34 A quadratic function $ax^2 + by^2 + cx + dy$ has the gradients shown in Figure B. Estimate a, b, c, d and sketch two level curves.

35 The level curves of $f(x, y)$ are circles around $(1, 1)$. The curve $f = c$ has radius $2c$. What is f? What is grad f at $(0, 0)$?

36 Suppose grad f is tangent to the hyperbolas $xy = $ constant in Figure C. Draw three level curves of $f(x, y)$. Is $|$grad $f|$ larger at P or Q? Is $|$grad $f|$ constant along the hyperbolas? Choose a function that could be f: $x^2 + y^2$, $x^2 - y^2$, xy, $x^2 y^2$.

37 Repeat Problem 36, if grad f is *perpendicular* to the hyperbolas in Figure C.

38 If $f = 0, 1, 2$ at the points $(0, 1), (1, 0), (2, 1)$, estimate grad f by assuming $f = Ax + By + C$.

39 What functions have the following gradients?
(a) $(2x + y, x)$ (b) $(e^{x-y}, -e^{x-y})$ (c) $(y, -x)$ (careful)

40 Draw level curves of $f(x, y)$ if grad $f = (y, x)$.

In 41–46 find the velocity v and the tangent vector T. Then compute the rate of change $df/dt = $ grad $f \cdot $ v **and the slope** $df/ds = $ grad $f \cdot $ T.

41 $f = x^2 + y^2$ $x = t$ $y = t^2$

42 $f = x$ $x = \cos 2t$ $y = \sin 2t$

43 $f = x^2 - y^2$ $x = x_0 + 2t$ $y = y_0 + 3t$

44 $f = xy$ $x = t^2 + 1$ $y = 3$

45 $f = \ln xyz$ $x = e^t$ $y = e^{2t}$ $z = e^{-t}$

46 $f = 2x^2 + 3y^2 + z^2$ $x = t$ $y = t^2$ $z = t^3$

47 (a) Find df/ds and df/dt for the roller-coaster $f = xy$ along the path $x = \cos 2t$, $y = \sin 2t$. (b) Change to $f = x^2 + y^2$ and explain why the slope is zero.

48 The distance D from (x, y) to $(1, 2)$ has $D^2 = (x - 1)^2 + (y - 2)^2$. Show that $\partial D/\partial x = (x - 1)/D$ and $\partial D/\partial y = (y - 2)/D$ and $|$grad $D| = 1$. The graph of $D(x, y)$ is a _____ with its vertex at _____.

49 If $f = 1$ and grad $f = (2, 3)$ at the point $(4, 5)$, find the tangent plane at $(4, 5)$. If f is a linear function, find $f(x, y)$.

50 Define the derivative of $f(x, y)$ in the direction $\mathbf{u} = (u_1, u_2)$ at the point $P = (x_0, y_0)$. What is Δf (approximately)? What is $D_{\mathbf{u}} f$ (exactly)?

51 The slope of f along a level curve is $df/ds = $ _____ $= 0$. This says that grad f is perpendicular to the vector _____ in the level direction.

13.5 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:

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

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:

$$\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = (y)(\cos \theta) + (x)(\sin \theta) = 2r \sin \theta \cos \theta.$$

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 $\dfrac{\partial f}{\partial x}\dfrac{\partial x}{\partial \theta} + \dfrac{\partial f}{\partial y}\dfrac{\partial y}{\partial \theta}$.

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:

$$\boxed{\textbf{13G} \quad \textit{Chain rule for } f(g(x, y)): \quad \frac{\partial f}{\partial x} = \frac{df}{dg}\frac{\partial g}{\partial x} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{df}{dg}\frac{\partial g}{\partial y}.} \qquad (1)$$

EXAMPLE 1   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 $\partial g/\partial y = c$:

$$\frac{\partial f}{\partial x} = \frac{df}{dg}\frac{\partial g}{\partial x} = 1\frac{df}{dg} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{df}{dg}\frac{\partial g}{\partial y} = c\frac{df}{dg} \quad \text{so} \quad \frac{\partial f}{\partial y} = c\frac{\partial f}{\partial x}. \qquad (2)$$

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 $f_g g_x$. 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.

THE DERIVATIVE 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,

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

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

13H *Chain rule for* $f(x(t), y(t))$: $\dfrac{df}{dt} = \dfrac{\partial f}{\partial x}\dfrac{dx}{dt} + \dfrac{\partial f}{\partial y}\dfrac{dy}{dt}.$	(3)

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 $(\partial f/\partial x)(dx/dt)$ we only get one of the terms!

We mention again that the true name for $f(x(t), y(t))$ 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).

EXAMPLE 2 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 .05 degrees per mile. Driving straight south at 70 miles per hour, the rate of increase is $(.05)(70) = 3.5$ 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 $f = 30 + .05x + .01y$. Now $x(t)$ is distance south and $y(t)$ is distance east. What is df/dt if the speed is still 70?

Solution $\dfrac{df}{dt} = \dfrac{\partial f}{\partial x}\dfrac{dx}{dt} + \dfrac{\partial f}{\partial y}\dfrac{dy}{dt} = .05\dfrac{70}{\sqrt{2}} + .01\dfrac{70}{\sqrt{2}} \approx 3$ degrees/hour.

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:

$$\text{For } f(x, y, t) \text{ the chain rule is } \frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial t}. \tag{4}$$

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.

SECOND DERIVATIVES

What is $d^2 f/dt^2$? We need the derivative of (4), which is painful. It is like acceleration in Chapter 12, with many terms. 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:

$$f_t = f_x x_t + f_y y_t = f_x \cos \theta + f_y \sin \theta. \tag{5}$$

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

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

Collect terms: $$f_{tt} = f_{xx} \cos^2\theta + 2f_{xy} \cos \theta \sin \theta + f_{yy} \sin^2\theta. \tag{6}$$

In polar coordinates change t to r. When we move in the r direction, θ is fixed. Equation (6) 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.

EXAMPLE 3 If 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.

Solution Equation (6) gives $|f_{tt}| \leqslant M \cos^2\theta + M \sin 2\theta + M \sin^2\theta \leqslant 2M$. This upper bound $2M$ was needed in equation 13.3.14, for the error in linear approximation.

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:

$$\boxed{\textbf{13I} \quad \text{Chain rule for } f(x(t, u), y(t, u)): \quad \frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.} \tag{7}$$

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 $\partial f/\partial u$: $f_u = f_x x_u + f_y y_u$.

EXAMPLE 4 In polar coordinates find f_θ and $f_{\theta\theta}$. Start from $f(x, y) = f(r \cos \theta, r \sin \theta)$.

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

$$\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = \left(\frac{\partial f}{\partial x}\right)(-r \sin \theta) + \left(\frac{\partial f}{\partial y}\right)(r \cos \theta). \tag{8}$$

The second θ derivative is harder, because (8) 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:

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

The θ derivative of $\partial f/\partial y$ is similar. So apply the product rule to (8):

$$f_{\theta\theta} = [f_{xx}(-r \sin \theta) + f_{xy}(r \cos \theta)](-r \sin \theta) + f_x(-r \cos \theta)$$
$$+ [f_{yx}(-r \sin \theta) + f_{yy}(r \cos \theta)](r \cos \theta) + f_y(-r \sin \theta). \tag{9}$$

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 for $f_{rr} + f_{\theta\theta}$, 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

$$f_{xx} + f_{yy} = f_{rr} + \frac{1}{r} f_r + \frac{1}{r^2} f_{\theta\theta}. \tag{10}$$

This equation is true. Add $(5) + (6) + (9)$ with t changed to r. **Laplace's equation** $f_{xx} + f_{yy} = 0$ *is now expressed in polar coordinates*: $f_{rr} + f_r/r + f_{\theta\theta}/r^2 = 0$.

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:

$$\frac{\partial r}{\partial x} = \frac{\partial}{\partial x} \sqrt{x^2 + y^2} = x/\sqrt{x^2 + y^2} = x/r. \tag{11}$$
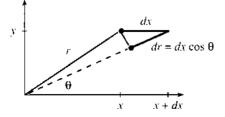
My problem is this. We know that x is $r \cos \theta$. So x/r on the right side is $\cos \theta$. On the other hand r is $x/\cos \theta$. So $\partial r/\partial x$ is also $1/\cos \theta$. **How can $\partial r/\partial x$ lead to $\cos \theta$ one way and $1/\cos \theta$ the other way?**

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 (11), y is constant. But when $r = x/\cos \theta$ gave $\partial r/\partial x = 1/\cos \theta$, θ 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 13.15 shows the difference.

