y-x\|,$$

and this proves (3) since $\|\lambda\| = 1$.

If you are comfortable with vector-valued (hence with map-valued) integration, then (3) can be obtained directly by

$$\begin{aligned} fy - fx = \int_0^1 Df(x+t(y-x))\,\mathrm{d}t\;(y-x) &\le \int_0^1 \|Df(x+t(y-x))\|\,\mathrm{d}t\;\|y-x\| \\ &\le \sup \|Df([x\,..\,y])\|\;\|y-x\|. \end{aligned}$$

H.P.(2) Let A be a boundedly invertible lm from the nls X to the nls Y, let K be a convex subset of X, and let $f : K \to Y$ be Fréchet differentiable. Prove that the map $(A - f) : K \to Y : x \mapsto Ax - f(x)$ is 1-1 in case $\sup_{x\in K} \|A^{-1}Df(x)\| < 1$.

We can improve this estimate in case Df has some smoothness, as follows.

(4) Lemma. *If $u \mapsto Df(u)$ is continuous on some convex set N with modulus of continuity ω, i.e.,*

$$\forall\{y, z \in N\} \quad \|Df(y) - Df(z)\| \le \omega(\|y-z\|),$$

then

$$\forall\{x, y \in N\} \quad E_f(x,y) := fy - \big(fx + Df(x)(y-x)\big) \le \int_0^1 \omega(t\|y-x\|)\,\mathrm{d}t\;\|y-x\|.$$

In particular,

$$\|E_f(x,y)\| \le \frac{K}{2}\|y-x\|^2$$

in case Df is Lipschitz continuous on N with constant K.

Proof: Let λ be a lfl of norm 1 that takes on its norm on the vector $E_f(x,y)$, and consider again the function $g : [0\,..\,1] \to Y : t \mapsto \lambda f(x+t(y-x))$. Now

$$\lambda fy - \lambda fx = g(1) - g(0) = \int_0^1 Dg(t)dt = \int_0^1 \lambda Df(x+t(y-x))(y-x)\,\mathrm{d}t,$$

hence

$$\begin{aligned} \|E_f(x,y)\| &= \lambda\big(fy - (fx + Df(x)(y-x)\big) \\ &= \int_0^1 \lambda\{Df\big(x+t(y-x)\big) - Df(x)\}(y-x)\,\mathrm{d}t \\ &\le \int_0^1 \|\lambda\|\omega(t\|y-x\|)\|y-x\|\,\mathrm{d}t = \int_0^1 \omega(t\|y-x\|)\,\mathrm{d}t\;\|y-x\|. \qquad \square \end{aligned}$$

This argument, too, can be simplified if you are willing to use map-valued integration, as follows:

$$\begin{aligned} fy - fx - Df(x)\,(y-x) &= \int_0^1 \big(Df(x+t(y-x)) - Df(x)\big)\,\mathrm{d}t\,(y-x) \\ &\le \int_0^1 \omega_{Df}(t\|y-x\|)\,\mathrm{d}t\;\|y-x\|. \end{aligned}$$

Newton's method

Assume that the map $f : X \to Y$ (for which we seek $z \in X$ s.t. $fz = 0$) is **continuously** Fréchet-differentiable **at** z, i.e., f is Fréchet-differentiable in some nbhd N of z and $\|(Df)(x) - (Df)(z)\| \le \omega_{Df}(\|x - z\|)$ for some modulus of continuity ω_{Df}. Then, for $x \in N$, we compute a (better?) approximation y to z by dropping all higher order terms from the expansion

$$0 = fz = fx + Df(x)(z - x) + \text{ higher order terms},$$

i.e., by solving

$$0 = fx + Df(x)(? - x) \tag{5}$$

thus getting the (improved?) approximation

$$y = x + h = x - Df(x)^{-1} fx.$$

Then

$$y - z = x - z - Df(x)^{-1}\,(fx - fz) = Df(x)^{-1}\Big(Df(x) - \int_0^1 Df(z + (x - z)s)\,\mathrm{d}s\Big)(x - z).$$

But

$$Df(x) - \int_0^1 Df(z + (x - z)s)\,\mathrm{d}s = \int_0^1 \big(Df(x) - Df(z + (x - z)s)\big)\,\mathrm{d}s \le 2\omega_{Df}(\|x - z\|).$$

