plane and the surface.

Figure 13 shows more detail about the tangent plane. The dotted lines are the x and y tangent lines. They lie in the plane. All tangent lines lie in the tangent plane! These particular lines are tangent to the «partial functions», where y is fixed at $y_0 = 2$ or x is fixed at $x_0 = 1$. The plane is balancing on the surface and touching at the tangent point.

More is true. In the surface, every curve through the point is tangent to the plane. Geometrically, the curve goes up to the point and «kisses»the plane[1]. The tangent $\boldsymbol{T}$ to the curve and the normal $\boldsymbol{N}$ to the surface are perpendicular: $\boldsymbol{T} \cdot \boldsymbol{N} = 0$.

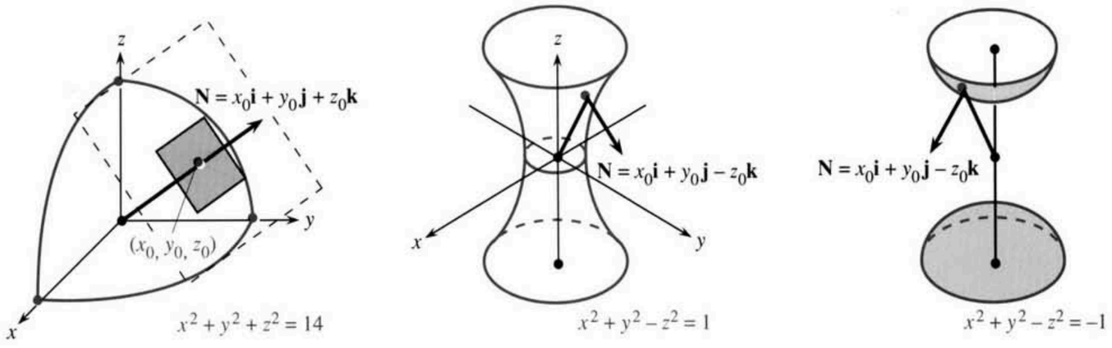

Figure 14: Tangent plane and normal $\boldsymbol{N}$ for a sphere. Hyperboloids of 1 and 2 sheets.

Exemple 20. Find the tangent plane to the sphere $z^2 = 14 - x^2 - y^2$ at $(1, 2, 3)$.

Answer 20 Instead of $z^2 = 14 - x^2 - y^2$ we have $z = \sqrt{14 - x^2 - y^2}$. At $x_0 = 1$, $y_0 = 2$ the height is now $z_0 = 3$. The surface is a sphere with radius $\sqrt{14}$. The only trouble from the square root is its derivatives:

$$\frac{\partial z}{\partial x} = \frac{\partial}{\partial x}\sqrt{14 - x^2 - y^2} = \frac{\frac{1}{2}(-2x)}{\sqrt{14 - x^2 - y^2}} \qquad \text{and} \qquad \frac{\partial z}{\partial y} = \frac{\frac{1}{2}(-2y)}{\sqrt{14 - x^2 - y^2}} \tag{5.3}$$

At $(1, 2)$ those slopes are $-1/3$ and $-2/3$. The equation of the tangent plane is linear: $z - 3 = -1/3(x - 1) - 2/3(y - 2)$. I cannot resist improving the equation, by multiplying through by 3 and moving all terms to the left side:

$$\textit{tangent plane to sphere:} \quad 1(x - 1) + 2(y - 2) + 3(z - 3) = 0. \tag{5.4}$$

If mathematics is the «science of patterns», equation (5.4) is a prime candidate for study. The numbers $1, 2, 3$ appear twice. The coordinates are $(x_0, y_0, z_0) = (1, 2, 3)$. The normal vector is $1\boldsymbol{i} + 2\boldsymbol{j} + 3\boldsymbol{k}$. The tangent equation is $1x + 2y + 3z = 14$. None of this can be an accident, but the square root of $14 - x^2 - y^2$ made a simple pattern look complicated.

This square root is not necessary. Calculus offers a direct way to find dz/dx - *implicit differentiation*. Just differentiate every term as it stands:

$$x^2 + y^2 + z^2 = 14 \qquad \text{leads to} \qquad 2x + 2z\,\partial z/\partial x = 0 \qquad \text{and} \qquad 2y + 2z\,\partial z/\partial y = 0 \tag{5.5}$$

[1]A safer word is «osculate». At saddle points the plane is kissed from both sides.

Canceling the 2's, the derivatives on a sphere are $-x/z$ and $-y/z$. Those are the same as in (5.3). The equation for the tangent plane has an extremely symmetric form:

$$z - z_0 = -\frac{x_0}{z_0}(x - x_0) - \frac{y_0}{z_0}(y - y_0) \qquad \text{or} \qquad x_0(x - x_0) + y_0(y - y_0) + z_0(z - z_0) = 0 \quad (5.6)$$

Reading off $\boldsymbol{N} = x_0\boldsymbol{i} + y_0\boldsymbol{j} + z_0\boldsymbol{k}$ from the last equation, calculus proves something we already knew: **The normal vector to a sphere points outward along the radius**.

5.1 The tangent plane to $F(x, y, z) = c$

The sphere suggests a question that is important for other surfaces. Suppose the equation is $F(x, y, z) = c$ instead of $z = f(x, y)$. Can the partial derivatives and tangent plane be found directly from F ? The answer is yes. It is not necessary to solve first for z. The derivatives of F, computed at (x_0, u_0, z_0), give a second formula for the tangent plane and normal vector.

The tangent plane to the surface $F(x, y, z) = c$ has the linear equation

$$\left(\frac{\partial F}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial F}{\partial y}\right)_0 (y - y_0) + \left(\frac{\partial F}{\partial z}\right)_0 (z - z_0) = 0 \qquad (5.7)$$

The normal vector is $\boldsymbol{N} = (\partial F/\partial x)_0\, \boldsymbol{i} + (\partial F/\partial y)_0\, \boldsymbol{j} + (\partial F/\partial z)_0\, \boldsymbol{k}$.

Notice how this includes the original case $z = f(x, y)$. **The function F becomes** $f(x, y) - z$. Its partial derivatives are $\partial f/\partial x$ and $\partial f/\partial y$ and -1 (the -1 is from the derivative of $-z$). Then equation (5.7) is the same as our original tangent equation (5.1).

Exemple 21. The surface $F = x^2 + y^2 - z^2 = c$ is a **hyperboloid**. Find its tangent plane.

Answer 21 The partial derivatives are $F_x = 2x, F_y = 2y, F_z = -2z$. Equation (5.7) is

$$\textit{tangent plane :} \qquad 2x_0(x - x_0) + 2y_0(y - y_0) - 2z_0(z - z_0) = 0 \qquad (5.8)$$

We can cancel the 2's. The normal vector is $\boldsymbol{N} = x_0\, \boldsymbol{i} + y_0\, \boldsymbol{j} - z_0\, \boldsymbol{k}$. For $c > 0$ this hyperboloid has one sheet (**Figure 14**). For $c = 0$ it is a cone and for $c < 0$ it breaks into two sheets.

5.2 Differentials

Come back to the linear equation $z - z_0 = \left(\frac{\partial f}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y - y_0)$ for the tangent plane. That may be the most important formula in this document. Move along the tangent plane instead of the curved surface. Movements in the plane are dx and dy and dz, while Δx and Δy and Δz are movements in the surface. The d's are governed by the tangent equation - the Δ's are governed by $z = f(x, y)$. Usually the d's were **differentials** along the tangent line:

$$dy = \frac{dy}{dx} dx \ \ \text{(straight line)} \qquad \text{and} \qquad df \approx \frac{dy}{dx} \Delta x \ \ \text{(on the curve)} \qquad (5.9)$$

Now y is independent like x. The dependent variable is z. The idea is the same. The distances $x - x_0$ and $y - y_0$ and $z - z_0$ (on the tangent plane) are dx and dy and dz. The equation of the plane is

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy \qquad \text{or} \qquad df = f_x dx + f_y dy \qquad (5.10)$$

This is the **total differential**. All letters dz and df and dw can be used, but ∂z and ∂f are not used. Differentials suggest small movements in x and y; then dz is the resulting movement in z.

On the tangent plane, equation (5.10) holds exactly.

A «centering transform» has put x_0, y_0, z_0 at the center of coordinates. Then the »zoom transform» stretches the surface into its tangent plane.

Exemple 22. The area of a triangle is $A = \frac{1}{2}ab\sin\theta$. Find the total differential dA.

Answer 22 The base has length b and the sloping side has length a. The angle between them is θ. You may prefer $A = \frac{1}{2}bh$, where h is the perpendicular height a $\sin\theta$. Either way we need the partial derivatives. If $A = \frac{1}{2}ab\sin\theta$, then

$$\frac{\partial A}{\partial a} = \frac{1}{2}\, b\, \sin\theta \qquad \frac{\partial A}{\partial b} = \frac{1}{2}\, a\, \sin\theta \qquad \frac{\partial A}{\partial \theta} = \frac{1}{2}\, ab\, \cos\theta \tag{5.11}$$

These lead immediately to the total differential dA (like a product rule):

$$dA = \frac{\partial A}{\partial a}\, da + \frac{\partial A}{\partial b}\, db + \frac{\partial A}{\partial \theta}\, d\theta = \frac{1}{2}\, b\, \sin\theta\, da + \frac{1}{2}\, a\, \sin\theta\, db + \frac{1}{2}\, ab\, \cos\theta\, d\theta \tag{5.12}$$

Exemple 23. The volume o fa cylinder is $V = \pi r^2 h$. Decide whether V is more sensitive to a change from $r = 1.0$ to $r = 1.1$ or from $h = 1.0$ to $h = 1.1$.

