11

Implicit Functions

11.1 Partial derivatives

To express the fact that z is a function of the two independent variables x and y we write $z = z(x, y)$. If variable y is fixed, then z becomes a function of x only, and if variable x is fixed, then z becomes a function of y only. The notation $\partial z/\partial x$, pronounced 'partial z with respect to x', is the derivative function of z with respect to x with y considered a constant. Similarly, $\partial z/\partial y$ is the derivative function of z with respect to y, with x held fixed. In any event, unlike dy/dx, the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ are not fractions. For the sake of notational succinctness the partial derivatives may be written as z_x and z_y.

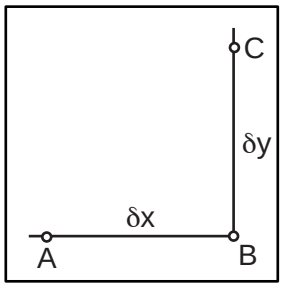
Examples.

1. For $z = x^2 + x^3y^{-1} + y^3 - 2x + 3y - 5$ we have

$$\frac{\partial z}{\partial x} = 2x + 3x^2y^{-1} - 2, \quad \frac{\partial z}{\partial y} = x^3(-y^{-2}) + 3y^2 + 3.$$

2. For $z = \sqrt{1 + x^2y^3}$ we have

$$\frac{\partial z}{\partial x} = \frac{2xy^3}{2\sqrt{1 + x^2y^3}}, \quad \frac{\partial z}{\partial y} = \frac{3y^2x^2}{2\sqrt{1 + x^2y^3}}.$$



We shall now linearize function $z(x, y)$ of two variables. For this we assume the function to be differentiable with respect to both x and y on a rectangle that includes points A, B, C in the figure to the right. Additionally we assume that the two partial derivatives of z, $\partial z/\partial x \equiv z_x$ and $\partial z/\partial y \equiv z_y$ are continuos on the rectangle.

Substitution of

$$F(B) = F(A) + (\frac{\partial F}{\partial x})_A \delta x + g_1 \delta x, \ g_1 \to 0 \text{ as } \delta x \to 0$$

into

$$F(C) = F(B) + (\frac{\partial F}{\partial y})_B \delta y + g_2 \delta y, \ g_2 \to 0 \text{ as } \delta y \to 0$$

results in

$$\delta z = F(C) - F(A) = ((\frac{\partial F}{\partial x})_A + g_1)\delta x + ((\frac{\partial F}{\partial y})_B + g_2)\delta y.$$

By continuity assumption on the partial derivatives we have that

$$\lim_{\delta x \to 0} (\frac{\partial F}{\partial x})_A + g_1 = (\frac{\partial F}{\partial x})_A \quad \text{and} \quad \lim_{\delta x, \delta y \to 0} ((\frac{\partial F}{\partial y})_B + g_2) = (\frac{\partial F}{\partial y})_A.$$

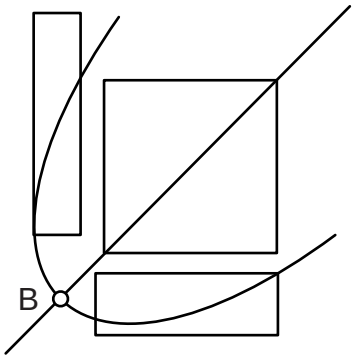
resulting in

$$dz = (\frac{\partial F}{\partial x})_A dx + (\frac{\partial F}{\partial y})_A dy$$

which is the equation of a plane.

11.2 Implicit Functions

The totality of points (x, y) satisfying the equation $F(x, y) = 0$ forms a curve. Given a value of the independent variable x, evaluation of y, supposing one exists, may require the approximate solution of $F(x, y) = 0$ by numerical means, such as the method of bisections or the metod of successive linearizations. It is possible that for one given x value there is a number of corresponding y values. Yet, it may happen that a restricted portion of the plane delineated by a (finite or infinite) rectangle contains an arc of the total curve that is the graph of some function $y = y(x)$. We say then that $F(x, y) = 0$ is an *implicit* representation of the function $y = y(x)$. The figure to the right shows the tilted ψ shaped curve implicit in some $F(x, y) = 0$. Intersection point B of the linear and parabolic branches of the curve is often referred to as a *bifurcation* point since the curve branches or forks at such a pont. The figure also shows three rectangles such that for any x in a rectangle there is only one corresponding $y(x)$ in the same rectangle, and the three arcs within the three rectangles shown are thus each the graphs of a function.



Examples.

1. The curve described by the equation $F(x, y) = y^2 - 2yx^2 + x^4 = 0$ is the parabola $y = x^2$, a fact discovered by observing that $F(x, y) = (y - x^2)^2 = 0$. Hence $F(x, y) = 0$ is the implicit representation of the (explicit) function $y(x) = x^2$.

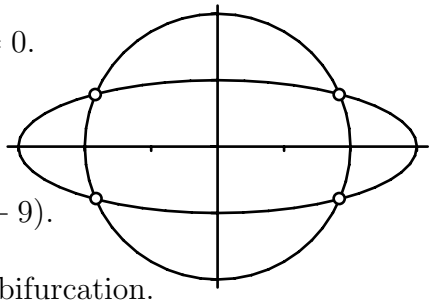
2. The curve of the equation $F(x, y) = x^2 + y^2 = 0$ consists of the single point $(0, 0)$.

3. The curve of $F(x, y) = x^2 + y^2 + 1$ is empty. Not one (real) pair (x, y) is found to satisfy this equation.

4. The curve of the equation $F(x, y) = x^2 - y^2 = (x - y)(x + y) = 0$ consists of the union of the two lines $y = x$ and $y = -x$. The intersection point $(0, 0)$ of the two lines is a bifurcation point. Under the restriction $y \geq 0$ the equation $F(x, y) = 0$ becomes the implicit representation of the sole explicit function $y(x) = |x|$, and under the restriction $y \leq 0$ the same equation $F(x, y) = 0$ becomes the implicit representation of the (single) explicit function $y(x) = -|x|$

5. The curve shown to the right is the locus of the totality of points satisfying

$$F(x, y) = x^4 + 10x^2y^2 + 9y^4 - 13x^2 - 45y^2 + 36 = 0.$$



It consists of the union of a circle and an ellipse, as in fact,

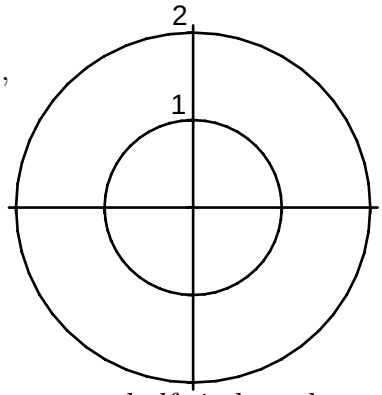
$$F(x, y) = G(x, y)H(x, y) = (x^2 + y^2 - 4)(x^2 + 9y^2 - 9).$$

The combined grapf of the circle and the ellipse shows four points of bifurcation.

6. The curve of the equation

$$F(x, y) = (x^2 + y^2)^2 - 3(x^2 + y^2) + 2 = 0$$

consists of two concentric circles as shown in the figure to the right, a fact discovered by observing that



$$F(x,y) = (x^2+y^2)^2 - 3(x^2+y^2) + 2 = (x^2+y^2-1)(x^2+y^2-2).$$

7. The curve described by the equation

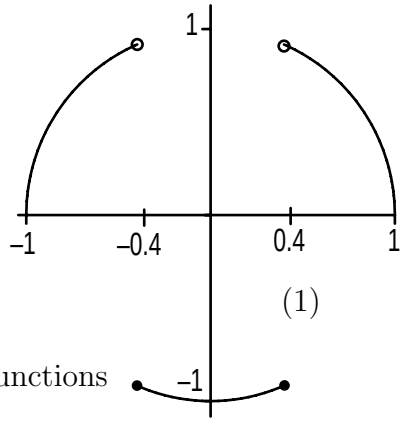
$$F(x,y) = x^2 + y^2 - 1 = 0,\ -1 \le x \le 1$$

is a full circle, which we may consider as consisting of two branches: an upper half circle and a lower half circle, represented by the two continuous functions, written explicitly as

$$y_1(x) = \sqrt{1-x^2} \quad \text{and} \quad y_2(x) = -\sqrt{1-x^2}$$

which are moreover differentiable both on $-1 < x < 1$. Under the additional restriction $y \ge 0$ the equation $F(x,y) = 0$ becomes the (unique) implicit representation of the function $y_1(x)$ only, and under the additional restriction $y \le 0$ the equation $F(x,y) = 0$ becomes the (unique) implicit representation of the function $y_2(x)$ only.