Remark 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! Section 14.4 gives the details:

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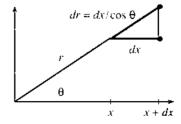

Fig. 13.15 $dr = dx \cos \theta$ when y is constant, $dr = dx/\cos \theta$ when θ is constant.

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.

EXAMPLE 5   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 $1/(\partial P/\partial V)$?

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):

$$(\partial r/\partial x)_y = \cos \theta \qquad (\partial r/\partial x)_\theta = 1/\cos \theta. \tag{12}$$

$(\partial f/\partial P)_V$ 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.

EXAMPLE 6   Let $f = 3x + 2y + z$. Compute $\partial f/\partial x$ on the plane $z = 4x + y$.

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

$$f = 3x + 2y + (4x + y) \quad \text{so} \quad (\partial f/\partial x)_y = 7.$$

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

$$(\partial f/\partial x)_y = \partial f/\partial x + (\partial f/\partial z)(\partial z/\partial x) = 3 + (1)(4) = 7.$$

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

$$(\partial f/\partial x)_z = \partial f/\partial x + (\partial f/\partial y)(\partial y/\partial x) = 3 + (2)(-4) = -5.$$

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 $4dx$. The change in $f = 3x + 2y + z$ is $df = 3dx + 0 + dz = 7dx$.

EXAMPLE 7   On the world line $x^2 + y^2 + z^2 = t^2$ find $(\partial f/\partial y)_{x,z}$ for $f = xyzt$.

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

$$(\partial f/\partial y)_{x,z} = \partial f/\partial y + (\partial f/\partial t)(\partial t/\partial y) = xzt + (xyz)(y/t). \quad \textit{Why } y/t?$$

EXAMPLE 8   From the law $PV = T$, compute the product $(\partial P/\partial V)_T(\partial V/\partial T)_P(\partial T/\partial P)_V$.

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

$$(\partial P/\partial V)_T = -T/V^2 \qquad (\partial V/\partial T)_P = 1/P \qquad (\partial T/\partial P)_V = V.$$

The product is $-T/PV$. *This is* -1 *not* $+1$! The chain rule is tricky (Problem 42).

EXAMPLE 9   Implicit differentiation was used in Chapter 4. The chain rule explains it:

$$\textit{If } F(x, y) = 0 \textit{ then } F_x + F_y y_x = 0 \textit{ so } dy/dx = -F_x/F_y. \tag{13}$$

13.5 EXERCISES

Read-through questions

The chain rule applies to a function of a __a__. The x derivative of $f(g(x, y))$ is $\partial f/\partial x =$ __b__. The y derivative is $\partial f/\partial y =$ __c__. The example $f = (x + y)^n$ has $g =$ __d__. Because $\partial g/\partial x = \partial g/\partial y$ we know that __e__ = __f__. This __g__ differential equation is satisfied by any function of $x + y$.

Along a path, the derivative of $f(x(t), y(t))$ is $df/dt =$ __h__. The derivative of $f(x(t), y(t), z(t))$ is __i__. If $f = xy$ then the chain rule gives $df/dt =$ __j__. That is the same as the __k__ rule! When $x = u_1 t$ and $y = u_2 t$ the path is __l__. The chain rule for $f(x, y)$ gives $df/dt =$ __m__. That is the __n__ derivative $D_u f$.

The chain rule for $f(x(t, u), y(t, u))$ is $\partial f/\partial t =$ __o__. We don't write df/dt because __p__. If $x = r \cos \theta$ and $y = r \sin \theta$, the variables t, u change to __q__. In this case $\partial f/\partial r =$ __r__ and $\partial f/\partial \theta =$ __s__. That connects the derivatives in __t__ and __u__ coordinates. The difference between $\partial r/\partial x = x/r$ and $\partial r/\partial x = 1/\cos \theta$ is because __v__ is constant in the first and __w__ is constant in the second.

With a relation like $xyz = 1$, the three variables are __x__ independent. The derivatives $(\partial f/\partial x)_y$ and $(\partial f/\partial x)_z$ and $(\partial f/\partial x)$ mean __y__ and __z__ and __A__. For $f = x^2 + y^2 + z^2$ with $xyz = 1$, we compute $(\partial f/\partial x)_z$ from the chain rule __B__. In that rule $\partial z/\partial x =$ __C__ from the relation $xyz = 1$.

Find f_x and f_y in Problems 1–4. What equation connects them?

1 $f(x, y) = \sin(x + cy)$ 2 $f(x, y) = (ax + by)^{10}$

3 $f(x, y) = e^{x - 7y}$ 4 $f(x, y) = \ln(x + 7y)$

5 Find both terms in the t derivative of $(g(x(t), y(t)))^3$.

6 If $f(x, y) = xy$ and $x = u(t)$ and $y = v(t)$, what is df/dt? Probably all other rules for derivatives follow from the chain rule.

7 The step function $f(x)$ is zero for $x < 0$ and one for $x > 0$. Graph $f(x)$ and $g(x) = f(x + 2)$ and $h(x) = f(x + 4)$. If $f(x + 2t)$ represents a wall of water (a tidal wave), which way is it moving and how fast?

8 The wave equation is $f_{tt} = c^2 f_{xx}$. (a) Show that $(x + ct)^n$ is a solution. (b) Find C different from c so that $(x + Ct)^n$ is also a solution.

9 If $f = \sin(x - t)$, draw two lines in the xt plane along which $f = 0$. Between those lines sketch a sine wave. Skiing on top of the sine wave, what is your speed dx/dt?

10 If you float at $x = 0$ in Problem 9, do you go up first or down first? At time $t = 4$ what is your height and upward velocity?

11 *Laplace's equation* is $f_{xx} + f_{yy} = 0$. Show from the chain rule that any function $f(x + iy)$ satisfies this equation if $i^2 = -1$. Check that $f = (x + iy)^2$ and its real part _____ and its imaginary part _____ all satisfy Laplace's equation.

12 Equation (10) gave the polar form $f_{rr} + f_r/r + f_{\theta\theta}/r^2 = 0$ of Laplace's equation. (a) Check that $f = r^2 e^{2i\theta}$ and its real part $r^2 \cos 2\theta$ and its imaginary part $r^2 \sin 2\theta$ all satisfy Laplace's equation. (b) Show from the chain rule that any function $f(re^{i\theta})$ satisfies this equation if $i^2 = -1$.

In Problems 13–18 find df/dt from the chain rule (3).

13 $f = x^2 + y^2$, $x = t$, $y = t^2$

14 $f = \sqrt{x^2 + y^2}$, $x = t$, $y = t^2$

15 $f = xy$, $x = 1 - \sqrt{t}$, $y = 1 + \sqrt{t}$

16 $f = x/y$, $x = e^t$, $y = 2e^t$

17 $f = \ln(x + y)$, $x = e^t$, $y = e^t$

18 $f = x^4$, $x = t$, $y = t$

19 If a cone grows in height by $dh/dt = 1$ and in radius by $dr/dt = 2$, starting from zero, how fast is its volume growing at $t = 3$?

20 If a rocket has speed $dx/dt = 6$ down range and $dy/dt = 2t$ upward, how fast is it moving away from the launch point at $(0, 0)$? How fast is the angle θ changing, if $\tan \theta = y/x$?

21 If a train approaches a crossing at 60 mph and a car approaches (at right angles) at 45 mph, how fast are they coming together? (a) Assume they are both 90 miles from the crossing. (b) Assume they are going to hit.

22 In Example 2 does the temperature increase faster if you drive due south at 70 mph or southeast at 80 mph?

23 On the line $x = u_1 t$, $y = u_2 t$, $z = u_3 t$, what combination of f_x, f_y, f_z gives df/dt? This is the *directional derivative* in 3D.

24 On the same line $x = u_1 t$, $y = u_2 t$, $z = u_3 t$, find a formula for $d^2 f/dt^2$. Apply it to $f = xyz$.

25 For $f(x, y, t) = x + y + t$ find $\partial f/\partial t$ and df/dt when $x = 2t$ and $y = 3t$. Explain the difference.

26 If $z = (x + y)^2$ then $x = \sqrt{z} - y$. Does $(\partial z/\partial x)(\partial x/\partial z) = 1$?

27 Suppose $x_t = t$ and $y_t = 2t$, not constant as in (5–6). For $f(x, y)$ find f_t and f_{tt}. The answer involves $f_x, f_y, f_{xx}, f_{xy}, f_{yy}$.

28 Suppose $x_t = t$ and $y_t = t^2$. For $f = (x + y)^3$ find f_t and then f_{tt} from the chain rule.

29 Derive $\partial f/\partial r = (\partial f/\partial x) \cos \theta + (\partial f/\partial y) \sin \theta$ from the chain rule. Why do we take $\partial x/\partial r$ as $\cos \theta$ and not $1/\cos \theta$?