Answer 23 The partial derivatives are $\partial V/\partial r = 2\pi r h$ and $\partial V/\partial h = \pi r^2$. **They measure the sensitivity to change**. Physically, they are the side area and base area of the cylinder. The volume differential dV comes from a shell around the side plus a layer on top:

$$dV = \text{shell} + \text{layer} = 2\pi r h\, dr + \pi r^2\, dh \tag{5.13}$$

Starting from $r = h = 1$, that differential is $dV = 2\pi\, dr + \pi\, dh$. With $dr = dh = 0.1$, the shell volume is $0.2 \times \pi$ and the layer volume is only $0.1 \times \pi$. So V is sensitive to dr.

For a short cylinder like a penny, the layer has greater volume. V is more sensitive to dh. In our case $V = \pi r^2 h$ increases from $\pi(1)^3$ to $\pi(1.1)^3$. **Compare ΔV to dV** :

$$\Delta V = \pi(1.1)^3 - \pi(1)^3 = 0.331 \times \pi \qquad \text{and} \qquad dV = 2\pi(0.1) + \pi(0.1) = (0.300)\pi \tag{5.14}$$

The difference is $\Delta V - dV = 0.031\,\pi$. The shell and layer missed a small volume in **Figure 15 (left)**, just above the shell and around the layer. The mistake is of order $(dr)^2 + (dh)^2$. For $V = \pi r^2 h$, the differential $dV = 2\pi r h\, dr + \pi r^2\, dh$ is a **linear approximation** to the true change ΔV. We now explain that properly.

5.3 Linear approximation

Tangents lead immediately to linear approximations. That is true of tangent planes as it was of tangent lines. The plane stays close to the surface, as the line stayed close to the curve. Linear functions are simpler than $f(x)$ or $f(x, y)$ or $F(x, y, z)$. All we need are first derivatives *at the point*. Then the approximation is good *near the point*.

This key idea of calculus is already present in differentials. On the plane, df equals $f_x\, dx + f_y\, dy$. On the curved surface that is a linear approximation to Δf :

The linear approximation to $f(x, y)$ near the point (x_0, y_0) is

$$f(x, y) \approx f(x_0, y_0) + \left(\frac{\partial f}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y - y_0) \tag{5.15}$$

In other words $\Delta f \approx f_x \Delta x + f_y \Delta y$.The right side of (5.15) is a linear function $f_x(x, y)$. At (x_0, y_0), the functions f and f_L have the same slopes. Then $f(x, y)$ curves away from f, with an error of « second order »:

$$|f(x,y) - f_L(x,y)| \leq M\left[(x - x_0)^2 + (y - y_0)^2\right] \tag{5.16}$$

This assumes that f_{xx}, f_{xy}, and f_{yy} are continuous and bounded by M along the line from (x_0, y_0) to (x, y). Example 3 of Section 13.5 shows that $|f_{tt}| \ll 2M$ along that line. A factor 1/2 comes from equation 3.8.12, for the error $f - f_L$, with one variable.

For the volume of a cylinder, r and h went from 1.0 to 1.1. The second derivatives of $V = \pi r^2 h$ are $V_{rr} = 2\pi h$ and $V_{rh} = 2\pi r$ and $V_{hh} = 0$. They are below $M = 2.2\pi$. Then (5.16) gives the error bound $2.2\pi(0.1^2 + 0.1^2) = 0.044\pi$, not far above the actual error 0.031π. The main point is that **the error in linear approximation comes from the quadratic terms**, those are the first terms to be ignored by f_L.

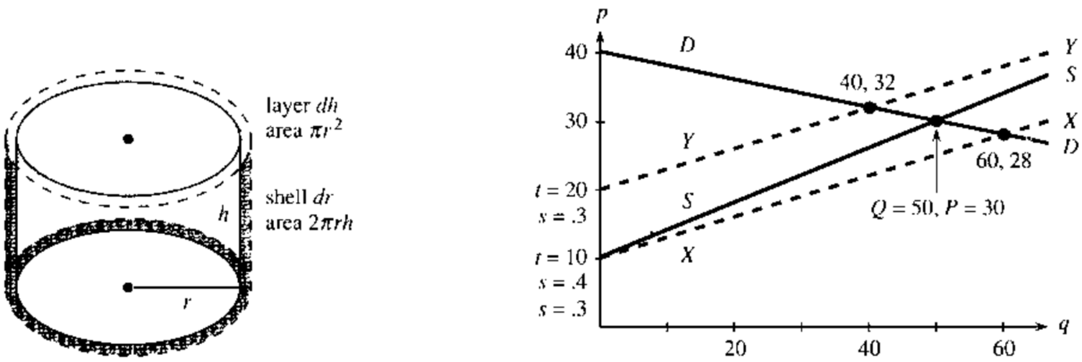

Figure 15: *Left* : Shell plus layer gives $dV = 0.300\pi$. Including top rings gives $\Delta V = 0.311\pi$. *Right* : Quantity Q and price P mov with the lines.

Exemple 24. Find a-linear approximation to the distance function $r = \sqrt{x^2 + y^2}$

Answer 24 The partial derivatives are x/r and y/r. Then $\Delta r \approx (x/r)\Delta x + (y/r)\Delta y$. For (x, y, r) near $(1, 2, \sqrt{5})$:

$$\sqrt{x^2 + y^2} \approx \sqrt{1^2 + 2^2} + (x - 1)/\sqrt{5} + 2(y - 2)/\sqrt{5} \tag{5.17}$$

If y is fixed at 2, this is a one-variable approximation to $\sqrt{x^2 + 2^2}$. If x is fixed at 1, it is a linear approximation in y. Moving both variables, you might think dr would involve dx and dy in a square root. It doesn't. Distance involves x and y in a square root, but *change of distance* is linear in Δx and Δy, to a first approximation.

There is a rough point at $x = 0, y = 0$. Any movement from $(0, 0)$ gives $\Delta r = \sqrt{(\Delta x)^2 + (\Delta y)^2}$. The square root has returned. The reason is that **the partial derivatives** x/r **and** y/r **are not continuous at** $(0, 0)$. The cone has a sharp point with no tangent plane. *Linear approximation breaks down.*

The next example shows how to approximate Δz from Δx and Δy, when the equation is $F(x, y, z) = c$. We use the implicit derivatives in (5.7) instead of the explicit derivatives in (5.1).

The idea is the same: Look at the tangent equation as a way to find Δz, instead of an equation for z. Here is an example with new letters.

Exemple 25. From $F = -x^2 - y^2 + z^2 = 0$ find a linear approximation to Δz.

Answer 25 (implicit derivatives) Use the derivatives of F : $-2x\,\Delta x - 2y\,\Delta y + 2z\,\Delta z \approx 0$. Then solve for Δz, which gives $\Delta z \approx (x/z)\Delta x + (y/z)\Delta y$, the same as in **Example 24**.

Exemple 26. How does the equilibrium price change when the supply curve changes?

Answer 26 The equilibrium price is at the intersection of the supply and demand curves (**supply = demand**). As the price p rises, the demand q drops (the slope is .2):

$$\text{demand line } DD: \quad p = -0.2\,q + 40 \tag{5.18}$$

The supply (also q) goes up with the price. The slope s is positive (here $s = 0.4$) :

$$\text{supply line } SS: \quad p = s\,q + t = 0.4\,q + 10 \tag{5.19}$$

Those lines are in **Figure 15**. They meet at the **equilibrium price** $P = \$30$. The quantity $Q = 50$ is available at P (on SS) and demanded at P (on DD). So it is sold.

Where do partial derivatives come in ? The reality is that those lines DD and SS are not fixed for all time. Technology changes, and competition changes, and the value of money changes. Therefore the lines move. Therefore the crossing point (Q, P) also moves. Please recognize that derivatives are hiding in those sentences.

Main point: **The equilibrium price P is a function of s and t.** Reducing s by better technology lowers the supply line to $p = 0.3\,q + 10$. The demand line has not changed. The customer is as eager or stingy as ever. But the price P and quantity Q are different. The new equilibrium is at $Q = 60$ and $P = \$28$, where the new line XX crosses DD.

If the technology is expensive, the supplier will raise t when reducing s. Line YY is $p = 0.3\,q{+}20$. That gives a higher equilibrium $P = \$32$ at a lower quantity $Q = 40$, the demand was too weak for the technology.

Calculus question Find $\partial P/\partial s$ and $\partial P/\partial t$. The difficulty is that P is not given as a function of s and t. So take implicit derivatives of the supply = demand equations :

$$\begin{array}{rlll} \text{supply} = \text{demand} : & P = -0.2\,Q + 40 & = sQ + t & \\ s\text{-derivative} : & P_s = -0.2\,Q_s & = sQ_s + Q & (\text{we note that } t_s = 0) \\ t\text{-derivative} : & P_t = -0/2\,Q_t & = sQ_t + 1 & (\text{we note that } t_t = 1) \end{array} \tag{5.20}$$

Now substitute $s = 0.4$, $t = 10$, $P = 30$, $Q = 50$. That is the starting point, around which we are finding a linear approximation. The last two equations give $P_s = 50/3$ and $P_t = 1/3$ (see **Exercise 15**). The linear approximation is

$$P = 30 + 50(s - 0.4)/3 + (t - 10)/3 \tag{5.21}$$

Comment This example turned out to be subtle (so is economics). I hesitated before including it. The equations are linear and their derivatives are easy, but something in the problem is hard, there is no explicit formula for P. The function $P(s,t)$ is not known. Instead of a point on a surface, we are following the intersection of two lines. *The solution changes as the equation changes.* **The**

derivative of the solution comes from the derivative of the equation.