Relinquishing continuity, the equation $F(x,y) = 0$ may be the implicit representation of a host of other, discontinuous, functions like the one chosen at random and shown in the figure to the right. Here $y(x) \ge 0$ if $-1 \le x < -0.4$ and $0.4 < x \le 1$; and $y < 0$ if $-0.4 \le x \le 0.4$.



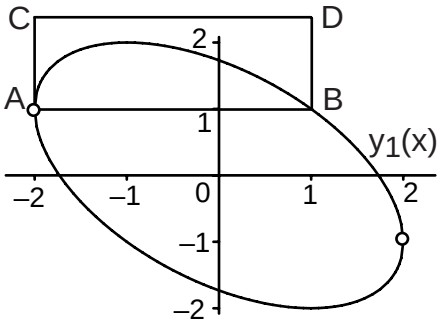
8. The graph of the equation

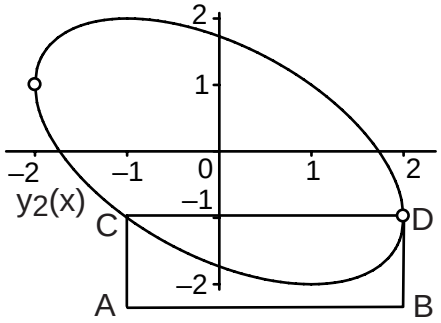
$$F(x,y) = x^2 + xy + y^2 - 3 = 0 \tag{1}$$

consists of the union of two curves explicitly described by the two functions

$$y_1(x) = \frac{1}{2}(-x + \sqrt{12-3x^2}),\quad y_2(x) = \frac{1}{2}(-x - \sqrt{12-3x^2}) \tag{2}$$

which are continuous on $-2 \le x \le 2$ and differentiable on $-2 < x < 2$. See the figure below.





With the restrictions $-2 \le x \le 1,\ 1 \le y \le \infty$ the equation $F(x,y) = 0$ becomes the implicit representation of the sole function $y = y(x)$ with the graph in the box $ABCD$ in the figure above to the left. By the restrictions $-1 \le x \le 2,\ -1 \le y \le -\infty$ the equation

$F(x,y)=0$ becomes the implicit representation of the sole function $y=y(x)$ with the graph in the box $ABCD$ in the figure above to the right.

We were able to produce the two explicit differentiable functions $y_1(x)$ and $y_2(x)$ out of implicit expression (1) because for any fixed x expression (1) reduces to a mere quadratic equation for y which can be solved by simple and formal algebraic means.

9. consider the function $F(x,y)=e^y+x^2-1=0$. We rewrite the equation as $e^y=1-x^2$ and restrict x to $(-1,1)$ and conclude that $y=\ln(1-x^2)$ is the sole explicit solution of $F(x,y)=0$.

10. If we choose in the equation

$$F(x,y)=xe^y+ye^x+x^2+y^2-9=0$$

$x=1$, then we are left with the **transcendental** equation $e^y+y+y^2-8=0$ which can be solved for y but numerically and approximately. Now, even if $F(x,y)=0$ is an implicit representation of the explicit $y(x)$, near $x=1$, there is no way this equation can be written explicitly in terms of elementary functions.

Before entertaining some general arguments, we shall theoretically demonstrate that the equation $F(x,y)=y^5+y-x+1=0$ is an implicit representation of one single function $y=f(x)$ for any x, inspite of the fact that it can not be turned explicit by any algebraic means. Indeed, for any x the equation $F(x,y)=0$ is a polynomial equation of odd degree in y and possesses at least one real root. Since $\partial F/\partial y=5y^4+1>0$ the function $F(x,y)$ is increasing for any fixed x and $F(x,y)=0$ only once. If $|x|>>1$, then $y=\sqrt[5]{x}$, nearly.

Definition *Let $F(x,y)=0$ is defined in the rectangle $a\le x\le b,\ a'\le y\le b'$. We shall say that $F(x,y)=0$ is an implicit representation of $y=f(x)$ in the rectangle, if for any $a\le x\le b$ there is a single $a'\le y\le b'$, such that the pair x,y satisfies the equation $F(x,y)=0$.*

Even if the explicit $y=f(x)$ can not be extracted from $F(x,y)=0$ we may still want to trace the graph of this function passing through point $P_0(x_0,y_0)$. To construct this curve we shall need a sequence of points $P_1(x_1,y_1),\ P_2(x_2,y_2),\ P_3(x_3,y_3)$ and so on closely strung on the curve. A way of moving from point P_0 on the curve to a nearby point P_1 consists of numerically solving the pair of equations

$$F(x,y)=0,\quad (x-x_0)^2+(y-y_0)^2=\epsilon^2 \tag{5}$$

in which $\epsilon>0$ is chosen small enough to produce close points for a smooth looking curve.

Is the general $F(x,y)=0$, actually represent $y=f(x)$ in the neighborhood of some $P(x_0,y_0)$, for which $F(x_0,y_0)=0$, even if this function can not be written explicitly? The following theorem provides a sufficient condition for the existence of a rectangle containing P inside which $F(x,y)=0$ defines the single function $y=y(x)$.

Implicit functions Theorem. *If function $F(x,y)$ of the two variables x and y satisfies the four conditions:*

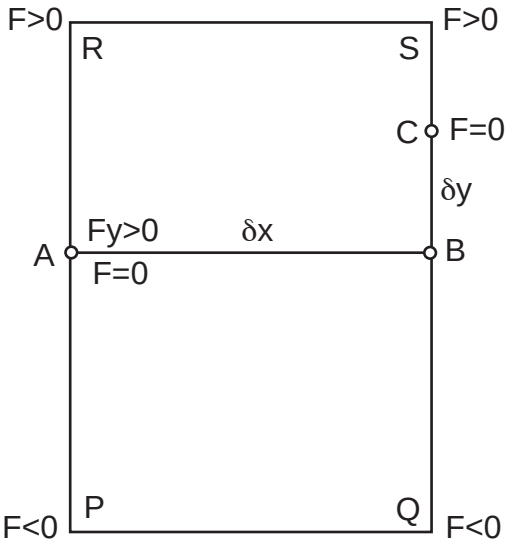
1. *$F(x_0,y_0)=0$ for $x=x_0,\ y=y_0$*

2. *$F(x,y)$ is continuous in some neighborhood of $x = x_0,\ y = y_0$, that is to say, $F(x,y) \to F(x_0,y_0)$ as $x \to x_0$ and $y \to y_0$*

3. *The partial derivatives F_x and F_y of F exist and are continuous in some neighborhood of $x = x_0,\ y = y_0$*

4. *$F_y(x_0,y_0) \neq 0$*

then there is a neighborhood of point (x_0,y_0) in which there is one and only one continuous and differentiable function $y = f(x),\ y_0 = f(x_0)$, that satisfies $F(x,f(x)) = 0$. Derivative function $f'(x)$ of the implicit function is continuous and is given by $f'(x) = -Fx(x,y)/Fy(x,y)$ with $y = f(x)$.



Proof. Refer to the figure above in which the sides of the rectangle $PQRS$ are parallel to the x and y axes. Assume that the function $F(x,y)$ satisfies the conditions of the theorem in and on the rectangle. At point A the partial derivative $F_y \neq 0$ and we assume that $(\partial F/\partial y)_A > 0$. Function $F(x,y)$ is thus increasing with y, and on some interval above point A, say up to point R, $F > 0$, and in some interval below point A, say up to pint P, $F < 0$. The continuity of F along x implies the existence of δx such that $F(S) > 0$ and $F(Q) < 0$. The continuity of F along y implies that $F = 0$ at point C strictly between S and Q. By the assumption that $F_y > 0$ on the rectangle, $F = 0$ on the side QS only once. Implicit function $y = f(x)$ is the y coordinate of point C.

We will show now that $y(x)$ is differentiable at point A and similarly that the function is differentiable at any other intermediate point on the curve. Using the fact that $F(A) = F(C) = 0$ we obtain from the previous paragraph that

$$0 = (\frac{\partial F}{\partial x})_A + (\frac{\partial F}{\partial y})_B \frac{\delta y}{\delta x} + g_1 + g_2 \frac{\delta y}{\delta x}.$$

The continuity assumption on the partial derivatives implies that

$$\lim_{\delta x \to 0} (\frac{\partial F}{\partial y})_B = (\frac{\partial F}{\partial y})_A$$

and

$$(\frac{\partial F}{\partial x})_A dx + (\frac{\partial F}{\partial y})_A dy = 0$$

or $y'(x) = -(\partial F/\partial x)/(\partial F/\partial y)$. Derivative function $y'(x)$ is continuous since it consists of the ratio of two continuous functions with $F_y \neq 0$. Function $y(x)$ is differentiable at point A and is therefore also continuous there. End of proof.