30 Compute f_{xx} for $f(x, y) = (ax + by + c)^{10}$. If $x = t$ and $y = t$ compute f_{tt}. True or false: $(\partial f/\partial x)(\partial x/\partial t) = \partial f/\partial t$.

31 Show that $\partial^2 r/\partial x^2 = y^2/r^3$ in two ways:
(1) Find the x derivative of $\partial r/\partial x = x/\sqrt{x^2 + y^2}$
(2) Find the x derivative of $\partial r/\partial x = x/r$ by the chain rule.

32 Reversing x and y in Problem 31 gives $r_{yy} = x^2/r^3$. But show that $r_{xy} = -xy/r^3$.

33 If $\sin z = x + y$ find $(\partial z/\partial x)_y$ in two ways:

(1) Write $z = \sin^{-1}(x + y)$ and compute its derivative.

(2) Take x derivatives of $\sin z = x + y$. Verify that these answers, explicit and implicit, are equal.

34 By direct computation find f_x and f_{xx} and f_{xy} for $f = \sqrt{x^2 + y^2}$.

35 Find a formula for $\partial^2 f/\partial r \partial \theta$ in terms of the x and y derivatives of $f(x, y)$.

36 Suppose $z = f(x, y)$ is solved for x to give $x = g(y, z)$. Is it true that $\partial z/\partial x = 1/(\partial x/\partial z)$? Test on examples.

37 Suppose $z = e^{xy}$ and therefore $x = (\ln z)/y$. Is it true or not that $(\partial z/\partial x) = 1/(\partial x/\partial z)$?

38 If $x = x(t, u, v)$ and $y = y(t, u, v)$ and $z = z(t, u, v)$, find the t derivative of $f(x, y, z)$.

39 The t derivative of $f(x(t, u), y(t, u))$ is in equation (7). What is f_u?

40 (a) For $f = x^2 + y^2 + z^2$ compute $\partial f/\partial x$ (no subscript, x, y, z all independent).

(b) When there is a further relation $z = x^2 + y^2$, use it to remove z and compute $(\partial f/\partial x)_y$.

(c) Compute $(\partial f/\partial x)_y$ using the chain rule $(\partial f/\partial x) + (\partial f/\partial z)(\partial z/\partial x)$.

(d) Why doesn't that chain rule contain $(\partial f/\partial y)(\partial y/\partial x)$?

41 For $f = ax + by$ on the plane $z = 3x + 5y$, find $(\partial f/\partial x)_z$ and $(\partial f/\partial x)_y$ and $(\partial f/\partial z)_x$.

42 The gas law in physics is $PV = nRT$ or a more general relation $F(P, V, T) = 0$. Show that the three derivatives in Example 8 still multiply to give -1. First find $(\partial P/\partial V)_T$ from $\partial F/\partial V + (\partial F/\partial P)(\partial P/\partial V)_T = 0$.

43 If Problem 42 changes to four variables related by $F(x, y, z, t) = 0$, what is the corresponding product of four derivatives?

44 Suppose $x = t + u$ and $y = tu$. Find the t and u derivatives of $f(x, y)$. Check when $f(x, y) = x^2 - 2y$.

45 (a) For $f = r^2 \sin^2\theta$ find f_x and f_y.

(b) For $f = x^2 + y^2$ find f_r and f_θ.

46 On the curve $\sin x + \sin y = 0$, find dy/dx and d^2y/dx^2 by implicit differentiation.

47 (horrible) Suppose $f_{xx} + f_{yy} = 0$. If $x = u + v$ and $y = u - v$ and $f(x, y) = g(u, v)$, find g_u and g_v. Show that $g_{uu} + g_{vv} = 0$.

48 A function has *constant returns to scale* if $f(cx, cy) = cf(x, y)$. When x and y are doubled so are $f = \sqrt{x^2 + y^2}$ and $f = \sqrt{xy}$. In economics, input/output is constant. In mathematics f is *homogeneous* of degree one.

Prove that $x \partial f/\partial x + y \partial f/\partial y = f(x, y)$, by computing the c derivative at $c = 1$. Test this equation on the two examples and find a third example.

49 *True or false*: The directional derivative of $f(r, \theta)$ in the direction of $\mathbf{u}_\theta$ is $\partial f/\partial\theta$.

13.6 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 $(f_{xy})^2$.

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 $f(x_0, y_0) \leqslant f(x, y)$ 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 of f. 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 $f(x_0, y_0)$.

13J If derivatives exist at an interior minimum or maximum, they are zero:

$$\partial f / \partial x = 0 \quad \text{and} \quad \partial f / \partial y = 0 \quad (\textit{together this is } \text{grad } f = 0). \tag{1}$$

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 grad $f = \mathbf{0}$.

Graphically, (x_0, y_0) is the low point of the surface $z = f(x, y)$. Both cross sections in Figure 13.16 touch bottom. The phrase "if derivatives exist" rules out the vertex of a cone, which is a ***rough point***. The absolute value $f = |x|$ has a minimum without $df/dx = 0$, and so does the distance $f = r$. The rough point is $(0, 0)$.
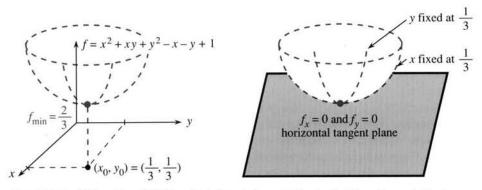

Fig. 13.16 $\partial f / \partial x = 0$ and $\partial f / \partial y = 0$ at the minimum. Quadratic f has linear derivatives.

EXAMPLE 1 Minimize the quadratic $f(x, y) = x^2 + xy + y^2 - x - y + 1$.

To locate the minimum (or maximum), set $f_x = 0$ and $f_y = 0$:

$$f_x = 2x + y - 1 = 0 \quad \text{and} \quad f_y = x + 2y - 1 = 0. \tag{2}$$

Notice what's important: **There are two equations for two unknowns x and y.** Since f is quadratic, the equations are linear. Their solution is $x_0 = \frac{1}{3}$, $y_0 = \frac{1}{3}$ (the stationary point). This is actually a minimum, but to prove that you need to read further.

The constant 1 affects the minimum value $f = \frac{2}{3}$—but not the minimum point. The graph shifts up by 1. The linear terms $-x - y$ affect f_x and f_y. They move the minimum away from $(0, 0)$. The quadratic part $x^2 + xy + y^2$ makes the surface curve upwards. Without that curving part, a plane has its minimum and maximum at boundary points.

EXAMPLE 2 *(Steiner's problem)   Find the point that is nearest to three given points.*

This example is worth your attention. We are locating an airport close to three cities. Or we are choosing a house close to three jobs. The problem is to get as near as possible to the corners of a triangle. The best point depends on the meaning of *"near."*

The distance to the first corner (x_1, y_1) is $d_1 = \sqrt{(x - x_1)^2 + (y - y_1)^2}$. The distances to the other corners (x_2, y_2) and (x_3, y_3) are d_2 and d_3. Depending on whether cost equals *(distance)* or *(distance)²* or *(distance)ᵖ*, our problem will be:

$$\textit{Minimize} \quad d_1 + d_2 + d_3 \quad \textit{or} \quad d_1^2 + d_2^2 + d_3^2 \quad \textit{or even} \quad d_1^p + d_2^p + d_3^p.$$

The second problem is the easiest, when d_1^2 and d_2^2 and d_3^2 are quadratics:

$$f(x, y) = (x - x_1)^2 + (y - y_1)^2 + (x - x_2)^2 + (y - y_2)^2 + (x - x_3)^2 + (y - y_3)^2$$

$$\partial f/\partial x = 2[x - x_1 + x - x_2 + x - x_3] = 0 \qquad \partial f/\partial y = 2[y - y_1 + y - y_2 + y - y_3] = 0.$$

Solving $\partial f/\partial x = 0$ gives $x = \frac{1}{3}(x_1 + x_2 + x_3)$. Then $\partial f/\partial y = 0$ gives $y = \frac{1}{3}(y_1 + y_2 + y_3)$. The best point is the **centroid of the triangle** (Figure 13.17a). It is the nearest point to the corners when the cost is (distance)². Note how squaring makes the derivatives linear. **Least squares** dominates an enormous part of applied mathematics.
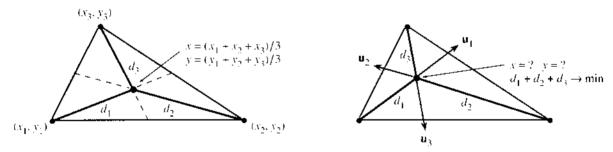

Fig. 13.17   The centroid minimizes $d_1^2 + d_2^2 + d_3^2$. The Steiner point minimizes $d_1 + d_2 + d_3$.