Summary The foundation of this section is equation (5.1) for the tangent plane. Everything builds on that, total differential, linear approximation, sensitivity to small change. Later sections go on to the chain rule and « directional derivatives » and « gradients ». The central idea of differential calculus is of $\Delta f \approx f_x \Delta x + f_y \Delta y$.

5.4 Newton's method for two equations

Linear approximation is used **to solve equations**. To find out where a function is zero, look first to see where its approximation is zero. To find out where a graph crosses the xy plane, look to see where its tangent plane crosses.

Remember Newton's method for $f(x) = 0$. The current guess is x_n. Around that point, $f(x)$ is close to $f(x_n) + (x - x_n)f'(x_n)$. This is zero at the next guess $x_n = x_n - f(x_n)/f'(x_n)$. That is where the tangent line crosses the x axis.

With two variables the idea is the same, but two unknowns x and y require *two equations*. We solve $g(x, y) = 0$ and $h(x, y) = 0$. Both functions have linear approximations that start from the current point (x_n, y_n), where derivatives are computed:

$$g(x,y) \approx g(x_n, y_n) + \left(\frac{\partial g}{\partial x}\right)(x - x_n) + \left(\frac{\partial g}{\partial y}\right)(y - y_n) \tag{5.22}$$

$$h(x,y) \approx h(x_n, y_n) + \left(\frac{\partial h}{\partial x}\right)(x - x_n) + \left(\frac{\partial h}{\partial y}\right)(y - y_n) \tag{5.23}$$

The natural idea is to **set these approximations to zero**. That gives linear equations for $x - x_n$ and $y - y_n$. Those are the steps Δx and Δy that take us to the next guess in Newton's method :

> Newton's method to solve $g(x, y) = 0$ and $h(x, y) = 0$ has linear equations for the steps Δx and Δy that go from (x_n, y_n) to (x_{n+1}, y_{n+1}) :
>
> $$\left(\frac{\partial g}{\partial x}\right)\Delta x + \left(\frac{\partial g}{\partial y}\right)\Delta y = -g(x_n, y_n) \quad \text{and} \quad \left(\frac{\partial h}{\partial x}\right)\Delta x + \left(\frac{\partial h}{\partial y}\right)\Delta y = -h(x_n, y_n) \tag{5.24}$$

Exemple 27. The equations $g = x^3 - y = 0$ and $h = y^3 - x = 0$ have 3 solutions $(1, 1)$, $(0, 0)$, $(-1, -1)$. I will start at different points (x_0, y_0). The next guess is $x_1 = x_0 + \Delta x$, $y_1 = y_0 + \Delta y$. It is of extreme interest to know which solution Newton's method will choose, if it converges at all. I made three small experiments.

1. Suppose $(x0, y_0) = (2, 1)$. At that point $g = 2^3 - 1 = 7$ and $h = 1^3 - 2 = -1$. The derivatives are $g_x = 3x^2 = 12$, $g_y = -1$, $h_x = -1$, $h_y = 3y^2 = 3$. The steps Δx and Δy come from solving (5.24) :

   $$\begin{cases} 12\Delta x - \Delta y = -7 \\ -\Delta x + 3\Delta y = +1 \end{cases} \quad \Rightarrow \quad \begin{cases} \Delta x = -4/7 \\ \Delta y = -1/7 \end{cases} \quad \Rightarrow \quad \begin{cases} x_1 = x_0 + \Delta x = 10/7 \\ y_1 = y_0 + \Delta y = 8/7 \end{cases}$$

   This new point $(10/7, 8/7)$ is closer to the solution at $(1, 1)$. The next point is $(1.1, 1.05)$ and convergence is clear. Soon convergence is fast.

2. Start at $(x_0, y_0) = (1/2, 0)$. There we find $g = 1/8$ and $h = -1/2$:

$$\begin{cases} (3/4)\Delta x - \Delta y = -1/8 \\ -\Delta x + 0\Delta y = +1/2 \end{cases} \Rightarrow \begin{cases} \Delta x = -1/2 \\ \Delta y = +1/4 \end{cases} \Rightarrow \begin{cases} x_1 = x_0 + \Delta x = 0 \\ y_1 = y_0 + \Delta y = -1/4 \end{cases}$$

   Newton has jumped from $(1/2, 0)$ on the x axis to $(0, -1/4)$ on the y axis. The next step goes to $(1/32, 0)$, back on the x axis. We are in the « basin of attraction »of $(0, 0)$.

3. Now start further out the axis at $(1, 0)$, where $g = 1$ and $h = -1$:

$$\begin{cases} 3\Delta x - \Delta y = -1 \\ -\Delta x + 0\Delta y = +1 \end{cases} \Rightarrow \begin{cases} \Delta x = -1 \\ \Delta y = -2 \end{cases} \Rightarrow \begin{cases} x_1 = x_0 + \Delta x = 0 \\ y_1 = y_0 + \Delta y = -2 \end{cases}$$

   Newton moves from $(1, 0)$ to $(0, -2)$ to $(16, 0)$. Convergence breaks down, the method blows up. This danger is ever-present, when we start far from a solution.

Please recognize that even a small computer will uncover amazing patterns. It can start from hundreds of points (x_0, y_0), and follow Newton's method. Each solution has a **basin of attraction**, containing all (x_0, y_0) leading to that solution. There is also a basin leading to infinity. The basins in **Figure 16** are completely mixed together, a color figure shows them as **fractals**. The most extreme behavior is on the borderline between basins, when Newton can't decide which way to go. Frequently we see chaos.

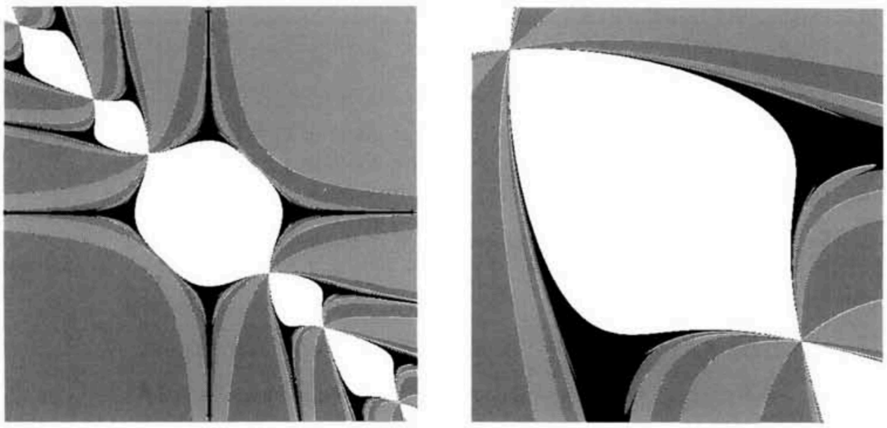

Figure 16: The basins of attraction to $(1, 1)$, $(0, 0)$, $(-1, -1)$, and infinity.

Chaos is irregular movement that follows a definite rule. Newton's method deter-mines an iteration from each point (x_n, y_0) to the next. In scientific problems it normally converges to the solution we want. (We start close enough.) But the computer makes it possible to study iterations from faraway points. This has created a new part of mathematics-so new that any experiments you do are likely to be original.

You can found chaos when trying to solve $x^2 + 1 = 0$. But don't think Newton's method is a failure. On the contrary, it is the best method to solve nonlinear equations. The error is squared as the algorithm converges, because linear approximations have errors of order $(\Delta x)^2 + (\Delta y)^2$. Each step doubles the number of correct digits, near the solution. The example shows why it is important to be near.

6 Directional Derivatives and Gradients

As x changes, we know how $f(x, y)$ changes. The partial derivative of ex treats y as constant. Similarly $\partial f/\partial y$ keeps x constant, and gives the slope in the y direction. But east-west and north-south are not the only directions to move. We could go along a 45° line, where $\Delta x = \Delta y$. In principle, before we draw axes, no direction is preferred. The graph is a surface with slopes in *all* directions.

On that surface, calculus looks for the rate of change (or the slope). There is a **directional derivative**, whatever the direction. In the 45° case we are inclined to divide Δf by Δx, but we would be wrong.

Let me state the problem. We are given $f(x, y)$ around a point $P = (x_0, y_0)$. We are also given a direction $\boldsymbol{u}$ (a unit vector). There must be a natural definition of $D_{\boldsymbol{u}}f$, the derivative off in the direction $\boldsymbol{u}$. To compute this slope at P, we need a formula. Preferably the formula is based on $\partial f/\partial x$ and $\partial f/\partial y$, which we already know.

Note that the 45° direction has $\boldsymbol{u} = \boldsymbol{i}/\sqrt{2} + \boldsymbol{j}/\sqrt{2}$. The square root of 2 is going to enter the derivative. This shows that dividing Δf by Δx is wrong. We should divide by the step length Δs.

Exemple 28. Stay on the surface $z = xy$. When (x, y) moves a distance Δs in the 45° direction from $(1, 1)$, what is $\Delta z/\Delta s$?

Answer 28 The step is As times the unit vector $\boldsymbol{u}$. Starting from $x = y = 1 thestependsat x = y = 1 + \Delta s/\sqrt{2}$. (The components of $\boldsymbol{u}\Delta s$ are $\Delta s/\sqrt{2}$.) Then $z = xy$ is

$$z = (1 + \Delta s/\sqrt{2})^2 = 1 + \sqrt{2}\,\Delta s + \frac{1}{2}\,(\Delta s)^2 \quad \text{which means} \quad \Delta s = \sqrt{2}\,\Delta s + \frac{1}{2}\,(\Delta s)^2 \tag{6.1}$$

The ratio $\Delta z/\Delta s$ approaches $\sqrt{2}$ as $\Delta s \to 0$. That is the slope in the 45° direction.