Examples and counterexamples.

1. For $y^2 - x = 0$ we obtain by implicit differentiation $2yy' - 1 = 0$ leading to the absurdity $0 = -1$ if $x = y = 0$ because the function $y = \sqrt{x}$ is not differentiable at $x = 0$.

2. For $(x - y)^2 = x^2 - 2xy + y^2 = 0$ we obtain by implicit differentiation $(x - y)(1 - y') = 0$ and $y' = 1$ if $x - y \neq 0$. Actually $y = x$ and $y' = 1$ for any x.

3. Expression $(x-y)(x+y) = x^2-y^2 = 0$ represents two orthogonal lines through $x = y = 0$. By implicit differentiation $x-yy' = 0$ and at $x = y = 0$ we get $0 = 0$ because of the ambiguity of the bifurcation.

4. Expression $x^2 + y^2 = 0$ represents the mere point $x = y = 0$. By implicit differentiation $x + yy'$=0, and at $x = y = 0$ this reduces to $0 = 0$.

5. The equation $F(x, y) = y^3 - x^3 = 0$ is equivalent to $y = x$ for any x. We have here that $3y^2y' - 3x^2 = 0$ which reduces to $0 = 0$ at $P(0, 0)$.

6. Consider

$$F(x, y) = x^3y + xy^3 + x + y - 4 = 0.$$

Recalling the chain rules $(y^2)' = 2yy'$ and $(y^3)' = 3y^2y'$, and diffrentiating both side of the above equation with respect to x we have

$$(3x^2 + x^3y') + (y^3 + 3xy^2y') + 1 + y' = 0$$

which is linear in y', and is readily solved to produce

$$y' = -\frac{3x^2y + y^3 + 1}{x^3 + 3xy^2 + 1} = -\frac{F_x(x, y)}{F_y(x, y)}.$$

For $x = 1,\ y = 1$, that satisfy the present $F(x, y) = 0$, yields $y'(1) = -1$

7. For $F(x, y) = y^3-6xy+5 = 0$ we shall compute y' and y'' at $P(1, 1)$. From $(y^3-6xy+5)' = 0$ we obtain $y^2y' - 2y - 2xy' = 0$, and $y'(1) = -2$. From $(y^2y' - 2y - 2xy')' = 0$ we obtain $2y'y' + y^2y'' - 4y' - 2xy'' = 0$, and $y''(1) = 16$.

8. Expression $x^2 + y^2 = r^2$ describes both the upper and the lower arc of a circle centered at $C(0, 0)$ and having radius r. Repeated differentiation produces here

$$x + yy' = 0 \qquad \text{and} \qquad 1 + (y')^2 + yy'' = 0$$

so that the critical points of both functions implicit in the equation of the circle happen to be at $x = 0,\ y = \pm r$. But if $y' = 0$, then $y'' = -1/y$, and $y'' = -1/r$ if $y = +r$, implying a local maximum, and $y'' = 1/r$ if $y = -r$ implying a local minimum.

9. For the function

$$F(x,y) = G(x,y)H(x,y) = (x^2+y^2-4)(x^2+9y^2-9)$$

considered before we have

$$\frac{\partial F}{\partial x} = \frac{\partial G}{\partial x}H + G\frac{\partial H}{\partial x} \quad \text{and} \quad \frac{\partial F}{\partial y} = \frac{\partial G}{\partial y}H + G\frac{\partial H}{\partial y}$$

and both $\partial F/\partial x = 0$ and $\partial F/\partial y = 0$ at the bifurcation points at which at once $G = H = 0$.

Exercises.

1. Find all critical points of the folium of Descartes $x^3 + y + 3 - 3xy = 0$. Ans. $y' = -(x^2-y)/(y^2-x)$. $x_0 = 2^{1/3}, y_0 = 4^{1/3}$. Max.

2. Show that if $f(x+y) = f(x)+f(y)$ for any x and y, then $f(x) = ax$ for constant a. Hint: First we observe that if $x = y = 0$, then $f(0) = 2f(0)$, and hence $f(0) = 0$. We introduce the auxiliary variable $u = x + y$, and have upon differentiation with respect to x and y that $f'(u) = f'(x)$ and $f'(u) = f'(y)$, or $f'(x) = f'(y)$ for any x and y, implying that $f'(x) = a$ for some constant a. Hence $f(x) = ax + b$, but $b = 0$.

3. Show that if function $f(x)$ is such that

$$f(x+y)f(x-y) = f^2(x) - f^2(y)$$

for any x and y, then $f(0) = 0$, and $f''(x)/f(x)$ =constant. Hint: Set in the above equation $x = (u+v)/2, y = (u-v)/2$ and differentiate it twice with respect to u and v.

4. Prove that if differentiable function $f(u)$ is such that $f(x+y) = f(x) + f(y)$ for any x and y, then $f'(x) = f'(y)$ for any x and y.

5. Prove that if differentiable function $f(u), u > 0$ is such that $f(xy) = f(x) + f(y)$ for any positive x and y, then $xf'(x) = yf'(y)$ for any positive x and y.

6. Prove that if differentiable function $f(u)$ is such that $f(x+y) = f(x)f(y)$ for any x and y, then $f'(x)/f(x) = f'(y)/f(y)$ for any x and y.

7. For what function does it happen that $f(xy) = f(x)f(y)$?

11.3 Tracking an implicit curve

Pursuing an implicit differentiable curve, or trajectory, entails placing close points on it. Presented here is a leap-and-land algorithm for doing that. Let $F(x,y) = 0$ be the implicit equation of the curve and let $A(x_0, y_0)$ be a point on it such that $F(A) = 0$. The equation of the tangent line to the curve at point A is

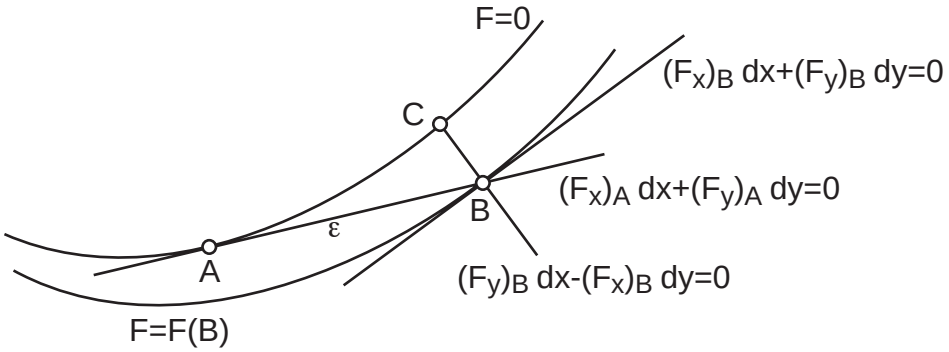
$$(\frac{\partial F}{\partial x})_A dx + (\frac{\partial F}{\partial y})_A dy = 0 \quad \text{or} \quad (\frac{\partial F}{\partial x})_A (x - x_0) + (\frac{\partial F}{\partial y})_A (y - y_0) = 0$$

We propose to leap forward from point A and place point B on the tangent line at distance ϵ from A. Point B is close to the curve (one must beware, though, of the danger of jumping from one branch of the function to another that may happen to be nearby) and from it we will attempt a short distance return that will land us back on the curve.

The tangential leap is restricted to distance ϵ with $dx^2 + dy^2 = \epsilon^2$, resulting in

$$dx = \epsilon \frac{F_y}{\sqrt{F_x^2 + F_y^2}} \ , \ dy = -\epsilon \frac{F_x}{\sqrt{F_x^2 + F_y^2}} \tag{1}$$

with the signs chosen to produce a clock-wise tracking, and with the partial derivatives evaluated all at point A. For instance, if $F(x, y) = x^2 + y^2 - 1$, then $F_x = 2x, F_y = 2y$, and $dx = \epsilon y_0$ and $dy = -\epsilon x_0$. A tangential leap originating at $A(x_0, y_0)$ terminates at point $B(x_0 + \epsilon y_0, y_0 - \epsilon x_0)$. Look an the figure below.