The real *"Steiner problem"* is to minimize $f(x, y) = d_1 + d_2 + d_3$. We are laying down roads from the corners, with cost proportional to length. The equations $f_x = 0$ and $f_y = 0$ look complicated because of square roots. But the nearest point in Figure 13.17b has a remarkable property, which you will appreciate.

Calculus takes derivatives of $d_1^2 = (x - x_1)^2 + (y - y_1)^2$. The x derivative leaves $2d_1(\partial d_1/\partial x) = 2(x - x_1)$. Divide both sides by $2d_1$:

$$\frac{\partial d_1}{\partial x} = \frac{x - x_1}{d_1} \quad \text{and} \quad \frac{\partial d_1}{\partial y} = \frac{y - y_1}{d_1} \quad \text{so grad } d_1 = \left(\frac{x - x_1}{d_1}, \frac{y - y_1}{d_1}\right). \tag{3}$$
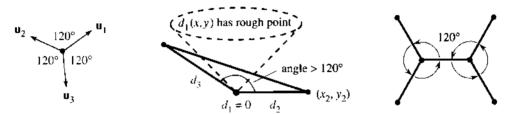
This gradient is a unit vector. The sum of $(x - x_1)^2/d_1^2$ and $(y - y_1)^2/d_1^2$ is $d_1^2/d_1^2 = 1$. This was already in Section 13.4: Distance increases with slope 1 away from the center. The gradient of d_1 (call it $\mathbf{u}_1$) is a unit vector from the center point (x_1, y_1).

Similarly the gradients of d_2 and d_3 are unit vectors $\mathbf{u}_2$ and $\mathbf{u}_3$. They point directly away from the other corners of the triangle. The total cost is $f(x, y) = d_1 + d_2 + d_3$, so we add the gradients. The equations $f_x = 0$ and $f_y = 0$ combine into the vector equation

$$\text{grad } f = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 = 0 \text{ at the minimum.}$$

The three unit vectors add to zero! Moving away from one corner brings us closer to another. The nearest point to the three corners is where those movements cancel. This is the meaning of "grad $f = 0$ at the minimum."

It is unusual for three unit vectors to add to zero—this can only happen in one way. *The three directions must form angles of* $120°$. The best point has this property, which is repeated in Figure 13.18a. The unit vectors cancel each other. At the "Steiner point," the roads to the corners make $120°$ angles. This optimal point solves the problem, except for one more possibility.

Fig. 13.18 Gradients $\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 = 0$ for $120°$ angles. Corner wins at wide angle. *Four corners.* In this case two branchpoints are better—still $120°$.

The other possibility is a minimum at a *rough point*. The graph of the distance function $d_1(x, y)$ is a cone. It has a sharp point at the center (x_1, y_1). All three corners of the triangle are rough points for $d_1 + d_2 + d_3$, so all of them are possible minimizers.

Suppose the angle at a corner exceeds $120°$. Then there is no Steiner point. Inside the triangle, the angle would become even wider. The best point must be a rough point—one of the corners. The winner is the corner with the wide angle. In the figure that means $d_1 = 0$. Then the sum $d_2 + d_3$ comes from the two shortest edges.

Summary The solution is at a $120°$ point or a wide-angle corner. That is the theory. The real problem is to compute the Steiner point—which I hope you will do.

Remark 1 Steiner's problem for *four points* is surprising. We don't minimize $d_1 + d_2 + d_3 + d_4$—there is a better problem. Connect the four points with roads, minimizing their total length, *and allow the roads to branch*. A typical solution is in Figure 13.18c. The angles at the branch points are $120°$. There are at most two branch points (two less than the number of corners).

Remark 2 For other powers p, the cost is $(d_1)^p + (d_2)^p + (d_3)^p$. The x derivative is

$$\partial f / \partial x = p(d_1)^{p-2}(x - x_1) + p(d_2)^{p-2}(x - x_2) + p(d_3)^{p-2}(x - x_3). \tag{4}$$

The key equations are still $\partial f / \partial x = 0$ and $\partial f / \partial y = 0$. Solving them requires a computer and an algorithm. To share the work fairly, I will supply the algorithm (Newton's method) if you supply the computer. Seriously, this is a terrific example. It is typical of real problems—we know $\partial f / \partial x$ and $\partial f / \partial y$ but not the point where they are zero. You can calculate that nearest point, which changes as p changes. You can also discover new mathematics, about how that point moves. I will collect all replies I receive to Problems 38 and 39.

MINIMUM OR MAXIMUM ON THE BOUNDARY

Steiner's problem had no boundaries. The roads could go anywhere. But most applications have restrictions on x and y, like $x \geqslant 0$ or $y \leqslant 0$ or $x^2 + y^2 \geqslant 1$. The minimum with these restrictions is probably higher than the absolute minimum. There are three possibilities:

(1) **stationary point** $f_x = 0, f_y = 0$ (2) **rough point**   (3) **boundary point**

That third possibility requires us to maximize or minimize $f(x, y)$ along the boundary.

EXAMPLE 3   Minimize $f(x, y) = x^2 + xy + y^2 - x - y + 1$ in the *half-plane* $x \geqslant 0$.

The minimum in Example 1 was $\frac{2}{3}$. It occurred at $x_0 = \frac{1}{3}$, $y_0 = \frac{1}{3}$. *This point is still allowed.* It satisfies the restriction $x \geqslant 0$. So the minimum is not moved.

EXAMPLE 4   Minimize the same $f(x, y)$ restricted to the *lower half-plane* $y \leqslant 0$.

Now the absolute minimum at $(\frac{1}{3}, \frac{1}{3})$ is not allowed. There are no rough points. We look for a minimum on the boundary line $y = 0$ in Figure 13.19a. Set $y = 0$, so f depends only on x. Then choose the best x:

$$f(x, 0) = x^2 + 0 - x - 0 + 1 \qquad \text{and} \qquad f_x = 2x - 1 = 0.$$

The minimum is at $x = \frac{1}{2}$ and $y = 0$, where $f = \frac{3}{4}$. This is up from $\frac{2}{3}$.
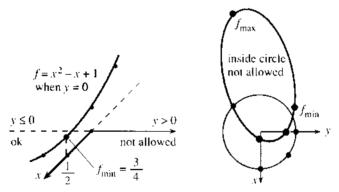

Fig. 13.19   The boundaries $y = 0$ and $x^2 + y^2 = 1$ contain the minimum points.

EXAMPLE 5   *Minimize the same $f(x, y)$ on or outside the circle $x^2 + y^2 = 1$.*

One possibility is $f_x = 0$ and $f_y = 0$. But this is at $(\frac{1}{3}, \frac{1}{3})$, inside the circle. The other possibility is a minimum at a boundary point, *on the circle*.

To follow this boundary we can set $y = \sqrt{1 - x^2}$. The function f gets complicated, and df/dx is worse. There is a way to avoid square roots: Set $x = \cos t$ and $y = \sin t$. Then $f = x^2 + xy + y^2 - x - y + 1$ is a function of the angle t:

$$f(t) = 1 + \cos t \sin t - \cos t - \sin t + 1$$

$$df/dt = \cos^2 t - \sin^2 t + \sin t - \cos t = (\cos t - \sin t)(\cos t + \sin t - 1).$$

Now $df/dt = 0$ locates a minimum or maximum along the boundary. The first factor $(\cos t - \sin t)$ is zero when $x = y$. The second factor is zero when $\cos t + \sin t = 1$, or $x + y = 1$. Those points *on the circle* are the candidates. Problem 24 sorts them out, and Section 13.7 finds the minimum in a new way   using "*Lagrange multipliers.*"

Minimization on a boundary is a serious problem—it gets difficult quickly—and multipliers are ultimately the best solution.

MAXIMUM VS. MINIMUM VS. SADDLE POINT

How to separate the maximum from the minimum? When possible, try all candidates and decide. Compute f at every stationary point and other critical point (maybe also out at infinity), and compare. Calculus offers another approach, based on *second derivatives*.

With one variable the second derivative test was simple: $f_{xx} > 0$ at a minimum, $f_{xx} = 0$ at an inflection point, $f_{xx} < 0$ at a maximum. This is a local test, which may not give a global answer. But it decides whether the slope is increasing (bottom of the graph) or decreasing (top of the graph). We now find a similar test for $f(x, y)$.

The new test involves all three second derivatives. It applies where $f_x = 0$ and $f_y = 0$. The tangent plane is horizontal. *We ask whether the graph of f goes above or below that plane.* The tests $f_{xx} > 0$ and $f_{yy} > 0$ guarantee a minimum in the x and y directions, but there are other directions.