Définition 11. *The derivative off in the direction $\boldsymbol{u}$ at the point P is $D_{\boldsymbol{u}}f(P)$ and is given by :*

$$D_{\boldsymbol{u}}f(P) = \lim_{\Delta s \to 0} \frac{\Delta f}{\Delta s} = \lim_{\Delta s \to 0} \frac{f(P + \boldsymbol{u}\,\Delta s) - f(P)}{\Delta s} \tag{6.2}$$

The step from $P = (x_0, y_0)$ has length Δs. It takes us to $(x_0 + u_1\Delta s, y_0 + u_2\Delta s)$. We compute the change Δf and divide by Δs. But formula (6.3) below saves time.

The x direction is $\boldsymbol{u} = (1, 0)$. Then $\boldsymbol{u}\,\Delta s$ is $(\Delta s, 0)$ and we recover $\partial f/\partial x$:

$$\frac{\Delta f}{\Delta s} = \frac{f(x_0 + \Delta s, y_0) - f(x_0, y_0)}{\Delta s} \qquad \text{approaches} \qquad D_{(1,0)}f = \frac{\partial f}{\partial x} \tag{6.3}$$

Similarly $D_{\boldsymbol{u}} = \partial f/\partial y$, when $\boldsymbol{u} = (0, 1)$ is in the y direction. What is $D_{\boldsymbol{u}}f$ when $\boldsymbol{u} = (0, -1)$? That is the negative y direction, so $D_{\boldsymbol{u}}f = -\partial f/\partial y$.

6.1 Calculating the directional derivative

$D_{\boldsymbol{u}}f$ is the slope of the surface $z = f(x, y)$ in the direction $\boldsymbol{u}$. How do you compute it ? From $\partial f/\partial x$ and $\partial f/\partial y$, in two special directions, there is a quick way to find $D_{\boldsymbol{u}}f$ in all directions. **Remember that $\boldsymbol{u}$ is a unit vector**.

The directional derivative $D_{\boldsymbol{u}}$ in the direction $u = (u_1, u_2)$ equals :

$$D_{\boldsymbol{u}} = \frac{\partial f}{\partial x} u_1 + \frac{\partial f}{\partial y} u_2 \tag{6.4}$$

The reasoning goes back to the linear approximation of Δf : of

$$\Delta f \approx \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y = \frac{\partial f}{\partial x} u_1 \Delta s + \frac{\partial f}{\partial y} u_2 \Delta s \tag{6.5}$$

Divide by Δs and let Δs approach zero. Formula (6.4) is the limit of $\Delta f/\Delta s$, as the approximation becomes exact. A more careful argument guarantees this limit provided f_x and f_y are continuous at the basepoint (x_0, y_0). The main point is : **Slopes in all directions are known from slopes in two directions**.

Exemple 29. (repeated) We have $f = xy$ and $P = (1, 1)$ and $\boldsymbol{u} = (1/\sqrt{2}, 1/\sqrt{2})$. Find $D_{\boldsymbol{u}}(P)$. The derivatives $f_x = y$ and $f_y = x$ equal 1 at P. The 45° derivative is

$$D_{\boldsymbol{u}}(P) = f_x u_1 + f_y u_2 = 1 \times (1/\sqrt{2}) + 1 \times (1/\sqrt{2}) = \sqrt{2} \text{ as before} \tag{6.6}$$

Exemple 30. The linear function $f = 3x + y + 1$ has slope $D_{\boldsymbol{u}} f = 3u_1 + u_2$. The x direction is $u = (1, 0)$, and $D_{\boldsymbol{u}} f = 3$. That is $\partial f/\partial x$. In the y direction $D_{\boldsymbol{u}} f = 1$. Two other directions are special-along the level lines of $f(x, y)$ and perpendicular :

- Level direction : $D_{\boldsymbol{u}} f$ is zero because f is constant.
- Steepest direction : $D_{\boldsymbol{u}} f$ is as large as possible (with $u_1^2 + u_2^2 = 1$).

To find those directions, look at $D_{\boldsymbol{u}} f = 3u_1 + u_2$. The level direction has $3u_1 + u_2 = 0$. Then u is proportional to $(1, -3)$. Changing x by 1 and y by -3 produces no change in $f = 3x + y + 1$.

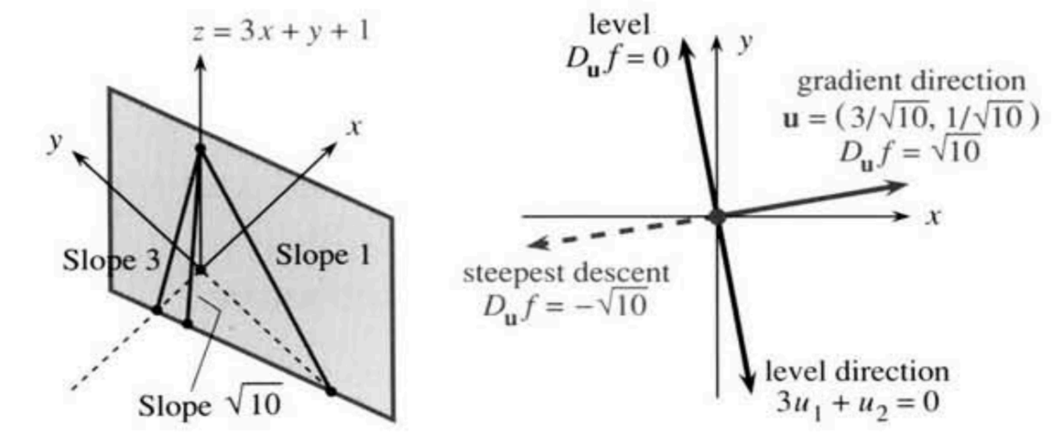

Figure 17: Steepest direction is along the gradient. Level direction is perpendicular.

In the steepest direction $\boldsymbol{u}$ is proportional to $(3, 1)$. Note the partial derivatives $f_x = 3$ and $f_y = 1$. The dot product of $(3, 1)$ and $(1, -3)$ is zero-steepest direction is perpendicular to level direction. To make $(3, 1)$ a unit vector, divide by $\sqrt{10}$.

- Steepest climb : $D_{\boldsymbol{u}} f = 3 \times (3/\sqrt{10}) + 1 \times (1/\sqrt{10}) = 10/\sqrt{10} = \sqrt{10}$.
- Steepest descent : Reverse to $\boldsymbol{u} = (-3/\sqrt{10}, -1/\sqrt{10})$ and $D_{\boldsymbol{u}} f = -\sqrt{10}$.

The contour lines around a mountain follow $D_{\boldsymbol{u}} f = 0$. The creeks are perpendicular. On a plane like $f = 3x + y + 1$, those directions stay the same at all points (**Figure 17**). On a mountain the steepest direction changes as the slopes change.

6.2 The gradient vector

Look again at $f_x u_1 + f_y u_2$, which is the directional derivative $D_{\boldsymbol{u}} f$. This is the dot product of two vectors. One vector is u = (ul, u2), which sets the direction. The other vector is (fx,f,), which comes from the function. This second vector is the gradient. of

Définition 12. *The* ***gradient*** *of $f(x, y)$ is the vector whose components are $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. We note :*

$$\boldsymbol{\nabla} f = \boldsymbol{\nabla} f = \frac{\partial f}{\partial x} \boldsymbol{i} + \frac{\partial f}{\partial y} \boldsymbol{j} \quad \text{(add } \frac{\partial f}{\partial z} \boldsymbol{k} \text{ in three dimensions)} \tag{6.7}$$

The space-saving symbol $\boldsymbol{\nabla}$ is read as « grad ». It can also be called « del ».

For the linear function $3x + y + 1$, the gradient is the constant vector $(3, 1)$. It is the way to climb the plane. For the nonlinear function $x^2 + xy$, the gradient is the non-constant vector $(2x + y, x)$. Notice that $\boldsymbol{\nabla} f$ shares the two derivatives in $\boldsymbol{N} = (f_x, f_y, -1)$. But the gradient is not the normal vector. $\boldsymbol{N}$ is in three dimensions, pointing away from the surface $z = f(x, y)$. **The gradient vector is in the xy plane !** The gradient tells which way on the surface is up, but it does that from down in the base.

The level curve is also in the xy plane, perpendicular to the gradient. The contour map is a projection on the base plane of what the hiker sees on the mountain. The vector $\boldsymbol{\nabla} f$ tells the **direction** of climb, and its length $|\boldsymbol{\nabla} f|$ gives the **steepness**.

> The directional derivative is $D_{\boldsymbol{u}} f = (\boldsymbol{\nabla} f) \cdot \boldsymbol{u}$. The level direction is perpendicular to $\boldsymbol{\nabla} f$, since $D_{\boldsymbol{u}} f = 0$. **The slope $D_{\boldsymbol{u}} f$ is largest when $\boldsymbol{u}$ is parallel to $\boldsymbol{\nabla} f$.** That maximum slope is the length $|\boldsymbol{\nabla} f| = \sqrt{f_x^2 + f_y^2}$:
>
> $$\text{For} \quad \boldsymbol{u} = \frac{\boldsymbol{\nabla} f}{|\boldsymbol{\nabla} f|} \quad \text{the slope is} \quad (\boldsymbol{\nabla} f) \cdot \boldsymbol{u} = \frac{|\boldsymbol{\nabla} f|^2}{|\boldsymbol{\nabla} f|} = \boldsymbol{\nabla} f$$

The example $f = 3x + y + 1$ had $\boldsymbol{\nabla} f = (3, 1)$. Its steepest slope was in the direction $\boldsymbol{u} = (3, 1)/\sqrt{10}$. The maximum slope was $\sqrt{10}$. That is $|\boldsymbol{\nabla} f| = \sqrt{9 + 1}$.