We consider point $B(x_1, y_1)$ as situated on the curve $F(x, y) = F(B) = F(x_1, y_1)$ having the tangent line

$$(\frac{\partial F}{\partial x})_B dx + (\frac{\partial F}{\partial y})_B dy = 0 \ \text{ or } \ (\frac{\partial F}{\partial x})_B (x - x_1) + (\frac{\partial F}{\partial y})_B (y - y_1) = 0.$$

Point C is the intersection point of the line orthogonal to this tangent line and the curve $F = 0$. To reach point C from point B we need the corrections dx and dy such that

$$F(x_1 + dx, y_1 + dy) = 0 \text{ under the restriction that } (F_y)_B dx - (F_x)_B dy = 0.$$

This system of equations is solved approximately by the linearization

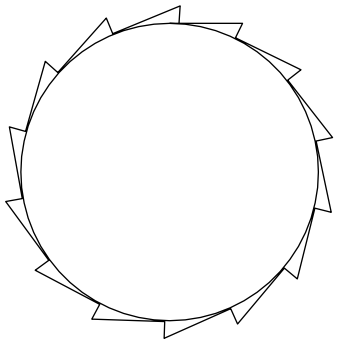
$$F(x_1 + dx, y_1 + dy) = F(x_1, y_1) + (F_x)_B dx + (F_y)_B dy = 0$$

to yield

$$dx = -\frac{F}{F_x^2 + F_y^2} F_x, \ dy = -\frac{F}{F_x^2 + F_y^2} F_y$$

in which function F and its partial derivatives F_x and F_y are evaluated at point $B(x_1, y_1)$.

If $|F(C)|$, evaluated at $C(x_1 + dx, y_1 + dy)$, is less than some prescribed tolerance, then a new leap is executed and a new landing on the curve attempted. If $|F(C)|$ is deemed not sufficiently small, then point C is taken instead of point B and the linearization is repeated.



The above figure shows a fourteen-step tracking of the unit circle $F(x,y) = x^2 + y^2 - 1$ by linear leaps of $\epsilon = 0.49$ and a single orthogonal correction.

11.4 The osculating line and circle

Let the equation $F(x,y) = 0$ be the implicit representation of the function $y = f(x)$, which we assume twice differentiable around some point P on its curve. The equation of the tangent (osculating) line to the graph of the function at point $P(x_1, y_1)$ on the curve is

$$(x - x_1)(F_x)_1 + (y - y_1)(F_y)_1 = 0$$

in which, for short, $F_x = \partial F/\partial x$ and $F_y = \partial F/\partial y$, and where the subscript 1, refers to evaluations at point $P(x_1, y_1)$. Subsequently we will drop the subscript 1 for notational neatness. The line is said to be osculating[1] to the graph since $y(x_1)$ and $y'(x_1)$ for the line are the same as for the function.

To obtain $y''(x)$ for the implicit function we start with

$$dF = 0 = F_x\,dx + F_y\,dy \quad \text{or} \quad G = \frac{dF}{dx} = 0 = F_x + F_y y'$$

and proceed to obtain

$$\frac{dG}{dx} = \frac{d^2F}{dx^2} = 0 = \frac{dF_x}{dx} + \frac{d(F_y y')}{dx} = \frac{dF_x}{dx} + \frac{dF_y}{dx}y' + F_y y''$$

and consequently

$$0 = F_{xx} + F_{xy} + (F_{yx} + F_{yy}y')y' + F_y y''.$$

Assuming that $F_{xy} = F_{yx}$ and using $y' = -F_x/F_y$ we obtain the second derivative in terms of the partial derivatives as

$$y'' = \frac{-F_{xx}F_y^2 + 2F_{xy}F_xF_y - F_{yy}F_x^2}{F_y^3}.$$

We write the equation of the general circle

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

1 Related to *scale*, to mount, to climb up, to cover

and repeatedly differentiate it to have

$$(x - x_0) + (y - y_0)y' = 0 \quad \text{and} \quad 1 + y'^2 + (y - y_0)y'' = 0$$

from which we obtain the center $C(x_0, y_0)$ and radius r as

$$x_0 = x - \frac{1 + y'^2}{y''}y', \quad y_0 = y + \frac{1 + y'^2}{y''} \quad r = \frac{(1 + y'^2)^{3/2}}{|y''|}.$$

Putting into the above expressions y' and y'' of any function we obtain the center and radius of the osculating circle to the curve at point $P(x, y)$ on it. If the function is given implicitly as $F(x, y)$, then in terms of the partial derivatives

$$x_0 = x + \frac{F_x^2 + F_y^2}{\Delta}F_x \quad y_0 = y + \frac{F_x^2 + F_y^2}{\Delta}F_y \quad r = \frac{\sqrt{F_x^2 + F_y^2}}{|\Delta|}$$
$$\Delta = -F_{xx}F_y^2 + 2F_{xy}F_xF_y - F_{yy}F_x^2.$$

Examples.

1. If $f(x) = x^2$, then $f'(x) = 2x$ and $f''(x) = 2$. For this parabola

$$x_0 = -x^3, y_0 = \frac{1}{2}(1 + 6x^2), \text{ and } r = \frac{1}{2}(1 + 4x^2)^{3/2}.$$

In particular, if $x = 0$, then $x_0 = 0$, $y_0 = 1/2$, and $r = 1/2$.

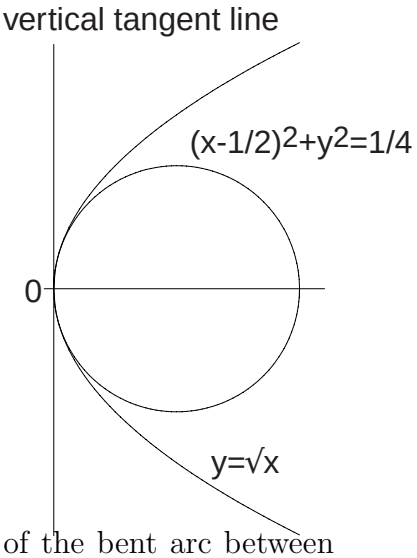
2. Take $F(x, y) = y^2 - x$ and $P(0, 0)$. Here, at P

$$F_x = -1, \ F_y = 0 \quad F_{xx} = 0, \ F_{xy} = 0, \ F_{yy} = 2 \quad \Delta = -2$$

so that

$$x_0 = 1/2, \ y_0 = 0, \ r = 1/2$$

and the osculating circle is as in the figure below.



In the tracking algorithm of the previous section instead of leaping from point A on the curve to point B nearby, moving on the tangent line, we could proceed along an arc of the osculating circle to be even closer to the curve before attempting to land on it.

Consider the figure below to the right. Geometrically, we my construct the osculating circle this way. The center of a circle, and consequently its radius is found at the intersection of two normals. Let the arc in the figure to the right be the graph of function f, twice differentiable in some neighborhood of x_1. A small portion of the bent arc between points A and B appears circular. The two normals raised at point A and B intersect at point C. If point C reaches a limit position as $B \to A$, then point C is the center of the osculating circle to the curve at point A.

The coordinates of point C are found from the intersection of the two normals

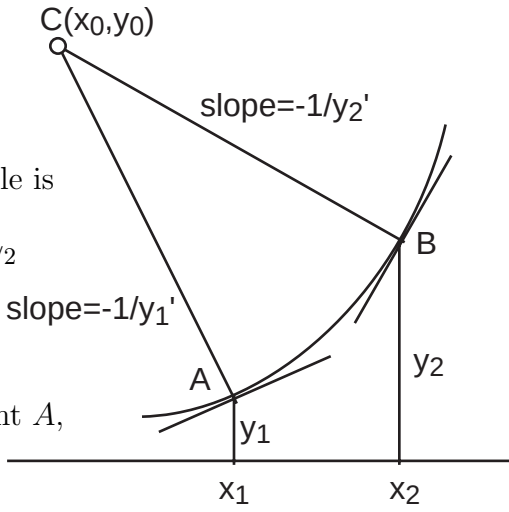
$$y - y_1 = -\frac{1}{y_1'}(x - x_1),\ y - y_2 = -\frac{1}{y_2'}(x - x_2)$$

and

$$x_0 = \lim_{x_2 \to x_1} \frac{(y_2 - y_1)y_1'y_2' - x_1y_2' + x_2y_1'}{y_1' - y_2'}.$$

Application of L'Hôpital's rule produces the limit values

$$x_0 = x - \frac{y'}{y''}(1 + y'^2),\ y_0 = y + \frac{1}{y''}(1 + y'^2)$$



in which the subscript 1 was removed. The radius r of the limit circle is

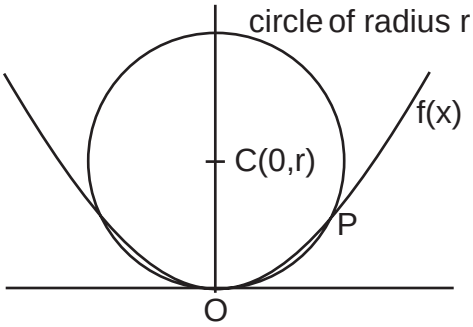
$$r = [(x - x_0)^2 + (y - y_0)^2]^{1/2} = \frac{1}{|y''|}(1 + y'^2)^{3/2}$$

as before.