EXAMPLE 6 $f(x, y) = x^2 + 10xy + y^2$ has $f_{xx} = 2$, $f_{xy} = 10$, $f_{yy} = 2$ (*minimum or not?*)

All second derivatives are positive—but wait and see. The stationary point is $(0, 0)$, where $\partial f/\partial x$ and $\partial f/\partial y$ are both zero. Our function is the sum of $x^2 + y^2$, which goes upward, and $10xy$ which has a saddle. The second derivatives must decide whether $x^2 + y^2$ or $10xy$ is stronger.

Along the x axis, where $y = 0$ and $f = x^2$, our point is at the bottom. The minimum in the x direction is at $(0, 0)$. Similarly for the y direction. But $(0, 0)$ is *not a minimum point* for the whole function, because of $10xy$.

Try $x = 1$, $y = -1$. Then $f = 1 - 10 + 1$, which is negative. The graph goes *below* the xy plane in that direction. The stationary point at $x = y = 0$ is a *saddle point*.
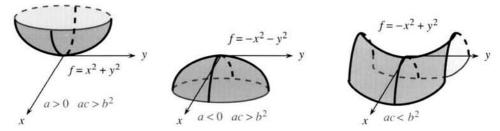

Fig. 13.20 Minimum, maximum, saddle point based on the signs of a and $ac - b^2$.

EXAMPLE 7 $f(x, y) = x^2 + xy + y^2$ has $f_{xx} = 2$, $f_{xy} = 1$, $f_{yy} = 2$ (*minimum or not?*)

The second derivatives 2, 1, 2 are again positive. The graph curves up in the x and y directions. But there is a big difference from Example 6: f_{xy} is reduced from 10 to 1. *It is the size of f_{xy} (not its sign!) that makes the difference.* The extra terms $-x - y + 4$ in Example 1 moved the stationary point to $(\frac{1}{3}, \frac{1}{3})$. The second derivatives are still 2, 1, 2, and they pass the test for a minimum:

13K At $(0, 0)$ the quadratic function $f(x, y) = ax^2 + 2bxy + cy^2$ has a

$$\text{minimum if } \begin{matrix} a > 0 \\ ac > b^2 \end{matrix} \qquad \text{maximum if } \begin{matrix} a < 0 \\ ac > b^2 \end{matrix} \qquad \text{saddle point if } ac < b^2.$$

For a direct proof, split $f(x, y)$ into two parts by "completing the square:"

$$ax^2 + 2bxy + cy^2 = a\left(x + \frac{b}{a}y\right)^2 + \frac{ac - b^2}{a}y^2.$$

That algebra can be checked (notice the $2b$). It is the conclusion that's important:

if $a > 0$ and $ac > b^2$, both parts are positive: **minimum** at $(0, 0)$

if $a < 0$ and $ac > b^2$, both parts are negative: **maximum** at $(0, 0)$

if $ac < b^2$, the parts have opposite signs: **saddle point** at $(0, 0)$.

Since the test involves the *square* of b, its sign has no importance. Example 6 had $b = 5$ and a saddle point. Example 7 had $b = \frac{1}{2}$ and a minimum. Reversing to $-x^2 - xy - y^2$ yields a maximum. So does $-x^2 + xy - y^2$.

Now comes the final step, from $ax^2 + 2bxy + cy^2$ to a general function $f(x, y)$. For all functions, quadratics or not, it is the *second order terms* that we test.

EXAMPLE 8 $f(x, y) = e^x - x - \cos y$ has a stationary point at $x = 0, y = 0$.

The first derivatives are $e^x - 1$ and $\sin y$, both zero. The second derivatives are $f_{xx} = e^x = 1$ and $f_{yy} = \cos y = 1$ and $f_{xy} = 0$. We only use the derivatives *at the stationary point*. The first derivatives are zero, so the second order terms come to the front in the series for $e^x - x - \cos y$:

$$(1 + x + \tfrac{1}{2}x^2 + \cdots) - x - (1 - \tfrac{1}{2}y^2 + \cdots) = \tfrac{1}{2}x^2 + \tfrac{1}{2}y^2 + \text{higher order terms.} \quad (7)$$

There is a *minimum* at the origin. The quadratic part $\tfrac{1}{2}x^2 + \tfrac{1}{2}y^2$ goes upward. The x^3 and y^4 terms are too small to protest. Eventually those terms get large, but near a stationary point it is the quadratic that counts. We didn't need the whole series, because from $f_{xx} = f_{yy} = 1$ and $f_{xy} = 0$ we knew it would start with $\tfrac{1}{2}x^2 + \tfrac{1}{2}y^2$.

> **13L**   The test in **13K** applies to the second derivatives $a = f_{xx}$, $b = f_{xy}$, $c = f_{yy}$ of any $f(x, y)$ at any stationary point. At all points the test decides whether the graph is concave up, concave down, or "indefinite."

EXAMPLE 9 $f(x, y) = e^{xy}$ has $f_x = ye^{xy}$ and $f_y = xe^{xy}$. The stationary point is $(0, 0)$.

The second derivatives at that point are $a = f_{xx} = 0$, $b = f_{xy} = 1$, and $c = f_{yy} = 0$. The test $b^2 > ac$ makes this a saddle point. Look at the infinite series:

$$e^{xy} = 1 + xy + \tfrac{1}{2}x^2y^2 + \cdots.$$

No linear term because $f_x = f_y = 0$: The origin is a *stationary point*. No x^2 or y^2 term (only xy): The stationary point is a *saddle point*.

At $x = 2, y = -2$ we find $f_{xx}f_{yy} > (f_{xy})^2$. The graph is concave up at that point— but it's not a minimum since the first derivatives are not zero.

The series begins with the constant term—not important. Then come the linear terms—extremely important. Those terms are decided by *first* derivatives, and they give the tangent plane. It is only at stationary points—when the linear part disappears and the tangent plane is horizontal—that second derivatives take over. Around any basepoint, *these constant-linear-quadratic terms are the start of the Taylor series*.

THE TAYLOR SERIES

We now put together the whole infinite series. It is a "Taylor series"—which means it is *a power series that matches all derivatives of f* (at the basepoint). For one variable, the powers were x^n when the basepoint was 0. For two variables, the powers are x^n times y^m when the basepoint is $(0, 0)$. Chapter 10 multiplied the nth derivative $d^n f/dx^n$ by $x^n/n!$ *Now every mixed derivative $(\partial/\partial x)^n(\partial/\partial y)^m f(x, y)$ is computed at the basepoint* (subscript $_0$).

We multiply those numbers by $x^n y^m/n!m!$ to match each derivative of $f(x, y)$:

13M When the basepoint is $(0, 0)$, the Taylor series is a double sum $\Sigma\Sigma\, a_{nm}x^n y^m$. The term $a_{nm}x^n y^m$ has the same mixed derivative at $(0, 0)$ as $f(x, y)$. The series is

$$f(0, 0) + x\left(\frac{\partial f}{\partial x}\right)_0 + y\left(\frac{\partial f}{\partial y}\right)_0 + \frac{x^2}{2}\left(\frac{\partial^2 f}{\partial x^2}\right)_0 + xy\left(\frac{\partial^2 f}{\partial x \partial y}\right)_0 + \frac{y^2}{2}\left(\frac{\partial^2 f}{\partial y^2}\right)_0$$

$$+ \sum_{n+m>2}\sum \frac{x^n y^m}{n!m!}\left(\frac{\partial^{n+m} f}{\partial x^n \partial y^m}\right)_0.$$

The derivatives of this series agree with the derivatives of $f(x, y)$ at the basepoint.

The first three terms are the linear approximation to $f(x, y)$. They give the tangent plane at the basepoint. The x^2 term has $n = 2$ and $m = 0$, so $n!m! = 2$. The xy term has $n = m = 1$, and $n!m! = 1$. *The quadratic part $\frac{1}{2}(ax^2 + 2bxy + cy^2)$ is in control when the linear part is zero.*

EXAMPLE 10 All derivatives of e^{x+y} equal one at the origin. The Taylor series is

$$e^{x+y} = 1 + x + y + \frac{x^2}{2} + xy + \frac{y^2}{2} + \cdots = \sum\sum \frac{x^n y^m}{n!m!}$$

This happens to have $ac = b^2$, the special case that was omitted in **13M** and **13N**. *It is the two-dimensional version of an inflection point.* The second derivatives fail to decide the concavity. When $f_{xx}f_{yy} = (f_{xy})^2$, the decision is passed up to the higher derivatives. But in ordinary practice, the Taylor series is stopped after the quadratics.

If the basepoint moves to (x_0, y_0), the powers become $(x - x_0)^n(y - y_0)^m$—and all derivatives are computed at this new basepoint.