Important point : **The maximum of $(\boldsymbol{\nabla} f) \cdot \boldsymbol{u}$ is the length $|\boldsymbol{\nabla} f|$.** In nonlinear examples, the gradient and steepest direction and slope will vary. But look at one particular point in **Figure 18**. Near that point, and near any point, the linear picture takes over.

On the graph of f the special vectors are the level direction $\boldsymbol{L} = (f_x, -f_x, 0)$ and the uphill direction $\boldsymbol{U} = (f_x, f_y, f_x^2 + f_y^2)$ and the normal $\boldsymbol{N} = (f_x, f_y, -1)$. Problem 18 checks that those are perpendicular.

Exemple 31. The gradient of $f(x, y) = (14 - x^2 - y^2)/3$ is $\boldsymbol{\nabla} f = (-2x/3, -2y/3)$. On the surface, the normal vector is $\boldsymbol{N} = (-2x/3, -4y/3, -1)$. At the point $(1, 2, 3)$, this perpendicular is $\boldsymbol{N} = (-2/3, -4/3, -1)$. At the point (1, 2) down in the base, the gradient is $(-2/3, -4/3)$. The length of $\boldsymbol{\nabla} f$ is the slope $\sqrt{20}/3$.

Probably a hiker does not go straight up. A « grade » of $\sqrt{20}/3$ is fairly steep (almost 150%). To estimate the slope in other directions, measure the distance along the path between two contour lines. If $\Delta f = 1$ in a distance $\Delta s = 3$ the slope is about 1/3. This calculation is not exact until the limit of $\Delta f / \Delta s$, which is $D_{\boldsymbol{u}} f$.

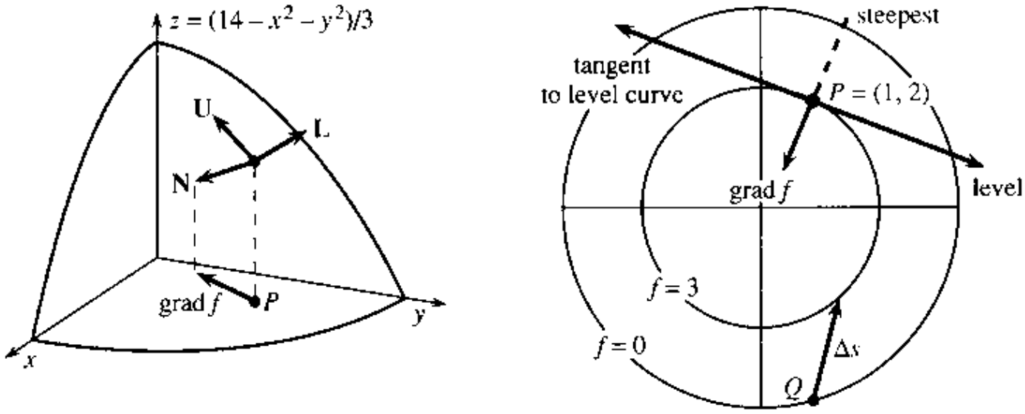

Figure 18: $\boldsymbol{N}$ perpendicular to surface and $\boldsymbol{\nabla} f$ perpendicular to level line (in the base).

Exemple 32. The gradient of $f(x, y, z) = xy + yz + xz$ has three components. The pattern extends from $f(x, y)$ to $f(x, y, z)$. The gradient is now the three-dimensional vector (f_x, f_y, f_z). For this function $\boldsymbol{\nabla} f$ is $(y + z, x + z, x + y)$. To draw the graph of $w = (x, y, z)$ would require a four-dimensional picture, with axes in the $xyzw$ directions.

Notice the dimensions. The graph is a 3-dimensional « surface » in 4-dimensional space. The gradient is down below in the 3-dimensional base. The level sets off come from $xy + yz + zx = c$, they are 2-dimensional. The gradient is perpendicular to that level set (still down in 3 dimensions). The gradient is not $\boldsymbol{N}$! The normal vector is $(f_x, f_y, f_z, -1)$, perpendicular to the surface up in 4-dimensional space.

Exemple 33. Find $\mathbf{grad}\, z$ when $z(x, y)$ is given implicitly : $F(x, y, z) = x^2 + y^2 - z^2 = 0$. In this case we find $z = \pm\sqrt{x^2 + y^2}$. The derivatives are $\pm x/\sqrt{x^2 + y^2}$ and $\pm y/\sqrt{x^2 + y^2}$, which go into $\mathbf{grad}\, z$. But the point is this : To find that gradient faster, differentiate $F(x, y, z)$ as it stands. Then divide by F_z :

$$F_x\, dx + F_y\, dy + F_z\, dz = 0 \qquad \text{or} \qquad dz = (-F_x\, dx - F_y\, dy)/F_z \tag{6.8}$$

The gradient is $(-F_x/F_z, -F_y/F_z)$. Those derivatives are evaluated at (x_0, y_0). The computation does not need the explicit function $z = f(x, y)$:

$$F = x^2 + y^2 - z^2 \quad \Rightarrow \quad F_x = 2x, F_y = 2y, F_z = -2z \quad \Rightarrow \quad \mathbf{grad}\, z = (x/z, y/z)$$

To go uphill on the cone, move in the direction $(x/z, y/z)$. That gradient direction goes radially outward. The steepness of the cone is the length of the gradient vector:

$$\textbf{grad}\, z = \sqrt{(x/z)^2 + (y/z)^2} = 1 \quad \text{because} \quad z^2 = x^2 + y^2 \quad \text{on the cone}$$

6.3 Derivatives along a curved paths

On a straight path the derivative of f is $D_{\boldsymbol{u}} f = (\textbf{grad} f) \cdot \boldsymbol{u}$. What is the derivative on a curved path ? **The path direction $\boldsymbol{u}$ is the tangent vector $\boldsymbol{T}$.** So replace $\boldsymbol{u}$ by $\boldsymbol{T}$, which gives the « direction » of the curve.

The path is given by the position vector $\boldsymbol{R}(t) = x(t)\,\boldsymbol{i} + y(t)\,\boldsymbol{j}$. The velocity is $\boldsymbol{v} = (dx/dt)\,\boldsymbol{i} + (dy/dt)\,\boldsymbol{j}$. The tangent vector is $\boldsymbol{T} = \boldsymbol{v}/|\boldsymbol{v}|$. Notice the choice, to move at any speed (with $\boldsymbol{v}$) or to go at unit speed (with $\boldsymbol{T}$). There is the same choice for the derivative of $f(x, y)$ along this curve :

$$\textbf{rate of change} = \frac{df}{dt} = (\textbf{grad} f) \cdot \boldsymbol{v} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \tag{6.9}$$

$$\textbf{slope} = \frac{df}{ds} = (\textbf{grad} f) \cdot \boldsymbol{T} = \frac{\partial f}{\partial x}\frac{dx}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} \tag{6.10}$$

The first involves *time*. If we move faster, df/dt increases. The second involves *distance*. If we move a distance ds, at any speed, the function changes by df. So the slope in that direction is df/ds.

Uniform motion on a straight line has $\boldsymbol{R} = \boldsymbol{R}_0 + \boldsymbol{v}t$. The velocity $\boldsymbol{v}$ is constant. The direction $\boldsymbol{T} = \boldsymbol{u} = \boldsymbol{v}/|\boldsymbol{v}|$ is also constant. The directional derivative is $(\textbf{grad} f) \cdot \boldsymbol{u}$, but the rate of change is $(\textbf{grad} f) \cdot \boldsymbol{v}$.

Equations (6.9) and (6.10) look like chain rules. They are chain rules. The next section extends $df/dt = (df/dx)(dx/dt)$ to more variables, proving (6.9) and (6.10). Here we focus on the meaning : **df/ds is the derivative off in the direction $\boldsymbol{u} = \boldsymbol{T}$ along the curve.**

Exemple 34. Find df/dt and df/ds for $f = r$. The curve is $x = t^2$, $y = t$ in **Figure 19 (left)** .

Answer 34 The velocity along the curve is $\boldsymbol{v} = 2t\,\boldsymbol{i} + \boldsymbol{j}$. At the typical point $t = 1$ it is $\boldsymbol{v} = 2\boldsymbol{i} + \boldsymbol{j}$. The unit tangent is $\boldsymbol{T} = \boldsymbol{v}/\sqrt{5}$. The gradient is a unit vector $\boldsymbol{i}/\sqrt{2} + \boldsymbol{j}/\sqrt{2}$ pointing outward, when $f(x, y)$ is the distance r from the center. The dot product with $\boldsymbol{v}$ is $df/dt = 3/\sqrt{2}$. The dot product with $\boldsymbol{T}$ is $df/ds = 3/\sqrt{10}$. When we slow down to speed 1 (with $\boldsymbol{T}$), the changes in $f(x, y)$ slow down too.