The inverse $\kappa = 1/r$ is called the **curvature** of the curve at point A,

$$\kappa = \frac{|y''|}{(1 + y'^2)^{3/2}}.$$

For a line $\kappa = 0$, and at a corner $\kappa = \infty$.



We can obtain the osculating circle also also this way. Let $f(x)$ be twice differentiable at $x = 0$ and such that $f(0) = f'(0) = 0$. The equation of a circle of radius r centered at $C(0, r)$ and passing through $O(0, 0)$ is $x^2 + (y - r)^2 = r^2$. The intersection of this circle and $f(x)$ occurs at points O and P. See the figure to the right. At an intersection point $y = f(x)$ and at such a point $x^2 + (f - r)^2 = r^2$ or

$$r = \frac{x^2 + f^2}{2f}.$$

Letting $x \to 0$, or $P \to O$, and using L'Hôpital's rule we get

$$\lim_{x \to 0} \frac{x^2 + f^2}{2f} = \lim_{x \to 0} \frac{2x + 2ff'}{2f'} = \lim_{x \to 0} \frac{1 + f'^2 + ff''}{f''} = \frac{1}{f''(0)}$$

which is the radius of the osculating circle at P.

For the function $f(x) = (e^x + e^{-x})/2 - 1$ this yields $r = 1$.

11.5 The osculating ellipse

We shall write the equation of the osculating ellipse for a symmetrical function at $P(0,0)$. In its canonical form the equation of the ellipse is

$$b^2(x-x_0)^2 + a^2(y-y_0)^2 = a^2b^2.$$

Differentiating the equation repeatedly we obtain the set of four equations

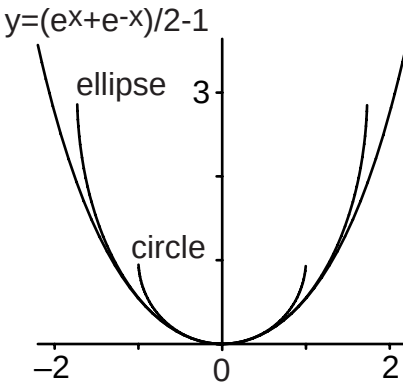
$$b^2(x-x_0) + a^2(y-b)y' = 0, \quad b^2 + a^2y'^2 + a^2(y-y_0)y'' = 0$$

$$3y'y'' + (y-y_0)y''' = 0, \quad 3yy''^2 + 4y'y''' + (y-y_0)y'''' = 0.$$

For $x = 0, y = 0, y' = 0, y'' > 0, y''' = 0, y'''' > 0$ the equations are reduced to

$$x_0 = 0,\ y_0 = b,\ b = 3y''/y'''' \quad \text{and} \quad a = \sqrt{3/y''''}.$$

For the function $f(x) = (e^x + e^{-x})/2 - 1$ the semi-axes are $a = \sqrt{3}, b = 3$. See the figure to the right.



11.6 From Fixed-point to Newton to Halley to Higher Order Iterations

11.6.1 Fixed point iteration

Definition. *Real number a is a fixed point of function $f(x)$ if $a = f(a)$.*

The, utterly simple, fixed point iteration method recursively gets x_{n+1} from x_n by $x_{n+1} = f(x_n)$, starting with an initial guess x_0. Under certain conditions $x_n \to a$, the fixed point of $f(x)$, as $n \to \infty$.

The First Fixed Point Iteration Theorem: *Let a be a fixed point of $f(x)$. If $|f'(x)| < 1$ on the open interval $I = (a-\delta, a+\delta),\ \delta > 0$, then the sequence $\{x_n\}$ generated by the fixed point iteration $x_{n+1} = f(x_n)$ converges to a, for any initial guess $x_0 \in I$.*

Proof. Let $x_0 \in I$. We write

$$x_1 - a = f(x_0) - a = f(x_0) - f(a)$$

apply the mean-value theorem in the form $f(x_0) = f(a) + (a - x_0)f'(\xi)$, and have

$$x_1 - a = (x_0 - a)f'(\xi) \text{ or } |x_1 - a| = |x_0 - a|\,|f'(\xi)|$$

where $\xi \in I$. By virtue of the fact that $|f(\xi)| < 1$ the next iterant $x_1 = f(x_0)$ is nearer to a then x_0 and is therefore also in I. End of proof.

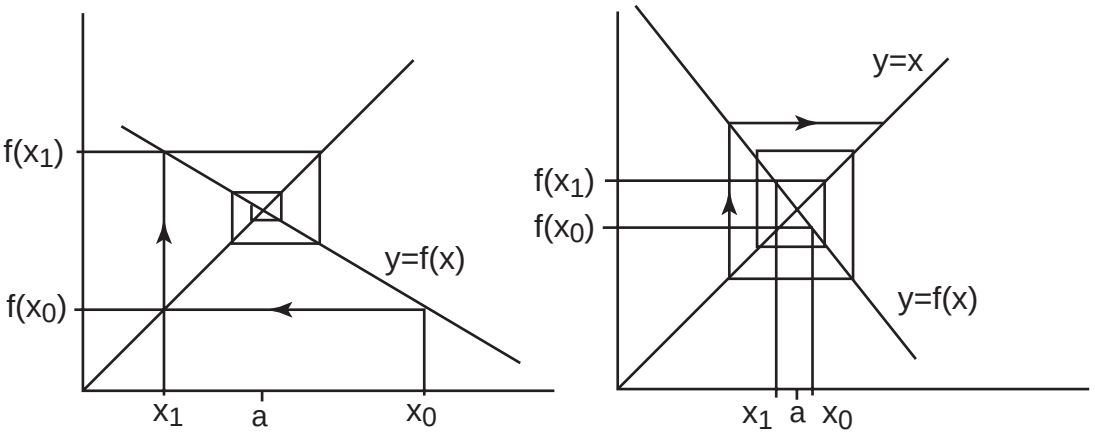
Example. Consider the linear function $f(x) = k(x-a) + a,\ k \neq 1$. for which

$$x_1 - a = k(x_0 - a).$$

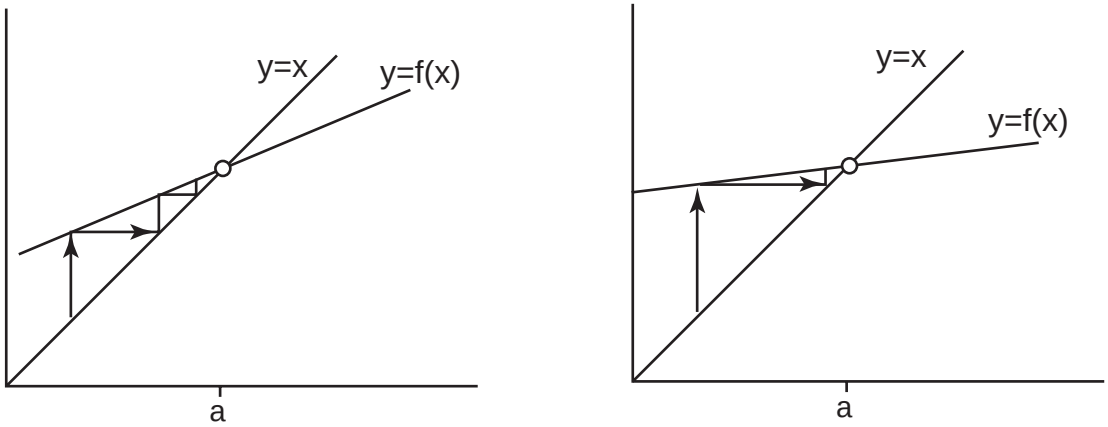
If $k > 0$, then $x_1 - a$ and $x_0 - a$ have the same sign, but if $k < 0$, then $x_1 - a$ and $x_0 - a$ have opposite signs.

Chapter 11

See the figure below.



Moreover, as $|k|$ becomes smaller, or as the linear function $f(x) = k(x-a)+a$ tilts closer to being horizontal, convergence becomes faster. See the figure below.



In the general case, we expect the fixed point iteration $x_1 = f(x_0)$ to be fast converging if function $f(x)$ is flat and nearly horizontal at $x = a$. Namely, if $f'(a) = f''(a) = \cdots = 0$, which is the assertion of the next theorem.