***Final question*: *How would you compute a minimum numerically*?** One good way is to solve $f_x = 0$ and $f_y = 0$. These are the functions g and h of Newton's method (Section 13.3). At the current point (x_n, y_n), the derivatives of $g = f_x$ and $h = f_y$ give linear equations for the steps Δx and Δy. Then the next point $x_{n+1} = x_n + \Delta x$, $y_{n+1} = y_n + \Delta y$ comes from those steps. The input is (x_n, y_n), the output is the new point, and the linear equations are

$$\begin{array}{ll}
(g_x)\Delta x + (g_y)\Delta y = -g(x_n, y_n) & (f_{xx})\Delta x + (f_{xy})\Delta y = -f_x(x_n, y_n) \\
(h_x)\Delta x + (h_y)\Delta y = -h(x_n, y_n) \quad \text{or} \quad & (f_{xy})\Delta x + (f_{yy})\Delta y = -f_y(x_n, y_n).
\end{array} \quad (5)$$

When the second derivatives of f are available, use Newton's method.

When the problem is too complicated to go beyond first derivatives, here is an alternative—*steepest descent*. The goal is to move down the graph of $f(x, y)$, like a boulder rolling down a mountain. The steepest direction at any point is given by the *gradient*, with a minus sign to go down instead of up. So move in the direction $\Delta x = -s\, \partial f/\partial x$ and $\Delta y = -s\, \partial f/\partial y$.

The question is: How far to move? Like a boulder, a steep start may not aim directly toward the minimum. The stepsize s is monitored, to end the step when the function f starts upward again (Problem 54). At the end of each step, compute first derivatives and start again in the new steepest direction.

13.6 EXERCISES

Read-through questions

A minimum occurs at a __a__ point (where $f_x = f_y = 0$) or a __b__ point (no derivative) or a __c__ point. Since $f = x^2 - xy + 2y$ has $f_x = $ __d__ and $f_y = $ __e__, the stationary point is $x = $ __f__, $y = $ __g__. This is not a minimum, because f decreases when __h__.

The minimum of $d^2 = (x - x_1)^2 + (y - y_1)^2$ occurs at the rough point __i__. The graph of d is a __j__ and grad d is a __k__ vector that points __l__. The graph of $f = |xy|$ touches bottom along the lines __m__. Those are "rough lines" because the derivative __n__. The maximum of d and f must occur on the __o__ of the allowed region because it doesn't occur __p__.

When the boundary curve is $x = x(t)$, $y = y(t)$, the derivative of $f(x, y)$ along the boundary is __q__ (chain rule). If $f = x^2 + 2y^2$ and the boundary is $x = \cos t$, $y = \sin t$, then $df/dt = $ __r__. It is zero at the points __s__. The maximum is at __t__ and the minimum is at __u__. Inside the circle f has an absolute minimum at __v__.

To separate maximum from minimum from __w__, compute the __x__ derivatives at a __y__ point. The tests for a minimum are __z__. The tests for a maximum are __A__. In case $ac < $ __B__ or $f_{xx}f_{yy} < $ __C__, we have a __D__. At all points these tests decide between concave up and __E__ and "indefinite." For $f = 8x^2 - 6xy + y^2$, the origin is a __F__. The signs of f at $(1, 0)$ and $(1, 3)$ are __G__.

The Taylor series for $f(x, y)$ begins with the six terms __H__. The coefficient of $x^n y^m$ is __I__. To find a stationary point numerically, use __J__ or __K__.

Find all stationary points ($f_x = f_y = 0$) in 1–16. Separate minimum from maximum from saddle point. Test 13K applies to $a = f_{xx}$, $b = f_{xy}$, $c = f_{yy}$.

1 $x^2 + 2xy + 3y^2$
2 $xy - x + y$
3 $x^2 + 4xy + 3y^2 - 6x - 12y$
4 $x^2 - y^2 + 4y$
5 $x^2 y^2 - x$
6 $xe^y - e^x$
7 $-x^2 + 2xy - 3y^2$
8 $(x + y)^2 + (x + 2y - 6)^2$
9 $x^2 + y^2 + z^2 - 4z$
10 $(x + y)(x + 2y - 6)$
11 $(x - y)^2$
12 $(1 + x^2)/(1 + y^2)$
13 $(x + y)^2 - (x + 2y)^2$
14 $\sin x - \cos y$

15 $x^3 + y^3 - 3x^2 + 3y^2$
16 $8xy - x^4 - y^4$

17 A rectangle has sides on the x and y axes and a corner on the line $x + 3y = 12$. Find its maximum area.

18 A box has a corner at $(0, 0, 0)$ and all edges parallel to the axes. If the opposite corner (x, y, z) is on the plane $3x + 2y + z = 1$, what position gives maximum volume? Show first that the problem maximizes $xy - 3x^2y - 2xy^2$.

19 (Straight line fit, Section 11.4) Find x and y to minimize the error
$$E = (x + y)^2 + (x + 2y - 5)^2 + (x + 3y - 4)^2.$$
Show that this gives a minimum not a saddle point.

20 (*Least squares*) What numbers x, y come closest to satisfying the three equations $x - y = 1$, $2x + y = -1$, $x + 2y = 1$? Square and add the errors, $(x - y - 1)^2 + $ _____ + _____. Then minimize.

21 Minimize $f = x^2 + xy + y^2 - x - y$ restricted by
  (a) $x \le 0$ (b) $y \ge 1$ (c) $x \le 0$ and $y \ge 1$.

22 Minimize $f = x^2 + y^2 + 2x + 4y$ in the regions
  (a) all x, y (b) $y \ge 0$ (c) $x \ge 0$, $y \ge 0$

23 Maximize and minimize $f = x + \sqrt{3}y$ on the circle $x = \cos t$, $y = \sin t$.

24 Example 5 followed $f = x^2 + xy + y^2 - x - y + 1$ around the circle $x^2 + y^2 = 1$. The four stationary points have $x = y$ or $x + y = 1$. Compute f at those points and locate the minimum.

25 (a) Maximize $f = ax + by$ on the circle $x^2 + y^2 = 1$.
  (b) Minimize $x^2 + y^2$ on the line $ax + by = 1$.

26 For $f(x, y) = \frac{1}{4}x^4 - xy + \frac{1}{4}y^4$, what are the equations $f_x = 0$ and $f_y = 0$? What are their solutions? What is $f_{\min}$?

27 Choose $c > 0$ so that $f = x^2 + xy + cy^2$ has a saddle point at $(0, 0)$. Note that $f > 0$ on the lines $x = 0$ and $y = 0$ and $y = x$ and $y = -x$, so checking four directions does not confirm a minimum.

Problems 28–42 minimize the Steiner distance $f = d_1 + d_2 + d_3$ and related functions. A computer is needed for 33 and 36–39.

28 Draw the triangle with corners at $(0, 0)$, $(1, 1)$, and $(1, -1)$. By symmetry the Steiner point will be on the x axis. Write down the distances d_1, d_2, d_3 to $(x, 0)$ and find the x that minimizes $d_1 + d_2 + d_3$. Check the $120°$ angles.

29 Suppose three unit vectors add to zero. Prove that the angles between them must be $120°$.

30 In three dimensions, Steiner minimizes the total distance $f(x, y, z) = d_1 + d_2 + d_3 + d_4$ from four points. Show that grad d_1 is still a unit vector (in which direction?) At what angles do four unit vectors add to zero?

31 With four points in a plane, the Steiner problem allows branches (Figure 13.18c). Find the shortest network connecting the corners of a rectangle, if the side lengths are (a) 1 and 2 (b) 1 and 1 (two solutions for a square) (c) 1 and 0.1.

32 Show that a Steiner point ($120°$ angles) can never be outside the triangle.

33 Write a program to minimize $f(x, y) = d_1 + d_2 + d_3$ by Newton's method in equation (5). Fix two corners at $(0, 0)$, $(3, 0)$, vary the third from $(1, 1)$ to $(2, 1)$ to $(3, 1)$ to $(4, 1)$, and compute Steiner points.

34 Suppose one side of the triangle goes from $(-1, 0)$ to $(1, 0)$. Above that side are points from which the lines to $(-1, 0)$ and $(1, 0)$ meet at a $120°$ angle. Those points lie on a circular arc—draw it and find its center and its radius.

35 Continuing Problem 34, there are circular arcs for all three sides of the triangle. On the arcs, every point sees one side of the triangle at a $120°$ angle. Where is the Steiner point? (Sketch three sides with their arcs.)

36 Invent an algorithm to converge to the Steiner point based on Problem 35. Test it on the triangles of Problem 33.

37 Write a code to minimize $f = d_1^4 + d_2^4 + d_3^4$ by solving $f_x = 0$ and $f_y = 0$. Use Newton's method in equation (5).