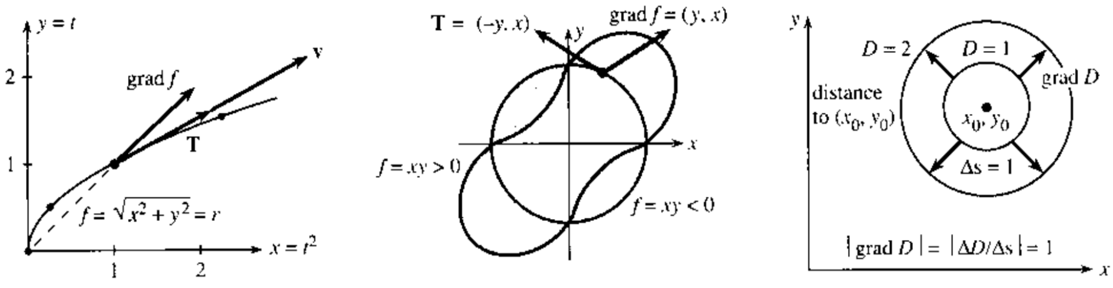

Figure 19: The distance $f = r$ changes along the curve. The slope of the roller-coaster is $(\boldsymbol{\nabla} f) \cdot \boldsymbol{T}$. The distance D from (x_0, y_0) has $\textbf{grad}\, D =$ unit vector.

Exemple 35. Find df/ds for $f = xy$ along the circular path $x = \cos t$, $y = \sin t$.

Answer 35 First take a direct approach. On the circle, xy equals $(\cos t)(\sin t)$. Its derivative comes from the product rule: $df/dt = \cos^2 t - \sin^2 t$. Normally this is different from df/ds, because the time t need not equal the arc length s. There is a speed factor ds/dt to divide by, but here the speed is 1. (A circle of length $s = 2\pi$ is completed at $t = 2\pi$). Thus the slope df/ds along the roller-coaster in **Figure 19** is $\cos^2 t - \sin^2 t$.

The second approach uses the vectors $\boldsymbol{\nabla} f$ and $\boldsymbol{T}$. The gradient of $f = xy$ is $(y, x) = (\sin t, \cos t)$. The unit tangent vector to the path is $\boldsymbol{T} = (-\sin t, \cos t)$. Their dot product is the same df/ds :

$$\textbf{slope along path} = (\boldsymbol{\nabla} f) \cdot \boldsymbol{T} = -\sin^2 t + \cos^2 t \tag{6.11}$$

6.4 Gradients without coordinates

This section ends with a little « philosophy ». What is the coordinate-free definition of the gradient ? Up to now, $\boldsymbol{\nabla} f = (f_x, f_y)$ depended totally on the choice of x and y axes. But the steepness of a surface is independent of the axes. Those are added later, to help us compute.

The steepness df/ds involves only f and the direction, nothing else. The gradient should be a « tensor », its meaning does not depend on the coordinate system. The gradient has different formulas in different systems (xy or $r\theta$ or ...), but the direction and length of grad! are determined by df/ds-without any axes:

- The **direction** of $\boldsymbol{\nabla} f$ is the one in which df/ds is largest.
- The **length** $|\boldsymbol{\nabla} f|$ is that largest slope.

The key equation is (**change in** f) $\approx$ (**gradient of** f) $\cdot$ (**change in position**). That is another way to write $\Delta f \approx f_x \, \Delta x + f_y \, \Delta y$. It is the multivariable form, we used two variables, of the basic linear approximation $\Delta y \approx (dy/dx)\Delta x$.

Exemple 36. $D(x, y) =$ distance from (x, y) to (x_0, y_0). Without derivatives prove $|\mathbf{grad}\, D| = 1$. The graph of $D(x, y)$ is a cone with slope 1 and sharp point (x_0, y_0).

First question In which direction does the distance $D(x, y)$ increase fastest ?
Answer Going directly away from (x_0, y_0). Therefore this is the direction of $\mathbf{grad}\, D$.

Second question How quickly does D increase in that steepest direction ?
Answer A step of length Δs increases D by Δs. Therefore $|\mathbf{grad}\, D| = \Delta s/\Delta s = 1$.

Conclusion **grad** D **is a unit rector.** The derivatives of D in **Exercise 19** are $(x - x_0)/D$ and $(y - y_0)/D$. The sum of their squares is 1, because $(x - x_0)^2 + (y - y_0)^2$ equals D^2.

7 The Chain Rule

Calculus goes back and forth between solving problems and getting ready for harder problems. The first is « application », the second looks like « theory ». If we minimize f to save time or money or energy, that is an application. If we don't take derivatives to find the minimum, maybe because f is a function of other functions, and we don't have a chain rule, then it is time for more theory. The chain rule is a fundamental working tool, because $f(g(x))$ appears all the time in applications. So do $f(g(x, y))$ and $f(x(t), y(t))$ and worse. We have to know their derivatives. Otherwise calculus can't continue with the applications.

You may instinctively say: Don't bother with the theory, just teach me the formulas. That is not possible. You now regard the derivative of $\sin 2x$ as a trivial problem, unworthy of an answer. That was not always so. Before the chain rule, the slopes of $\sin 2x$ and $\sin x^2$ and $\sin^2 x^2$ were hard to compute from $\Delta f/\Delta x$. We are now at the same point for $f(x, y)$. We know the *meaning* of $\partial f/\partial x$, but if $f = r \tan\theta$ and $x = r\cos\theta$ and $y = r\sin\theta$, we need a way to *compute* $\partial f/\partial x$. A little theory is unavoidable, if the problem-solving part of calculus is to keep going.

To repeat: **The chain rule applies to a function of a function**. In one variable that was $f(g(x))$. With two variables there are more possibilities :

1. $f(z)$ with $z = g(x, y)$. Find $\partial f/\partial x$ and $\partial f/\partial y$.
2. $f(x, y)$ with $x = x(t)$, $y = y(t)$. Find df/dt.
3. $f(x, y)$ with $x = x(t, u)$, $y = y(t, u)$. Find $\partial f/\partial t$ and $\partial f/\partial u$.

All derivatives are assumed continuous. More exactly, the *input* derivatives like $\partial g/\partial x$ and dx/dt and $\partial x/\partial u$ are continuous. Then the output derivatives like $\partial f/\partial x$ and df/dt and $\partial f/\partial u$ will be continuous from the chain rule. We avoid points like $r = 0$ in polar coordinates, where $\partial r/\partial x = x/r$ has a division by zero.

A Typical Problem Start with a function of x and y, for example x times y. Thus $f(x, y) = xy$. Change x to $r\cos\theta$ and y to $r\sin\theta$. The function becomes $(r\cos\theta)$ times $(r\sin\theta)$. We want its derivatives with respect to r and θ. First we have to decide on its *name*.

To be correct, we should not reuse the letter f. The new function can be F :

$$f(x, y) = xy \qquad f(r\cos\theta, r\sin\theta) = (r\cos\theta)(r\sin\theta) = F(r, \theta). \tag{7.1}$$

Why not call it $f(r, \theta)$? Because strictly speaking that is r times θ ! If we follow the rules, then $f(x, y)$ is xy and $f(r, \theta)$ should be $r\theta$. The new function F does the right thing, it multiplies $(r\cos\theta)(r\sin\theta)$. But in many cases, the rules get bent and the letter F is changed back to f.

This crime has already occurred. The end of the last page ought to say $\partial F/\partial t$. Instead the printer put $\partial f/\partial t$. The purpose of the chain rule is to find derivatives in the new variables t and u (or r and θ). In our example we want **the derivative of F with respect to r**. Here is the chain rule :

$$\frac{\partial F}{\partial x} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = y\,\cos\theta + x\,\sin\theta = 2r\,\sin\theta\,\cos\theta \tag{7.2}$$

I substituted $r\sin\theta$ and $r\cos\theta$ for y and x. You immediately check the answer: $F(r, \theta) = r^2\cos\theta\sin\theta$ does lead to $\partial F/\partial r = 2r\cos\theta\sin\theta$. The derivative is correct. The only incorrect thing, but we do it anyway, is to write f instead of F.

Question What is $\dfrac{\partial F}{\partial \theta}$? *Answer* It is $\dfrac{\partial f}{\partial x}\dfrac{\partial x}{\partial \theta} + \dfrac{\partial f}{\partial y}\dfrac{\partial y}{\partial \theta}$.

7.1 The derivatives of $f(g(x, y))$

Here g depends on x and y, and f depends on g. Suppose x moves by dx, while y stays constant. Then g moves by $dg = (\partial g/\partial x)dx$. When g changes, f also changes: $df = (df/dg)dg$. Now substitute for dg to make the chain: $df = (df/dg)(\partial g/\partial x)dx$. This is the first rule :

$$\frac{\partial f}{\partial x} = \frac{df}{dg}\frac{\partial g}{\partial x} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = \frac{df}{dg}\frac{\partial g}{\partial y} \tag{7.3}$$

Exemple 37. Every $f(x+cy)$ satisfies the 1-way wave equation $\partial f/\partial y = c\partial f/\partial x$. The inside function is $g = x+cy$. The outside function can be anything, g^2 or $\sin g$ or e^g. The composite function is $(x+cy)^2$ or $\sin(x+cy)$ or e^{x+cy}. In each separate case we could check that $\partial f/\partial y = c\partial f/\partial x$. The chain rule produces this equation in all cases at once, from $\partial g/\partial x = 1$ and $\partial g/\partial y = c$:

$$\frac{\partial f}{\partial x} = \frac{df}{dg}\frac{\partial g}{\partial x} = 1\,\frac{\partial f}{\partial g} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{df}{dg}\frac{\partial g}{\partial y} = c\,\frac{\partial f}{\partial g} \quad \text{so} \quad \frac{\partial f}{\partial y} = c\,\frac{\partial f}{\partial x} \tag{7.4}$$

This is important : $\partial f/\partial y = c\partial f/\partial x$ is our first example of a **partial differential equation**. The unknown $f(x,y)$ has two variables. Two partial derivatives enter the equation.