The second, general, fixed point theorem: *Let a be a fixed point of function $f(x)$, $a = f(a)$. Suppose $f(x)$ has a bounded derivative of order $m+1$ in an open interval containing fixed point a. If $f'(a) = f''(a) = \ldots = f^{(m)}(a) = 0$, but $f^{(m+1)}(a) \neq 0$, then the sequence x_n produced by the recursion $x_{n+1} = f(x_n)$ is such that $|a - x_{n+1}| < c|a - x_n|^m$, for some constant $c > 0$, provided that x_0 is taken close enough to a.*

Proof: We shall prove the theorem for the specific case of $m = 2$. Let $f''(x)$ be bounded on the interval $I = (a-\delta, a+\delta)$, $\delta > 0$, and assume x to be in this interval. Taylor's expansion of $f(x)$ around point a is

$$f(x) = f(a) + (x-a)f'(a) + \frac{1}{2}(x-a)^2 f''(a) + \frac{1}{6}(x-a)^3 f'''(\xi),\ a < \xi < x$$

if x is to the right of a. The assumptions $a = f(a)$, $f'(a) = 0$, $f''(a) = 0$ reduce this equality to $f(x) - a = (1/6)(x-a)^3 f'''(\xi)$ or $|f(x) - a| \leq (M/6)|x-a|^3$, where M is an upper bound

on $|f'''(x)|$ in I. For $x_1 = f(x_0)$ the inequality becomes

$$|x_1 - a| \le c|x_0 - a|^3, \quad c = \frac{M}{6}$$

and if $|x_0 - a| < 1$, then $|x_0 - a|^3$ is much smaller than one. For x_0 sufficiently close to a, x_1 is closer to a then x_0, and hence it is also in I, and so on. End of proof.

Exercises.

1. If the fixed-point iteration

$$x_1 = \frac{x_0^2 + x_0 + 2}{x_0^2 + x_0 + 1}, \quad x_1 = f(x_0)$$

converges, to what number does it converge? Show that $x_1 > 0$ for any x_0. Consider $f'(x)$ and prove that it converges from any starting point. What is x_1 if x_0 is very large?

2. If the fixed-point iteration

$$x_1 = \frac{x_0}{1 + \sqrt{1 + x_0^2}}, \quad x_1 = f(x_0)$$

converges, to what number does it converge? Does the scheme converge from any starting point? What is x_1 if x_0 is very large?

3. Study the convergence, or divergence, of

$$x_n = e^{x_0 - 0.5} - 0.5.$$

Start from $x_0 = 0.45$, then from $x_0 = 0.55$.

4. Find fixed points a of

$$f(x) = \alpha x(1 - x).$$

For what α is $|f'(a)| < 1$?

5. Study the behavior of the fixed point iterates of $f(x) = \alpha x(1 - x)$ for $a = 3.45$ and $a = 3.5699465$. Start with x_0 only slightly different than the fixed point.

6. Study the behavior of the iterates of $x_1 = 1 - x_0 + x_0^2$. Start with $x_0 = 0.99$ and $x_0 = 1.01$.

7. Study the behavior of the iterates of $x_1 = 1/(2 - x_0)$. Start with $x_0 = 0.99$ and $x_0 = 1.01$.

8. Study the behavior of the iterates of $x_1 = 1/(2 - x_0)^2$. Start with $x_0 = 0.99$ and $x_0 = 1.01$.

9. For what c does $f(x) = x^2 + c$ have a fixed point? Study the behavior of the iterates of $x_{n+1} = f(x_n)$ for $c = -1, -1.5$ and $c = -2$. Start with $x_0 = 0$ and carry at least 20 iterative steps.

10. Apply the fixed point iteration $x_{n+1} = f(x_n)$ to the function $f(x) = kx + (1 - k)a$ with $0 \le k < 1$. Show that $x_n = k^n(x_0 - a) + a$ so that $x_n \to a = f(a)$ as $n \to \infty$.

Theorem: *Let function $y = f(x)$ be continuous and such that for any $a \le x \le b$ also $a \le y \le b$ (concisely put $f : [a, b] \to [a, b]$.) Then $f(x)$ has at least one fixed point $c = f(c)$.*

Proof. By the assumptions $f(a) \geq a$ and $f(b) \leq b$, hence according to the Intermediate Value Theorem there exists a number $c \in [a, b]$ such that $c = f(c)$. End of proof.

The Banach fixed point theorem: *Let $I = [a, b]$ be a closed, finite or infinite, interval. If function $f(x)$ is such that:*

1. *for any $x \in I$ also $f(x) \in I$*

2. *$|f(x) - f(y)| \leq q|x - y|$ with $0 \leq q < 1$*

Then

1 *function $f(x)$ has a unique fixed point $a = f(a)$*

2 *for any $x_0 \in I$ the sequence $x_0, x_1, x_2, \ldots$ generated by $x_{n+1} = f(x_n)$ converges to the fixed point a.*

Remark. Notice that $f(x)$ is not assumed continuous in this theorem.

Proof. To prove uniqueness suppose $a \neq b$ are two fixed points of $f(x)$ in I. The assumptions on a and b imply that

$$|a - b| = |f(a) - f(b)| \leq q|a - b| < |a - b|$$

which is a contradiction unless $a = b$.

We will prove the second part of the theorem by showing that:

1. The sequence $\{x_n\}_{n=0}^{n=\infty}$ generated by $x_{n+1} = f(x_n)$ is a Chaucy sequence and hence converging to some real a.

2. $a \in I$.

3. $a = f(a)$

We have that

$$|x_{n+1} - x_n| = |f(x_n) - f(x_{n-1})| \leq q|x_n - x_{n-1}| \leq \cdots \leq q^n|x_1 - x_0|.$$

Also, since $x_n - x_0 = (x_n - x_{n-1}) + (x_{n-1} - x_{n-2}) + \cdots + (x_1 - x_0)$, then

$$|x_n - x_0| \leq |x_n - x_{n-1}| + |x_{n-1} - x_{n-2}| + \cdots + |x_1 - x_0|$$

or

$$|x_n - x_0| \leq (q^{n-1} + q^n + \cdots + 1)|x_1 - x_0| = \frac{1 - q^n}{1 - q}|x_1 - x_0|.$$

Similarily,

$$|x_m - x_n| \leq q^n \frac{1 - q^{m-n}}{1 - q}|x_1 - x_0|, \quad m > n$$

implying that indeed $\{x_n\}_{n=0}^{n=\infty}$ is a Cauchy sequence which has a limit, say a.

Since $x_0 \in I$, then also $x_1 = f(x_0) \in I$ and so on for all x_n. Hence $a \in I$.

To prove that a is a fixed point of $f(x)$ assume it is not so, that $|f(a) - a| = \epsilon > 0$. Since $x_n \to a$, natural number N exists such that $|x_n - a| \leq \epsilon/2$ for all $n \geq N$. But

$$|f(a) - a| = |(f(a) - x_{N+1}) + (x_{N+1} - a)| \leq |f(a) - x_{N+1}| + |a - x_{N+1}|$$

$$= |f(a) - f(x_N)| + |a - x_{N+1}|$$
$$\leq q|a - x_N| + |a - x_{N+1}| < \epsilon$$

which is a contradiction, and a is a fixed point of $f(x)$. End of proof.

Theorem: *Let a be the fixed point of $f(x)$ under the assumptions of the previous theorem. Then for any natural n*

$$|x_n - a| \leq \frac{q^n}{1-q}|x_1 - x_0|$$

or equivalently

$$|x_{n+1} - a| \leq \frac{q}{1-q}|x_{n+1} - x_n|$$

and

$$|x_{n+1} - a| \leq q|x_n - a|.$$

Proof. From the properties of $f(x)$ and the way x_{n+1} is obtained from x_n we have that

$$|x_n - a| = |f(x_{n-1} - f(a)| \leq q|x_{n-1} - a|$$
$$\leq q^2|x_{n-2} - a| \leq \cdots \leq q^n|x_0 - a|.$$

But

$$|x_0 - a| = |x_0 - x_1 + x_1 - a| \leq |x_0 - x_1| + |x_1 - a|$$
$$\leq |x_1 - x_0| + q|x_0 - a|$$

and

$$|x_0 - a| \leq \frac{1}{1-q}|x_0 - x_1|.$$

End of proof.