38 Extend the code to allow all powers $p \geq 1$, not only $p = 4$. Follow the minimizing point from the centroid at $p = 2$ to the Steiner point at $p = 1$ (try $p = 1.8, 1.6, 1.4, 1.2$).

39 Follow the minimizing point with your code as p increases: $p = 2, p = 4, p = 8, p = 16$. Guess the limit at $p = \infty$ and test whether it is equally distant from the three corners.

40 At $p = \infty$ we are making the largest of the distances d_1, d_2, d_3 as small as possible. The best point for a 1, 1, $\sqrt{2}$ right triangle is _____.

41 Suppose the road from corner 1 is wider than the others, and the total cost is $f(x, y) = \sqrt{2}\, d_1 + d_2 + d_3$. Find the gradient of f and the angles at which the best roads meet.

42 Solve Steiner's problem for *two* points. Where is $d_1 + d_2$ a minimum? Solve also for three points if only the three corners are allowed.

Find all derivatives at $(0, 0)$. Construct the Taylor series:

43 $f(x, y) = (x + y)^3$ **44** $f(x, y) = xe^y$

45 $f(x, y) = \ln(1 - xy)$

Find $f_x, f_y, f_{xx}, f_{xy}, f_{yy}$ at the basepoint. Write the quadratic approximation to $f(x, y)$ — the Taylor series through second-order terms:

46 $f = e^{x+y}$ at $(0, 0)$ **47** $f = e^{x+y}$ at $(1, 1)$

48 $f = \sin x \cos y$ at $(0, 0)$ **49** $f = x^2 + y^2$ at $(1, -1)$

50 The Taylor series around (x, y) is also written with steps h and k: $f(x + h, y + k) = f(x, y) + h$ _____ $+ k$ _____ $+ \frac{1}{2}h^2$ _____ $+ hk$ _____ $+ \cdots$. Fill in those four blanks.

51 Find lines along which $f(x, y)$ is constant (these functions have $f_{xx}f_{yy} = f_{xy}^2$ or $ac = b^2$):
     (a) $f = x^2 - 4xy + 4y^2$ (b) $f = e^x e^y$

52 For $f(x, y, z)$ the first three terms after $f(0, 0, 0)$ in the Taylor series are _____. The next six terms are _____.

53 (a) For the error $f - f_L$ in linear approximation, the Taylor series at $(0, 0)$ starts with the quadratic terms _____.
     (b) The graph of f goes up from its tangent plane (and $f > f_L$) if _____. Then f is concave upward.
     (c) For $(0, 0)$ to be a minimum we also need _____

54 The gradient of $x^2 + 2y^2$ at the point $(1, 1)$ is $(2, 4)$. Steepest descent is along the line $x = 1 - 2s, y = 1 - 4s$ (minus sign to go downward). Minimize $x^2 + 2y^2$ with respect to the stepsize s. That locates the next point _____, where steepest descent begins again.

55 Newton's method minimizes $x^2 + 2y^2$ in one step. Starting at $(x_0, y_0) = (1, 1)$, find Δx and Δy from equation (5).

56 If $f_{xx} + f_{yy} = 0$, show that $f(x, y)$ cannot have an interior maximum or minimum (only saddle points).

57 The value of x theorems and y exercises is $f = x^2 y$ (maybe). The most that a student or author can deal with is $4x + y = 12$. Substitute $y = 12 - 4x$ and maximize f. Show that the line $4x + y = 12$ is tangent to the level curve $x^2 y = f_{max}$.

58 The desirability of x houses and y yachts is $f(x, y)$. The constraint $px + qy = k$ limits the money available. The cost of a house is _____, the cost of a yacht is _____. Substitute $y = (k - px)/q$ into $f(x, y) = F(x)$ and use the chain rule for dF/dx. Show that the slope $-f_x/f_y$ at the best x is $-p/q$.

59 At the farthest point in a baseball field, explain why the fence is perpendicular to the line from home plate. Assume it is not a rough point (corner) or endpoint (foul line).

13.7 Constraints and Lagrange Multipliers

This section faces up to a practical problem. We often minimize one function $f(x, y)$ while another function $g(x, y)$ is fixed. There is a **constraint** on x and y, given by $g(x, y) = k$. This restricts the material available or the funds available or the energy available. With this constraint, the problem is to do the best possible (f_{max} or f_{min}).

At the absolute minimum of $f(x, y)$, the requirement $g(x, y) = k$ is probably violated. In that case the minimum point is not allowed. We cannot use $f_x = 0$ and $f_y = 0$—those equations don't account for g.

Step 1 Find equations for the **constrained minimum** or **constrained maximum**. They will involve f_x and f_y and also g_x and g_y, which give local information about f and g. To see the equations, look at two examples.

EXAMPLE 1 Minimize $f = x^2 + y^2$ subject to the constraint $g = 2x + y = k$.

Trial runs The constraint allows $x = 0$, $y = k$, where $f = k^2$. Also $(\frac{1}{2}k, 0)$ satisfies the constraint, and $f = \frac{1}{4}k^2$ is smaller. Also $x = y = \frac{1}{3}k$ gives $f = \frac{2}{9}k^2$ (best so far).

Idea of solution Look at the level curves of $f(x, y)$ in Figure 13.21. They are circles $x^2 + y^2 = c$. When c is small, the circles do not touch the line $2x + y = k$. There are no points that satisfy the constraint, when c is too small. *Now increase c.*

Eventually the growing circles $x^2 + y^2 = c$ will *just touch the line* $x + 2y = k$. The point where they touch is the winner. It gives the smallest value of c that can be achieved on the line. The touching point is (x_{min}, y_{min}), and the value of c is f_{min}.

What equation describes that point? When the circle touches the line, they are *tangent*. They have the same slope. *The perpendiculars to the circle and the line go in the same direction.* That is the key fact, which you see in Figure 13.21a. The direction perpendicular to $f = c$ is given by $\text{grad} f = (f_x, f_y)$. The direction perpendicular to $g = k$ is given by $\text{grad} g = (g_x, g_y)$. The key equation says that those two vectors are parallel. One gradient vector is a multiple of the other gradient vector, with a multiplier λ (called lambda) that is unknown:

13N At the minimum of $f(x, y)$ subject to $g(x, y) = k$, the gradient of f is parallel to the gradient of g—with an unknown number λ as the multiplier:

$$\text{grad} f = \lambda \text{ grad } g \quad \text{so} \quad \frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x} \quad \text{and} \quad \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}. \tag{1}$$

Step 2 There are now three unknowns x, y, λ. There are also *three equations*:

$$\partial f / \partial x = \lambda \partial g / \partial x \text{ is} \qquad 2x = 2\lambda$$

$$\partial f / \partial y = \lambda \partial g / \partial y \text{ is} \qquad 2y = \lambda \tag{2}$$

$$g(x, y) = k \text{ is} \qquad 2x + y = k.$$

In the third equation, substitute 2λ for $2x$ and $\frac{1}{2}\lambda$ for y. Then $2x + y$ equals $\frac{5}{2}\lambda$ equals k. Knowing $\lambda = \frac{2}{5}k$, go back to the first two equations for x, y, and f_{min}:

$$x = \lambda = \frac{2}{5}k, \quad y = \frac{1}{2}\lambda = \frac{1}{5}k, \quad f_{min} = \left(\frac{2}{5}k\right)^2 + \left(\frac{1}{5}k\right)^2 = \frac{5}{25}k^2 = \frac{1}{5}k^2.$$

The winning point (x_{min}, y_{min}) is $(\frac{2}{5}k, \frac{1}{5}k)$. It minimizes the "distance squared," $f = x^2 + y^2 = \frac{1}{5}k^2$, from the origin to the line.

Question *What is the meaning of the Lagrange multiplier λ?*

Mysterious answer *The derivative of $\frac{1}{3}k^2$ is $\frac{2}{3}k$, which equals λ. The multiplier λ is the derivative of f_{min} with respect to k.* Move the line by Δk, and f_{min} changes by about $\lambda \Delta k$. Thus the Lagrange multiplier measures the *sensitivity* to k.

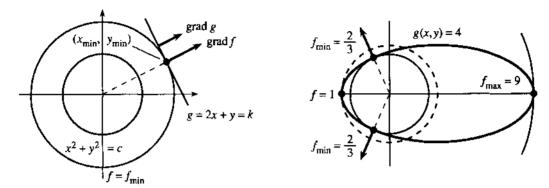Pronounce his name "Lagronge" or better "Lagrongh" as if you are French.

Fig. 13.21 Circles $f = c$ tangent to line $g = k$ and ellipse $g = 4$: parallel gradients.

EXAMPLE 2 Maximize and minimize $f = x^2 + y^2$ on the ellipse $g = (x-1)^2 + 4y^2 = 4$.