Hence, altogether,

$$\|y - z\| \le \|Df(x)^{-1}\| 2\omega_{Df}(\|x - z\|)\|x - z\|.$$

This assumes that $Df(x)$ is boundedly invertible, as it would have to be for any sufficiently small neighborhood N' of z since we assume that Df is continuous at z, provided we assume that $Df(z)$ is boundedly invertible. But, in that case, we can choose $N' \subseteq N$ small enough so that also $\sup_{x\in N'} \|Df(x)^{-1}\| =: \|Df(N')^{-1}\| < \infty$. Therefore, for $x \in N'$, the solution y of the linear system (5) satisfies

$$\|y - z\| \le \|Df(N')^{-1}\| 2\omega_{Df}(\|x - z\|)\|x - z\| \ \xrightarrow[x\to z]{}\ 0.$$

This implies the existence of $r > 0$ so that the **Newton map**

$$T : x \mapsto x - Df(x)^{-1} fx$$

carries $B_r(z)$ into itself, and

$$\exists\{q < 1\}\ \forall\{x \in B_r(z)\}\ \ \|z - Tx\| \le q\|z - x\|.$$

Hence, the Newton iteration, started "sufficiently close to" z (i.e., in $B_r(z)$), stays in $B_r(z)$ and converges at least linearly to z.

Note that continuity of $x \mapsto Df(x)$ at z is only used to conclude the uniform existence of $Df(x)^{-1}$ for all x near z. This could have been concluded from the continuity at any nearby point. In other words, continuity of $x \mapsto Df(x)$ at $x = z$ implies that f maps some nbhd of z 1-1 *onto* a nbhd of 0. In fact, the same is then true for any g sufficiently close to f in the sense that

$$\|gx - fx\| + \|Dg(x) - Df(x)\| \ll 1 \quad \forall x \in N'.$$

Under the assumption that f is continuously Fréchet-differentiable at the solution z, the more general iteration function

$$\widetilde{T}x := x - A_x^{-1} fx$$

also generates a sequence converging to z, as long as A_x stays close enough to $Df(z)$. Precisely, with $y := \widetilde{T}x$, we have

$$0 = fx + A_x(y - x)$$

while (with $E_f(x,z) := f(z) - f(x) - Df(x)(z - x)$ as in (4)Lemma)

$$0 = fz = fx + A_x(z - x) + \Big(Df(x) - A_x\Big)(z - x) + E_f(x,z),$$

therefore

$$0 = A_x(z - y) + \Big(Df(x) - A_x\Big)(z - x) + E_f(x,z).$$

Consequently

$$z - y = -A_x^{-1}\big(\Big(Df(x) - A_x\Big)(z - x) + E_f(x,z)\big)$$

or

$$\|z - y\| \le \|A_x^{-1}\|\big(\|Df(x) - A_x\| + \omega_{Df}(\|z - x\|)\big)\|z - x\|.$$

Here, the expression multiplying $\|z - x\|$ can be made small on some nontrivial ball around z by ensuring that that ball is small enough so that $Df(x)$ is close to $Df(z)$, as long as also A_x is chosen close enough to $Df(z)$.

The well-known "quadratic convergence", though, is obtained only if $A_x \xrightarrow[x\to z]{} Df(z)$, i.e., essentially only for Newton's method, and this needs further smoothness assumptions. E.g., for Newton's method, the assumption that $x \mapsto Df(x)$ is Lipschitz continuous in a nbhd of z is sufficient, since the above combined with (4)Lemma gives the following

(6) Proposition. *If $x \mapsto Df(x)$ is Lipschitz continuous in some convex neighborhood N of z, with constant K, then $\|z - Tx\| \le \|Df(N)^{-1}\|(K/2)\|z - x\|^2$.*

In practice, though, it is tough to come up with estimates for ω and/or $\|Df(N)^{-1}\|$ and/or K since they are likely to hold only locally, near the solution, and we don't know the solution in the first place. The real value of the analysis is to demonstrate that Newton's method converges **quadratically**. This is a condition that can be checked for the Newton iterates computed. In fact, it constitutes a very important check. For, the Fréchet derivative is not easy to get right (by hand), and any mistake in the $Df(x)$ is sure to kill the quadratic convergence, leaving you, usually, with linear convergence. Hence, once you detect linear convergence, it is time to check your formulae or program for $Df(x)$.