We briefly consider the appliction of the fixed point iteration to the solution of the linear system of equations $Ax = f$. We write $x = x + \alpha(Ax - f), \alpha \neq 0$ which is equivalent to the original system. How to fix α so that $x_1 = x_0 + \alpha(Ax_0 - f)$ converges, and at a brisk pace, is not a simple matter, but an example will do for now. Take $\alpha = 1/2$, start with the initial vector $x_0 = (0, 0)$ and carry out several iteratios towrd the solution of the linear system of two equations in two unknowns

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ -5 \end{bmatrix}, \quad Ax = f, \quad x = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

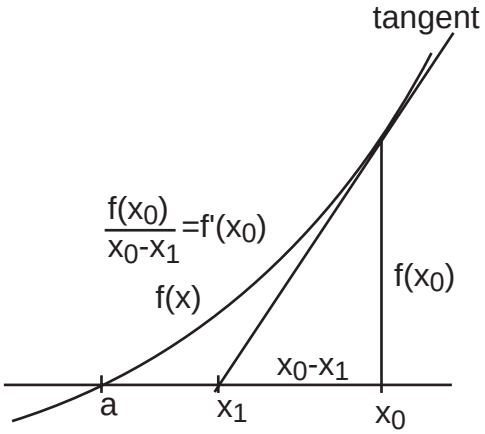
11.6.2 The Newton-Raphson iteration

Finding the root a of the nonlinear equation $f(x) = 0$. is equivalent to locating the fixed-point of $x = x + g(x)f(x)$, $g(a) \neq 0$. We write $F(x) = x + gf$ and seek $g(x)$ so that $F'(a) = 0$ to secure, in view of the previous theorem, a quadratic convergence for $x_{n+1} = x_n + F(x_n)$. Differentiating $F(x)$ by the product rule we have $F'(x) = 1 + g'f + gf'$, and $F'(a) = 1 + g'f(a) + gf'(a) = 1 + gf'(a)$ since $f(a) = 0$. Conceding that we do not know

fixed point a we settle for $a = x_n$ to have $g(x_n) = -1/f'(x_n)$, or in short $g = -1/f'_n$. With this g the fixed-point iteration becomes

$$x_{n+1} = x_n - \frac{f_n}{f'_n}$$

which is the Newton-Raphson method, presently shown to quadratically converge to a simple root of $f(x) = 0$. To prove the quadratic convergence of the NR method we derive it from the fixed point iteration by taking $F(x) = x + g(x)f(x), g(x) = A/f'(x)$ for constant A, and fix it from $F'(a) = 0$ as $A = -1$ independently of a. For a geometrical interpretation of the NR method see the figure below.



To observe the quadratic convergence of the NR method we select $f(x) = x^2 - 1$ for which $x_1 = x_0 - (x_0^2 - 1)/(2x_0)$, or $x_1 - 1 = x_0 - 1 - (x_0^2 - 1)/(2x_0)$ resulting in $x_1 - 1 = (1/2x_0)(x_0 - 1)^2$ or $x_1 - 1 = (1/2)(x_0 - 1)^2$ if $x_0 = 1$, nearly.

Exercises.

1. Show that the Newton-Raphson method converges to the (repeating) root of $f(x) = x^2 = 0$ only linearly.

2. Fix constant A so as to make the modified Newton-Raphson method $x_{n+1} = x_n - Af(x_n)/f'(x_n)$ converge quadratically to the zero root of the cubic $f(x) = a_2x^2 + a_3x^3$.

3. Study the convergence of the Newton-Raphson method in locating the root of $f(x) = x^{1/3}$.

It may happen that even if root a of $f(x) = 0$ is unknown still $f'(a)$ is known via a differential equation for f. Consider using the NR method for computing the natural logarithm. Here $f(x) = e^x - \alpha$, the root of which is $a = \ln \alpha$, and $f'(a) = \alpha$. Using $x_{n+1} = x_n - f_n/f'_n$ we have $x_{n+1} = x_n - (e^{x_n} - \alpha)/e^{x_n}$, while using $x_{n+1} = x_n - f_n/f'(a)$ we have $x_{n+1} = x_n - (e^{x_n} - \alpha)/\alpha$.

11.6.3 The Halley iteration

We write $x = F(x)$, $F(x) = x + g(x)f(x)$ so that the fixed point a of F is a root of $f, f(a) = 0$, if $g(a) \neq 0$, and we prepare to consider the fixed point iterative method $x_{n+1} = F(x_n)$. We differentiate $F(x)$ twice to have

$$F'(x) = 1 + gf' + g'f \quad \text{and} \quad F''(x) = gf'' + 2g'f' + g''f$$

and if $F'(a) = F''(a) = 0$, then the sequence generated by the recursion $x_{n+1} = F(x_n)$ is, under propitious circumstances, cubically converging to fixed point a. But we do not know a. The correct NR method is obtained from the fixed point iteration method under the assumption that $g(x_n)$ is a constant, or $g' = g'(x_n) = 0$. Here we assume $g''(x_n) = 0$. Not knowing fixed point a we replace the conditions $F'(a) = 0$ and $F''(a) = 0$ by $F'(x_n) = 0$ and $F''(x_n) = 0$. By this mitigating conditions g and g' are obtained from the linear system

$$\begin{bmatrix} f' & f \\ f'' & 2f' \end{bmatrix} \begin{bmatrix} g \\ g' \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \text{and} \quad g = \frac{\det \begin{bmatrix} -1 & f \\ 0 & 2f' \end{bmatrix}}{\det \begin{bmatrix} f' & f \\ f'' & 2f' \end{bmatrix}} = \frac{-2f'}{2f'^2 - ff''}$$

where f, g and their derivatives are all evaluated at x_n. Now

$$x_{n+1} = x_n - \frac{2f'_n}{2f'^2_n - f_n f''_n} f_n$$

which is Halley's method.

To observe the cubic convergence Halley's method we take $f(x) = x^2 - 1$, for which $f' = 2x$ and $f'' = 2$, and we readily ascertain that here $x_{n+1} - 1 = (1/(1 + 3x_n^2))(x_n - 1)^3$, or $x_{n+1} - 1 = (1/4)(x_n - 1)^3$ if $x_n = 1$, nearly.

To prove the cubic convergence of Halley's method write it as $x_{n+1} = F(x_n)$ for $F(x) = x + g(x)f(x)$ and $g(x) = -2f/(2f'^2 - ff'')$. We verify that $g(a) = -1/f'(a)$ and $g'(a) = f''(a)/(2f'^2(a))$ for a such that $f(a) = 0$. It follows that $F'(a) = F''(a) = 0$, and we conclude that convergence is indeed cubic for $f(x)$ satisfying the hypotheses of the general fixed point iteration theorem.

11.6.4 Still higher order iterative methods

A, hopefully quartic, iterative method is created by letting $F'(x_n) = 0, F''(x_n) = 0$ and $F'''(x_n) = 0$. Differentiating $F(x) = x + gf$ thrice to have

$$F' = 1 + gf' + g'f, \quad F'' = gf'' + 2g'f' + g''f, \quad F''' = gf''' + 3g'f'' + 3g''f' + g'''f$$

we set $g''' = g'''(x_n) = 0$ and obtain from $F' = F'' = F''' = 0$ the linear system of three equations in three unknowns

$$\begin{bmatrix} f' & f & 0 \\ f'' & 2f' & f \\ f''' & 3f'' & 3f' \end{bmatrix} \begin{bmatrix} g \\ g' \\ g'' \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$$

for g, g', g''. This system is solved as

$$g = \frac{\det \begin{bmatrix} -1 & f & 0 \\ 0 & 2f' & f \\ 0 & 3f'' & 3f' \end{bmatrix}}{\det \begin{bmatrix} f' & f & 0 \\ f'' & 2f' & f \\ f''' & 3f'' & 3f' \end{bmatrix}} = -\frac{6f'^2 - 3ff''}{6f'^3 - 6ff'f'' + f^2 f'''}$$

and our proposed, hopefully quartic, iterative method, becomes

$$x_{n+1} = x_n - \frac{6f_n'^2 - 3f_n f_n''}{6f_n'^3 - 6f_n f_n' f_n'' + f_n^2 f_n'''} f_n$$

implicit in a formula of Householder [3]. To observe the order of convergence of this method we select the function $f(x) = x^2 - 1$ for which $f'(x) = 2x, f''(x) = 2, f'''(x) = 0$, and determine that $x_{n+1} - 1 = (1/(4x_n^3 + 4x_n))(x_n - 1)^4$, or $x_{n+1} - 1 = (1/8)(x_n - 1)^4$, nearly, if $x_n = 1$, nearly.

To prove the quartic convergence of this method we write it as $x_{n+1} = F(x_n)$ for $F(x) = x + g(x)f(x)$ and $g(x) = (-6f'^2 + 3ff'')/(6f'^3 - 6ff'f'' + f^2 f''')$, where f is short for $f(x)$. We verify (using Mathematica) that $g(a) = -1/f'$, $g'(a) = f''/(2f'^2)$ and $g''(a) = -f''^2/(2f'^3) + f'''/(3f'^2)$ where f', f'', f''' are short for $f'(a), f''(a)', f'''(a)$. It readily results that $F'(a) = F''(a) = F'''(a) = 0$, proving that convergence is indeed quartic for $f(x)$ satisfying the hypotheses of the general fixed point iteration theorem.