Up to now we have worked with dy/dt and **ordinary differential equations**. The independent variable was time or space (and only one dimension in space). For partial differential equations the variables are time and space (possibly several dimensions in space). The great equations of mathematical physics, heat equation, wave equation, Laplace's equation, are partial differential equations.

Notice how the chain rule applies to $f = \sin xy$. Its x derivative is $y\cos xy$. A patient reader would check that f is $\sin g$ and g is xy and f_x is $f_x g_x$. Probably you are not so patient-you know the derivative of $\sin xy$. Therefore we pass quickly to the next chain rule. Its outside function depends on two inside functions, and each of those depends on t. We want df/dt.

7.2 The derivatives of $f(x(t), y(t))$

Before the formula, here is the idea. Suppose t changes by Δt. That affects x and y; they change by Δx and Δy. There is a domino effect on f; it changes by Δf. Tracing backwards,

$$\Delta f \approx \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y \quad \text{and} \quad \Delta x \approx \frac{dx}{dt}\Delta t \quad \text{and} \quad \Delta y \approx \frac{dy}{dt}\Delta t \tag{7.5}$$

Substitute the last two into the first, connecting Δf to Δt. Then let $\Delta t \to 0$:

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \tag{7.6}$$

This is close to the one-variable rule $dz/dx = (dz/dy)(dy/dx)$. There we could «cancel» dy. (We actually canceled Δy in $(\Delta z/\Delta y)(\Delta y/\Delta x)$, and then approached the limit.) Now Δt affects Δf in two ways, through x and through y. *The chain rule has two terms.* If we cancel in $(\partial f/\partial x)(\partial x/\partial t)$ we only get one of the terms!

We mention again that the true name for $f(x(t), yt))$ is $F(t)$ not $f(t)$. For $f(x,y,z)$ the rule has three terms : $f_x x_t + f_y y_t + f_z z_t$ is f_t (or better dF/dt).

Exemple 38. How quickly does the temperature change when you drive to Florida? Suppose the Midwest is at 30°F and Florida is at 80°F. Going 1000 miles south increases the temperature $f(x,y)$ by 50°, or 0.05 degrees per mile. Driving straight south at 70 miles per hour, the rate of increase is $(0.05)(70) = 3.5$ degrees per hour. Note how (degrees/mile) times (miles/hour) equals (degrees/hour). That is the ordinary chain rule $(df/dx)(dx/dt = (df/dt)$, there is no y variable going south.

If the road goes southeast, the temperature is $f = 30 + 0.05\,x + 0.01\,y$. Now $x(t)$ is distance south and $y(t)$ is distance east. What is df/dt if the speed is still 70 ?

Answer 38 We have

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = 0.05\,\frac{70}{\sqrt{2}} + 0.01\,\frac{70}{\sqrt{2}} \approx 3\,\text{degrees/hour} \tag{7.7}$$

In reality there is another term. The temperature also depends directly on t, because of night and day. The factor $\cos(2\pi t/24)$ has a period of 24 hours, and it brings an extra term into the chain rule :

$$\textbf{For } f(x,y,t) \textbf{ the chain rule is } \frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial t} \tag{7.8}$$

This is the **total derivative** df/dt, from all causes. Changes in x, y, t all affect f. The partial derivative $\partial f/\partial t$ is only one part of df/dt. (Note that $dt/dt = 1$). If night and day add $12\cos(2\pi t/24)$ to f, the extra term is $\partial f/\partial t == \pi \sin(2\pi t/24)$. At nightfall that is $-\pi$ degrees per hour. You have to drive faster than 70 mph to get warm.

7.3 Second derivatives

What is $d^2 f/dt^2$? We need the derivative of (7.8), which is painful. So start with movement in a straight line.

Suppose $x = x_0 + t\cos\theta$ and $y = y_0 + t\sin\theta$. We are moving at the fixed angle θ, with speed 1. The derivatives are $x_t = \cos\theta and y_t = \sin\theta and \cos^2\theta + \sin^2\theta = 1$. Then df/dt is immediate from the chain rule:

$$f_t = f_x x_t + f_y y_t = f_x \cos\theta + f_y \sin\theta \tag{7.9}$$

For the second derivative f_{tt} apply this rule to f_t. Then f_t is

$$(f_t)_x \cos\theta + (f_t)_y \sin\theta = (f_{xx}\cos\theta + f_{yx}\sin\theta)\cos\theta + (f_{xy}\cos\theta + f_{yy}\sin\theta)\sin\theta \tag{7.10}$$

Collect terms :

$$f_{tt} = f_{xx}\cos^2\theta + 2f_{xy}\cos\theta\sin\theta + f_{yy}\sin^2\theta \tag{7.11}$$

In polar coordinates change t to r. When we move in the r direction, θ is fixed. Equation (7.11) gives f_{rr} from f_{xx}, f_{xy}, f_{yy}. Second derivatives on curved paths (with new terms from the curving) are saved for the exercises.

Exemple 39. f_{xx}, f_{xy}, f_{yy} are all continuous and bounded by M, find a bound on f_{tt},. This is the second derivative along any line.

Answer 39 Equation (7.11) gives $|f_{tt}| \le M\cos^2\theta + M\sin 2\theta + M\sin^2\theta \le 2M$. This upper bound $2M$ was needed in equation (5.16), for the error in linear approximation.

7.4 The derivatives of $f(x(t,u), y(t,u))$

Suppose there are two inside functions x and y, each depending on t and u. When t moves, x and y both move: $dx = x_t dt$ and $dy = y_t dt$. Then dx and dy force a change in f : $df = f_x dx + f_y dy$.

The chain rule for $\partial f/\partial t$ is no surprise :

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} \tag{7.12}$$

This rule has $\partial/\partial t$ instead of d/dt, because of the extra variable u. The symbols remind us that u is constant. Similarly t is constant while u moves, and there is a second chain rule for $\partial f/\partial u$: $f_u = f_x x_u + f_x y_u$.

Exemple 40. In polar coordinates find f_θ and $f_{\theta\theta}$. Start from $f(x, y) = f(r\cos\theta, r\sin\theta)$.

Answer 40 The chain rule uses the θ derivatives of x and y:

$$\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = \frac{\partial f}{\partial x}(-r\sin\theta) + \frac{\partial f}{\partial y}(r\cos\theta) \tag{7.13}$$

The second θ derivative is harder, because (7.13) has four terms that depend on θ. Apply the chain rule to the first term $\partial f/\partial x$. It is a function of x and y, and x and y are functions of θ :

$$\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\frac{\partial x}{\partial \theta} + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\frac{\partial y}{\partial \theta} = f_{xx}(-r\sin\theta) + f_{yy}(r\cos\theta) \tag{7.14}$$

The θ derivative of $\partial f/\partial y$ is similar. So apply the product rule to (7.13) :

$$\begin{aligned} f_{\theta\theta} &= [f_{xx}(-r\sin\theta) + f_{xy}(r\cos\theta)](-r\sin\theta) + f_x(-r\cos\theta) \\ &\quad + [f_{yx}(-r\sin\theta) + f_{yy}(r\cos\theta)](r\cos\theta) + f_y(-r\sin\theta) \end{aligned} \tag{7.15}$$

This formula is not attractive. In mathematics, a messy formula is almost always a signal of asking the wrong question. In fact the combination $f_{xx} + f_{yy}$ is much more special than the separate derivatives. We might hope the same $f_{rr} + f_{\theta\theta}$, but dimensionally that is impossible, since r is a length and θ is an angle. The dimensions of f_{xx} and f_{yy} are matched by f_{rr} and f_r/r and $f_{\theta\theta}/r^2$. We could even hope that

$$f_{xx} + f_{yy} = f_{rr} + \frac{1}{r}f_r + \frac{1}{r^2}f_{\theta\theta} \tag{7.16}$$

This equation is true. Add (7.9) + (7.11) + (7.15) with t changed to r. **Laplace's equation $f_{xx} + f_{yy} = 0$ is now expressed in polar coordinates :** $f_{rr} + f_r/r + f_{\theta\theta}/r^2 = 0$.

7.5 A paradox

Before leaving polar coordinates there is one more question. It goes back to $\partial r/\partial x$, which was practically the first example of partial derivatives:

$$\frac{\partial r}{\partial x} = \frac{\partial}{\partial x}\sqrt{x^2+y^2} = \frac{x}{\sqrt{x^2+y^2}} = \frac{x}{r} \tag{7.17}$$

My problem is this. We know that x is $r\cos\theta$. So x/r on the right side is $\cos\theta$. On the other hand r is $x/\cos\theta$. So $\partial r/\partial x$ is also $1/\cos\theta$. **How can $\partial r/\partial x$ lead to $\cos\theta$ one way and $1/\cos\theta$ the other way ?**

I will admit that this cost me a sleepless night. There must be an explanation, we cannot end with $\cos\theta = 1/\cos\theta$. This paradox brings a new respect for partial derivatives. May I tell you what I finally noticed? You could cover up the next paragraph and think about the puzzle first.

The key to partial derivatives is to ask : **Which variable is held constant ?** In equation (7.17), y is constant. But when $r = x/\cos\theta$ gave $\partial r/\partial x = 1/\cos\theta$ was constant. In both cases we change x and look at the effect on r. The movement is on a horizontal line (constant y) or on a radial line (constant θ). **Figure 20** shows the difference.