This leaves open the question of how to get close, i.e., how to obtain a 'sufficiently close' initial guess. In a way, it is reasonable for this to be a problem since there may be many solutions, hence by picking an initial guess we are picking a particular solution.

**** a posteriori error estimates ****

This finishes the standard local convergence theory for Newton's method and its variants. There is an elaborate theory, associated with the name of **Kantorovich**, to allow the conclusion of convergence from numerical evidence computed in the first Newton step. This includes a proof that the given map f has a zero near the initial guess. The idea is a generalization of the well known univariate observation that a continuously differentiable f for which $Tx := x - Df(x)^{-1}fx$ lies close to x must have a zero near x in case f doesn't curve too much, e.g., if Df is Lipschitz continuous with a sufficiently small constant K.

**** infinite-dim. problems also need discretization ****

When the underlying Bs X is infinite-dimensional, then linearization (i.e., Newton's method and its variants) is only half the battle since the linear systems to be solved will in general be infinite-dimensional. **Discretization**, i.e., reduction to an approximate linear problem in finitely many unknowns, needs to be used. Of course, one could also discretize the original problem and thereby obtain right away a finite-dimensional problem, but now that problem is nonlinear in general, hence must be linearized, e.g., by Newton's method. When the discretization is done by projection, then it doesn't matter in which order we do this: The Newton equation

$$Df(x)h = -fx$$

for the correction h to the current guess x, when projected by P becomes

$$P(Df(x)h) = P(-fx)$$

with h to be found in some finite-dimensional F, assumed from now on to be $\operatorname{ran} P$ for simplicity, while the Newton equation for the projected equation $Pfx = 0$ for $x \in F$ is

$$D(Pf)(x)h = -Pfx,$$

with h again sought in F. But, for any bounded lm P, $D(Pf)(x) = PDf(x)$ (as you should *verify*!). It is usually easier, though, to carry out the details by linearizing (symbolically, e.g., using `Maple`) f itself, and then solving the resulting linear problem by projection. A *proof* of convergence of such a double iteration requires some uniformity of f. Typically, the problem of solving $fx = 0$ for x can be rewritten as a fixed point equation

$$x = gx,$$

and, in some nbhd N of the sought-for solution z, g is Fréchet-differentiable, with $Dg(x)$ compact uniformly for $x \in N$. Further,

$$E_g(x, y) := gy - gx - Dg(x)(y - x) \le \omega(\|y - x\|)\|y - x\|$$

for some modulus of continuity ω that depends only on the nbhd N. Finally, $1 - Dg(x)$ should be bounded and bounded below uniformly for $x \in N$. If also $P := P_n \xrightarrow{s} 1$, then, $1 - PDg(x)$ is boundedly invertible for all sufficiently large n, hence the Newton iteration step

$$y = Tx := x - (1 - PDg(x))^{-1}(x - Pgx)$$

can be carried out for any x sufficiently close to z and the resulting approximation y satisfies

$$\|z_P - y\| \le \text{const}\ \omega(\|z_P - x\|)\|z_P - x\|$$

with z_P the unique solution in N of the projected equation $x = Pgx$.

All of this you should (and could by now) verify!

** **example: solving a second-order non-linear ode by collocation** **

Consider the second-order non-linear ode

$$D^2 z = z/2 - 2(Dz)^2 \quad \text{on } [0\,..\,1]; \quad z(0)^2 = 1,\ z(1) = 1.5$$

to be solved for some $z \in X := C^{(2)}[0\,..\,1]$. This means that we are trying to find a zero of the map

(7) $$f : C^{(2)}[0\,..\,1] \to \mathbb{R} \times C[0\,..\,1] \times \mathbb{R} : x \mapsto (x(0)^2 - 1, gx, x(1) - 1.5),$$

with

$$gx := D^2 x + 2(Dx)^2 - x/2.$$

We try to solve this problem by **collocation**. This means that we look for a zero of the (non)linear map

$$\Lambda : x \mapsto (x(0)^2 - 1, (gx)(t_2), \ldots, (gx)(t_{n-1}), x(1) - 1.5) \in \mathbb{R}^n$$

in some n-dimensional lss F of X, hoping that, for an appropriate choice of the collocation points $t_2, \ldots, t_{n-1}$ in $[0\,..\,1]$, Λ is 1-1 on a suitable part of some such F.