11.6.5 Direct derivation of high order iterative methods from the general fixed point iteration theorem

For $f(x) = x^2 - \alpha = 0,\ a = \sqrt{\alpha}$ the NR method is $x_{n+1} = x_n - 1/(2x_n)f_n,\ f_n = f(x_n)$. For $\alpha = 2$ it becomes the ubiquitous recursion $x_{n+1} = x_n/2 + 2/x_n$, and $x_{n+1} - \sqrt{2} = (1/2x_n)(x_n - \sqrt{2})^2$ or $x_{n+1} - \sqrt{2} = (\sqrt{2}/4)(x_n - \sqrt{2})^2$, nearly, if x_n is close to $\sqrt{2}$, implying that convergence is quadratic, and from above. We suggest to reform the NR method, writing it as $x_{n+1} = x_n - x_n(x_n^2 - 2)/(2x_n^2)$ and set $x_n^2 = 2$ so as to have $x_{n+1} = (x_n/4)(6 - x_n^2)$. Division by x_n is thereby replaced by a multiplication. Some algebra leads to $x_{n+1} - \sqrt{2} = -(1/4)(x_n + 2\sqrt{2})(x_n - \sqrt{2})^2$ or $x_{n+1} - \sqrt{2} = (-3\sqrt{2}/4)(x_n - \sqrt{2})^2$, nearly, if x_n is close to $\sqrt{2}$, implying that convergence is quadratic, and from below. Yet we notice that the factor $-3\sqrt{2}/4$ is three times bigger in magnitude then the corresponding factor in the unaltered NR method.

It occurs to us now that since the two methods converge from opposite directions their average weighted at the ratio of 3/4 to 1/4 should do better. Taking

$$x_{n+1} = \frac{3}{4}\frac{x_n^2 + 2}{2x_n} + \frac{1}{4}x_n(6 - x_n^2)$$

we obtain

$$x_{n+1} = -\frac{1}{16x_n}(x_n^4 - 12x_n^2 - 12)$$

or

$$x_{n+1} - \sqrt{2} = -\frac{x_n + 3\sqrt{2}}{16x_n}(x_n - \sqrt{2})^3$$

demonstrating that the we have constructed in this way a cubic method. If $x_n = \sqrt{2}$, nearly, then $x_{n+1} - \sqrt{2} = -(1/4)(x_n - \sqrt{2})^3$, nearly, implying that the error of this cubic method alternates in sign.

To directly obtain a quartic method for $f(x) = x^2 - \alpha = 0,\ a = \sqrt{\alpha}$ from the general fixed point iteration theorem we propose to write $f(x) = 0$ as $x = x + g(x)f(x)$, or shortly

$x = F(x)$, for $g = x^{-1}(A + Bx^2 + Cx^4)$, and fix constants A, B, C so that $F'(a) = F''(a) = F'''(a) = 0$. The special choice of weight function g is designed to assure the explicit dependence of A, B, C on α but not on a. In fact,

$$A = -\frac{5}{16},\ B = -\frac{1}{4\alpha},\ \text{and } C = \frac{B}{4\alpha} = -\frac{1}{16\alpha^2}.$$

For the choice

$$g(x) = \frac{1}{x}\frac{A + Bx^2}{1 + Cx^2}$$

we find

$$A = -\frac{1}{4},\ B = -\frac{3}{4\alpha},\ \text{and } C = \frac{1}{\alpha}.$$

To numerically compare the NR method to the quartic methods we select $\alpha = 2$ and $x_0 = 1$. For this values the NR method converges in five steps, but the quartic in three. Of course, the computational efficiency of the different methods must also be taken into consideration, but it is not for now.

For

$$g(x) = x^{-1}(A + Bx^2 + Cx^4 + Dx^6)$$

we obtain

$$A = -\frac{35}{128},\ B = -\frac{47}{128\alpha},\ C = \frac{23}{128\alpha^2}\ \text{ and }\ D = -\frac{5}{128\alpha^3}.$$

For

$$g(x) = \frac{1}{x}\,\frac{A + Bx^2 + Cx^4}{1 + Dx^2 + Ex^4}$$

we obtain

$$a = -\frac{1}{6},\ B = -\frac{5}{3\alpha},\ C = -\frac{5}{6\alpha^2},\ D = \frac{10}{3\alpha}\ \text{ and }\ E = \frac{1}{\alpha^2}.$$

Convergence with these last two methods is in two steps.

For computing the root of $f(x) = e^x - \alpha$ we propose the weight function $g(x) = A + Be^x + Ce^{2x}$ and fix the constants A, B and C so that $F'(a) = F''(a) = F'''(a) = 0$. The success of this choice of $g(x)$ hinges on the fact that A, B and C depend on α but are independent of the root a of $f(x) = 0$, here $a = \ln\alpha$. Repeatedly differentiating $F(x) = x + g(x)f(x)$ we have $F' = 1 + g'f + gf'$, $F'' = g''f + 2g'f' + gf''$, and $F''' = g'''f + 3g''f' + 3g'f'' + gf''$, from which we obtain by means of some simple algebra

$$A = -\frac{11}{6\alpha},\ B = \frac{7}{6\alpha^2},\ C = -\frac{1}{3\alpha^3}$$

and forthwith the quartic method

$$x_{n+1} = x_n + (-\frac{11}{6\alpha} + \frac{7}{6\alpha^2}e_n - \frac{1}{3\alpha^3}e_n^2)(e_n - \alpha),\ e_n = e^{x_n}$$

that requires the computation of e^x in each iterative cycle only once.

Chapter 11

For the rational

$$x_{n+1} = x_n + \frac{A + Be_n}{1 + Ce_n}(e_n - \alpha),\ e_n = e^{x_n}$$

we have

$$A = -\frac{5}{2\alpha},\ B = -\frac{1}{2\alpha^2},\ C = \frac{2}{\alpha}.$$

To find the root of $f(x) = \ln x - \alpha$ we propose

$$x_{n+1} = x_n + x_n(A + B\ln x_n + C\ln^2 x_n)(\ln x_n - \alpha)$$

but decide to take $C = 0$ and find $A = -1 - \alpha/2,\ B = 1/2$. Halley's method applied to this function yields the recursion

$$x_{n+1} = x_n - \frac{2x_n}{2 - \alpha + l_n}(l_n - \alpha),\ l_n = \ln x_n.$$

Correspondingly, we propose

$$x_{n+1} = x_n + x_n\frac{A + Bl_n}{1 + Cl_n}(l_n - \alpha),\ l_n = \ln x_n$$

and find

$$A = \frac{6 + \alpha}{-6 + 2\alpha},\ B = \frac{1}{6 - 2\alpha},\ C = \frac{1}{3 - \alpha}.$$

To numerically compare the two methods we choose $\alpha = 0.5$, so that $a = \sqrt{e}$, and take as starting value $x_0 = 1$. Once more, Halley's method converged in three steps and our rational method converged in two steps.

Once a good program is available for the evaluation of $\sin x$ and $\cos x,\ 0 < x < \pi/2$, the inverse trigonometric function $\arcsin x$ can be obtained as the solution of $\sin x - \alpha = 0$. For a cubic iterative solution method we propose to take $g(x) = A + B\cos x$, and we ascertain that $F(a) = F''(a) = 0$ if $A = (3\alpha^2 - 2)/(2\beta^3)$ and $B = -\alpha/(2\beta^3)$, where $\beta = \sqrt{1 - \alpha^2}$. To have a quartic method we suggest $g(x) = (A + B\cos x)/(1 + C\cos x)$ and ascertain that $F(a) = F''(a) = F'''(a) = 0$ if

$$A = \frac{\beta}{1 + 2\alpha^2} - \frac{2}{\beta},\ B = \frac{1}{1 + 2\alpha^2},\ C = \frac{-2\beta}{1 + 2\alpha^2} + \frac{1}{\beta},\ \text{where } \beta = \sqrt{1 - \alpha^2}.$$

Halley's method is here

$$x_{n+1} = x_n - \frac{2c_n}{c_n^2 - \alpha s_n}(s_n - \alpha) \text{ where } c_n = \cos x_n, \text{ and } s_n = \sin x_n.$$

We numerically compare the two methods by choosing $\alpha = 0.5$, so that $a = \arcsin(0.5) = \pi/6$, and taking as starting value $x_0 = 1$. Using high precision computation to suppress the ill effect of arithmetical round-off, we have Halley's method converge in four steps, and our rational method in only two.