Remarque 1. This example shows that $\partial r/\partial x$ is different from $1/(\partial x/\partial r)$. The neat formula $(\partial r/\partial x)(\partial x/\partial r) = 1$ is not generally true. May I tell you what takes its place? We have to include $(\partial r/\partial y)(\partial y/\partial r)$. With two variables xy and two variables $r\theta$, we need 2 by 2 matrices! The next section gives the details:

$$\begin{bmatrix} \partial r/\partial x & \partial r/\partial y \\ \partial\theta/\partial x & \partial\theta/\partial y \end{bmatrix} \begin{bmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

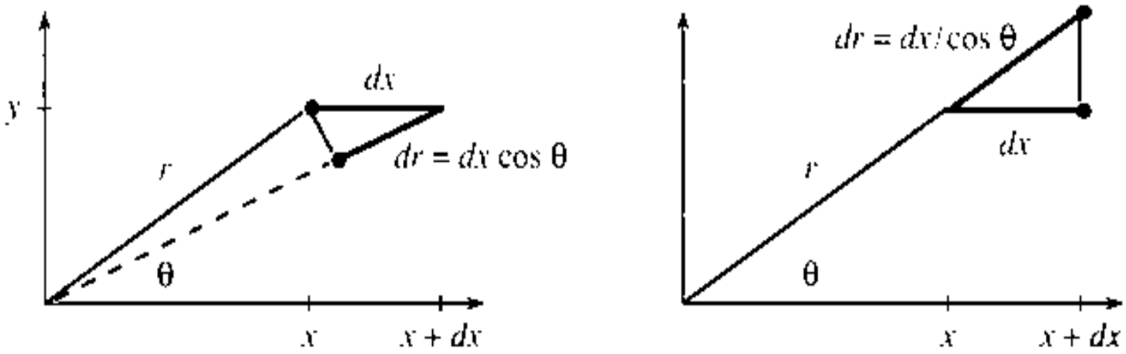

Figure 20: $dr = dx\ \cos\theta$ when y is constant, $dr = dx/\cos\theta$ when θ is constant.

7.6 Non-independent variables

This paradox points to a serious problem. In computing partial derivatives of $f(x, y, z)$, we assumed that x, y, z were independent. Up to now, x could move while y and z were fixed. In physics and chemistry and economics that may not be possible. If there is a relation between x, y, z, then x can't move by itself.

Exemple 41. The gas law $PV = nRT$ relates pressure to volume and temperature. P, V, T are not independent. What is the meaning of $\partial V/\partial P$? Does it equal $1/(\partial P/\partial V)$?

Answer 41 Those questions have no answers, until we say what is held constant. In the paradox, $\partial r/\partial x$ had one meaning for fixed y and another meaning for fixed θ. To indicate what is held constant, use an extra subscript (not denoting a derivative):

$$\left(\frac{\partial r}{\partial x}\right)_y = \cos\theta \qquad\qquad \left(\frac{\partial r}{\partial x}\right)_\theta = \frac{1}{\cos\theta} \tag{7.18}$$

$(\partial f/\partial P)_V$ has constant volume and $(\partial f/\partial P)_T$ has constant temperature. The usual $\partial f/\partial P$ has both V and T constant. But then the gas law won't let us change P.

Exemple 42. Let $f = 3x + 2y + z$. Compute $\partial f/\partial x$ on the plane $z = 4x + y$.

Answer 42 We have three possible approach :

1. Think of x and y as independent. Replace z by $4x + y$:

$$f = 3x + 2y + (4x + y) \qquad \text{so} \qquad \left(\frac{\partial f}{\partial x}\right)_y = 7. \tag{7.19}$$

2. Keep x and y independent. Deal with z by the chain rule :

$$\left(\frac{\partial f}{\partial x}\right)_y = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 3 + 1 \times 4 = 7 \tag{7.20}$$

3. (different) Make x and z independent. Then $y = z - 4x$:

$$\left(\frac{\partial f}{\partial x}\right)_z = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 3 + 2 \times (-4) = -5 \tag{7.21}$$

Without a subscript, $\partial f/\partial x$ means: Take the x derivative the usual way. The answer is $\partial f/\partial x = 3$, when y and z don't move. But on the plane $z = 4x + y$, one of them must move! 3 is only part of the total answer, which is $(\partial f/\partial x)_y = 7$ or $(\partial f/\partial x)_z = -5$.

Here is the geometrical meaning. We are on the plane $z = 4x + y$. The derivative $(\partial f/\partial x)_y$ moves x but not y. To stay on the plane, dz is $4\,dx$. The change in $f = 3x + 2y + z$ is $df = 3\,dx + 0 + dz = 7\,dx$.

Exemple 43. On the world line $x^2 + y^2 + z^2 = t^2$ find $(\partial f/\partial y)_{x,z}$ for $f = xyzt$.

Answer 43 The subscripts x, z mean that x and z are fixed. The chain rule skips $\partial f/\partial y$ and $\partial f/\partial z$:

$$\left(\frac{\partial f}{\partial y}\right)_{x,z} = \frac{\partial f}{\partial y} + \frac{\partial f}{\partial t}\frac{\partial t}{\partial y} = xzt + xyz \times \frac{y}{t} \qquad \left(\text{why } \frac{y}{t}\ ?\right) \tag{7.22}$$

Exemple 44. From the law $PV = T$, compute the product $(\frac{\partial P}{\partial V})_T(\frac{\partial V}{\partial T})_P(\frac{\partial T}{\partial P})_V$.

Answer 44 Any intelligent person cancels ∂V's, ∂T's, and ∂P's to get 1. The right answer is -1 :

$$\left(\frac{\partial P}{\partial V}\right)_T = -\frac{T}{V^2} \qquad \left(\frac{\partial V}{\partial T}\right)_P = \frac{1}{P} \qquad \left(\frac{\partial T}{\partial P}\right)_V = V \tag{7.23}$$

The product is $-T/PV$. **This is** -1 **not** $+1$? The chain rule is tricky (see **Exercise 23**).

Exemple 45. Implicit differentiation cane explained by the chain rule :

$$\text{If } \ F(x, y) = 0 \ \text{ then } \ F_x + F_y y_x = 0 \ \text{ so } \ \frac{dy}{dx} = -\frac{F_x}{F_y} \tag{7.24}$$

8 Maxima, minima and saddle points

The outstanding equation of differential calculus is also the simplest: $df/dx = 0$. The slope is zero and the tangent line is horizontal. Most likely we are at the top or bottom of the graph, a maximum or a minimum. This is the point that the engineer or manager or scientist or investor is looking for, maximum stress or production or velocity or profit. With more variables in $f(x, y)$ and $f(x, y, z)$, the problem becomes more realistic. The question still is: **How to locate the maximum and**

minimum ?

The answer is in the partial derivatives. When the graph is level, they are zero. Deriving the equations $f_x = 0$ and $f_y = 0$ is pure mathematics and pure pleasure. Applying them is the serious part. We watch out for saddle points, and also for a minimum at a boundary point, this section takes extra time. Remember the steps for $f(x)$ in one-variable calculus:

1. The leading candidates are **stationary** points (where $df/dx = 0$).
2. The other candidates are **rough points** (no derivative) and **endpoints** (a or b).
3. Maximum vs. minimum is decided by the sign of the **second derivative**.

In two dimensions, a stationary point requires $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$. The tangent line becomes a tangent plane. The endpoints a and b are replaced by a boundary curve. In practice boundaries contain about 40% of the minima and 80% of the work.

Finally there are three second derivatives f_{xx}, f_{xy}, and f_{yy}. They tell how the graph bends away from the tangent plane, up at a minimum, down at a maximum, both ways at a *saddle point*. This will be determined by comparing $(f_{xx})(f_{yy})$ with $(f_{xy})^2$.

8.1 Stationary point → Horizontal tangent → Zero derivatives

Suppose f has a minimum at the point (x_0, y_0). This may be an **absolute minimum** or only a **local minimum**. In both cases $f(x_0, y_0) \le f(x, y)$ near the point. For an absolute minimum, this inequality holds wherever f is defined. For a local minimum, the inequality can fail far away from (x_0, y_0). The bottom of your foot is an absolute minimum, the end of your finger is a local minimum.

We assume for now that (x_0, y_0) is an *interior point* of the domain off. At a *boundary point*, we cannot expect a horizontal tangent and zero derivatives.

Main conclusion: At a minimum or maximum (absolute or local) a nonzero derivative is impossible. The tangent plane would tilt. In some direction f would decrease. Note that the minimum *point* is (x_0, y_0), and the minimum *value* is $f(x_0, y_0)$.

If derivatives exist at an interior minimum or maximum, they are zero :

$$\frac{\partial f}{\partial x} = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y} = 0 \qquad \text{(together this is } \boldsymbol{\nabla} f = \mathbf{0}\text{)}$$

For a function $f(x, y, z)$ of three variables, add the third equation $\partial f/\partial z = 0$.

The reasoning goes back to the one-variable case. That is because we look along the lines $x = x_0$ and $y = y_0$. The minimum of $f(x, y)$ is at the point where the lines meet. So this is also the minimum **along each line separately**.

Moving in the x direction along $y = y_0$, we find $\partial f/\partial x = 0$. Moving in the y direction, $\partial f/\partial y = 0$ at the same point. **The slope in every direction is zero**. In other words $\boldsymbol{\nabla} f = \mathbf{0}$.

Graphically, (x_0, y_0) is the low point of the surface $z = f(x, y)$. Both cross sections in **Figure 21** touch bottom. The phrase «if derivatives exist» rules out the vertex of a cone, which is a