From the earlier example, we read off that the Fréchet derivative of g is the linear map

$$Dg(x) : h \mapsto D^2 h + 4(Dx)Dh - h/2,$$

while $x \mapsto x(t)$ is linear, hence its own Fréchet derivative. Therefore (by the chain rule),

$$D\Lambda(x) : h \mapsto (2x(0)h(0), \ldots, (Dg(x)h)(t_j), \ldots, h(1)).$$

Thus, with x our current guess for the solution of $\Lambda? = 0$ in F, Newton's method would provide the improved(?) guess $y := x + h$, with $h \in F$ solving the *linear* problem

(8) $$(D\Lambda)(x)h = -\Lambda x.$$

Now note that, in this derivation, we made no use of the fact that we are seeking a solution in F, nor did we pay particular attention to the collocation points. In fact, for the map f (see (7)) for which we are trying to find a zero, we have

$$(Df)(x) : h \mapsto (2x(0)h(0), Dg(x)h, h(1)).$$

This means that, with x our current guess for the solution z of $f? = 0$, Newton's method would provide the improved(?) guess $y := x + h$, with h solving the *linear* problem

$$(Df)(x)h = -fx,$$

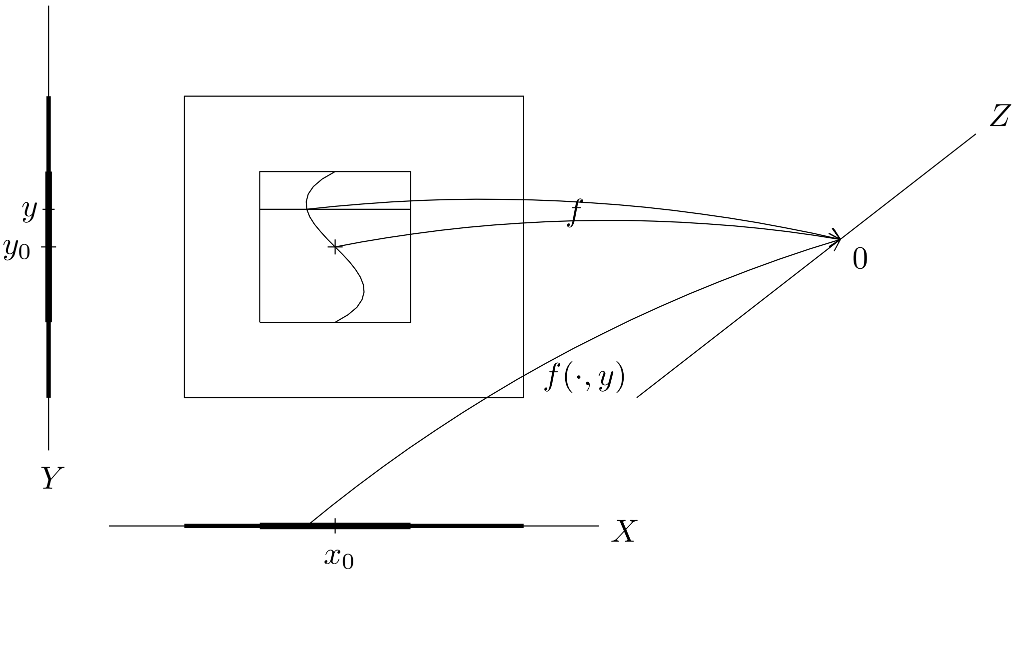
i.e., the *linear* second order ordinary boundary value problem

$$D^2h + 4(Dx)Dh - h/2 = -gx \quad \text{on } [0\,.\,.\,1], \quad 2x(0)h(0) = 1 - x(0)^2,\ h(1) = 1.5 - x(1).$$

If we now try to solve this ode problem by collocation at the points $t_2, \ldots, t_{n-1}$ with $h \in F$, we are back at (8), provided our current guess x is also in F.

**** implicit function theorem ****

There is no time to go into these theories. Instead, I bring quickly an important application of the contraction map idea and Newton's method, viz. the



(9) Figure. The Implicit Function Theorem

(10) Implicit Function Theorem. *X, Y, Z Bs's, $f : X \times Y \to Z$, $f(x_0, y_0) = 0$, f continuous on $N := B_r(x_0) \times B_s(y_0)$ for some $r, s > 0$. Further, $\forall \{y \in B_s(y_0)\}$ $f(\cdot, y)$ is Fréchet-differentiable on $B_r(x_0)$, and the resulting map $(x, y) \mapsto Df(\cdot, y)(x)$ is continuous at (x_0, y_0). Also, $A := Df(\cdot, y_0)(x_0)$ is boundedly invertible. Then, for some $r', s' > 0$, and for all $y \in B_{s'}(y_0)$, the equation*

$$f(x, y) = 0$$

has exactly one solution $x = x(y)$ in $B_{r'}(x_0)$, and the resulting map

$$B_{s'}(y_0) \to X : y \mapsto x(y)$$

is continuous.

Proof: The equation $f(x, y) = 0$ is equivalent to the fixed point equation

$$x = T(x, y) := x - A^{-1} f(x, y).$$

Its iteration function, $T(\cdot, y)$, is a strict contraction near x_0 and uniformly so for y near y_0 since, by assumption,

$$DT(\cdot, y)(x) = 1 - A^{-1} Df(\cdot, y)(x)$$

is a continuous function of $(x, y) \in N$, and $DT(\cdot, y_0)(x_0) = 0$. Precisely, this implies that, for some $r' > 0$ and some $q < 1$, $\|DT(\cdot, y)(x)\| \le q$ on $B_{r'}(x_0, y_0)$. Thus $\forall \{(x, y), (x', y) \in B_{r'}(x_0, y_0)\}$

$$\|T(x', y) - T(x, y)\| \le \sup \|DT(\cdot, y)([x \,.\,.\, x'])\| \|x' - x\| \le q\|x' - x\|,$$

by the Meanvalue estimate. This shows that $T(\cdot, y)$ is a strict contraction on $B_{r'}(x_0)$ uniformly in $y \in B_{r'}(y_0)$. It remains to show that $T(\cdot, y)$ maps some closed subset of $B_{r'}(x_0)$ into itself. For this, observe that

$$\begin{array}{rcccc} \|T(x, y) - x_0\| & \le & \|T(x, y) - T(x_0, y)\| & + & \|T(x_0, y) - x_0\| \\ & \le & q\|x - x_0\| & + & (1 - q)r' \end{array}$$

for all $y \in B_{s'}(y_0)$ for some positive $s' \le r'$ so choosable since T is continuous and $T(x_0, y_0) = x_0$. For any such y, $T(\cdot, y)$ is a proper contraction on $B_{r'}(x_0)$ into $B_{r'}(x_0)$, hence has a unique fixed point there. Call this fixed point $x(y)$. Then

$$\begin{aligned} \|x(y) - x(y')\| &= \|T(x(y), y) - T(x(y'), y')\| \\ &\le \|T(x(y), y) - T(x(y), y')\| + \underbrace{\|T(x(y), y') - T(x(y'), y')\|}_{\le q\|x(y) - x(y')\|}. \end{aligned}$$

Therefore

$$\|x(y) - x(y')\| \le \frac{1}{1 - q} \|T(x(y), y) - T(x(y), y')\|$$

which implies the continuity of $y \mapsto x(y)$ even if we only know that $T(x(y), \cdot)$ is continuous. $\square$

H.P.(3) Prove the following stronger version of the Implicit Function Theorem which merely assumes the existence of an approximate inverse for $Df(\cdot,y)(x)$ uniformly in (x,y): *Let X,Y,Z be Bs's, $r,s>0$, $f: N := B_r(x_0)\times B_s(y_0)\to Z$ continuous, $f(x_0,y_0)=0$. Assume further that (i)* $\sup\{\|Df(\cdot,y)(x)\| : (x,y)\in N\}<\infty$; *(ii) for some boundedly invertible $A\in bL(X,Z)$,* $\sup\{\|1-A^{-1}Df(\cdot,y)(x)\| : (x,y)\in N\}<1$. *Then there exists $r',s'>0$ and exactly one function $g: B_{s'}(y_0)\to B_{r'}(x_0)$, necessarily continuous, so that $g(y_0)=x_0$ and $f(g(y),y)=0$ for all $y\in B_{s'}(y_0)$.*

Basic subjects not covered

Details of the representation of $C(T)^*$ by functions of bounded variation.
Brouwer and Schauder fixed point theorems.
Discretization of functional equations.
Stability of difference schemes for PDEs.
In addition, there is the whole richness of nonlinear functional analysis.