Idea and equations The circles $x^2 + y^2 = c$ grow until they touch the ellipse. The touching point is (x_{min}, y_{min}) and that smallest value of c is f_{min}. As the circles grow they cut through the ellipse. Finally there is a point (x_{max}, y_{max}) where the last circle touches. That largest value of c is f_{max}.

The minimum and maximum are described by the same rule: the circle is tangent to the ellipse (Figure 13.21b). *The perpendiculars go in the same direction.* Therefore (f_x, f_y) is a multiple of (g_x, g_y), and the unknown multiplier is λ:

$$f_x = \lambda g_x: \qquad\qquad 2x = \lambda 2(x-1)$$
$$f_y = \lambda g_y: \qquad\qquad 2y = \lambda 8y \qquad\qquad (3)$$
$$g = k: \qquad (x-1)^2 + 4y^2 = 4.$$

Solution The second equation allows two possibilities: $y = 0$ or $\lambda = \frac{1}{4}$. Following up $y = 0$, the last equation gives $(x-1)^2 = 4$. Thus $x = 3$ or $x = -1$. Then the first equation gives $\lambda = 3/2$ or $\lambda = 1/2$. The values of f are $x^2 + y^2 = 3^2 + 0^2 = 9$ and $x^2 + y^2 = (-1)^2 + 0^2 = 1$.

Now follow $\lambda = 1/4$. The first equation yields $x = -1/3$. Then the last equation requires $y^2 = 5/9$. Since $x^2 = 1/9$ we find $x^2 + y^2 = 6/9 = 2/3$. This is f_{min}.

Conclusion The equations (3) have four solutions, at which the circle and ellipse are tangent. The four points are $(3, 0)$, $(-1, 0)$, $(-1/3, \sqrt{5}/3)$, and $(-1/3, -\sqrt{5}/3)$. The four values of f are 9, 1, $\frac{2}{3}$, $\frac{2}{3}$.

Summary The three equations are $f_x = \lambda g_x$ and $f_y = \lambda g_y$ and $g = k$. The unknowns are x, y, and λ. There is no absolute system for solving the equations (unless they are linear; then use elimination or Cramer's Rule). Often the first two equations yield x and y in terms of λ, and substituting into $g = k$ gives an equation for λ.

At the minimum, the level curve $f(x, y) = c$ is tangent to the constraint curve $g(x, y) = k$. If that constraint curve is given parametrically by $x(t)$ and $y(t)$, then

minimizing $f(x(t), y(t))$ uses the chain rule:

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = 0 \quad \text{or} \quad (\text{grad } f)\cdot(\text{tangent to curve}) = 0.$$

This is the calculus proof that grad f is perpendicular to the curve. *Thus* grad f *is parallel to* grad g. This means $(f_x, f_y) = \lambda(g_x, g_y)$.

We have lost $f_x = 0$ and $f_y = 0$. But a new function L has *three* zero derivatives:

13O   The Lagrange function is $L(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - k)$. Its three derivatives are $L_x = L_y = L_\lambda = 0$ at the solution:

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} - \lambda\frac{\partial g}{\partial x} = 0 \qquad \frac{\partial L}{\partial y} = \frac{\partial f}{\partial y} - \lambda\frac{\partial g}{\partial y} = 0 \qquad \frac{\partial L}{\partial \lambda} = -g + k = 0. \tag{4}$$

Note that $\partial L/\partial \lambda = 0$ automatically produces $g = k$. The constraint is "**built in**" to L. *Lagrange has included a term* $\lambda(g - k)$, which is destined to be zero—but its derivatives are absolutely needed in the equations! At the solution, $g = k$ and $L = f$ and $\partial L/\partial k = \lambda$.

What is important is $f_x = \lambda g_x$ and $f_y = \lambda g_y$, coming from $L_x = L_y = 0$. In words: The constraint $g = k$ forces $dg = g_x dx + g_y dy = 0$. *This restricts the movements dx and dy.* They must keep to the curve. The equations say that $df = f_x dx + f_y dy$ is equal to λdg. Thus df is zero *in the allowed direction*—which is the key point.

MAXIMUM AND MINIMUM WITH TWO CONSTRAINTS

The whole subject of min(max)imization is called **optimization**. Its applications to business decisions make up *operations research*. The special case of linear functions is always important—in this part of mathematics it is called *linear programming*. A book about those subjects won't fit inside a calculus book, but we can take one more step—to allow a second constraint.

The function to minimize or maximize is now $f(x, y, z)$. The constraints are $g(x, y, z) = k_1$ and $h(x, y, z) = k_2$. The multipliers are λ_1 and λ_2. We need at least three variables x, y, z because two constraints would completely determine x and y.

13P   To minimize $f(x, y, z)$ subject to $g(x, y, z) = k_1$ and $h(x, y, z) = k_2$, solve five equations for $x, y, z, \lambda_1, \lambda_2$. Combine $g = k_1$ and $h = k_2$ with

$$\frac{\partial f}{\partial x} = \lambda_1\frac{\partial g}{\partial x} + \lambda_2\frac{\partial h}{\partial x}, \qquad \frac{\partial f}{\partial y} = \lambda_1\frac{\partial g}{\partial y} + \lambda_2\frac{\partial h}{\partial y}, \qquad \frac{\partial f}{\partial z} = \lambda_1\frac{\partial g}{\partial z} + \lambda_2\frac{\partial h}{\partial z}. \tag{5}$$

Figure 13.22a shows the geometry behind these equations. For convenience f is $x^2 + y^2 + z^2$, so we are minimizing distance (squared). The constraints $g = x + y + z = 9$ and $h = x + 2y + 3z = 20$ are linear—their graphs are planes. The constraints keep (x, y, z) on both planes—and therefore on the line where they meet. We are finding *the squared distance from* $(0, 0, 0)$ *to a line*.

What equation do we solve? The level surfaces $x^2 + y^2 + z^2 = c$ are spheres. They grow as c increases. The first sphere to touch the line is tangent to it. That touching point gives the solution (the smallest c). *All three vectors* grad f, grad g, grad h *are perpendicular to the line*:

line tangent to sphere $\Rightarrow$ grad f perpendicular to line

line in both planes $\Rightarrow$ grad g and grad h perpendicular to line.

Thus grad f, grad g, grad h are *in the same plane*—perpendicular to the line. With three vectors in a plane, grad f is a combination of grad g and grad h:

$$(f_x, f_y, f_z) = \lambda_1(g_x, g_y, g_z) + \lambda_2(h_x, h_y, h_z). \tag{6}$$

This is the key equation (5). It applies to curved surfaces as well as planes.

EXAMPLE 3 Minimize $x^2 + y^2 + z^2$ when $x + y + z = 9$ and $x + 2y + 3z = 20$.
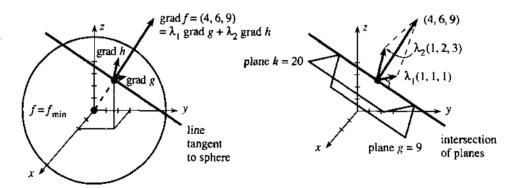
In Figure 13.22b, the normals to those planes are grad $g = (1, 1, 1)$ and grad $h = (1, 2, 3)$. The gradient of $f = x^2 + y^2 + z^2$ is $(2x, 2y, 2z)$. The equations (5)–(6) are

$$2x = \lambda_1 + \lambda_2, \quad 2y = \lambda_1 + 2\lambda_2, \quad 2z = \lambda_1 + 3\lambda_2.$$

Substitute these x, y, z into the other two equations $g = x + y + z = 9$ and $h = 20$:

$$\frac{\lambda_1 + \lambda_2}{2} + \frac{\lambda_1 + 2\lambda_2}{2} + \frac{\lambda_1 + 3\lambda_2}{2} = 9 \quad \text{and} \quad \frac{\lambda_1 + \lambda_2}{2} + 2\frac{\lambda_1 + 2\lambda_2}{2} + 3\frac{\lambda_1 + 3\lambda_2}{2} = 20.$$

After multiplying by 2, these simplify to $3\lambda_1 + 6\lambda_2 = 18$ and $6\lambda_1 + 14\lambda_2 = 40$. The solutions are $\lambda_1 = 2$ and $\lambda_2 = 2$. Now the previous equations give $(x, y, z) = (2, 3, 4)$.

The Lagrange function with two constraints is $L(x, y, z, \lambda_1, \lambda_2) = f - \lambda_1(g - k_1) - \lambda_2(h - k_2)$. Its five derivatives are zero—those are our five equations. Lagrange has increased the number of unknowns from 3 to 5, by adding λ_1 and λ_2. The best point $(2, 3, 4)$ gives $f_{\min} = 29$. The λ's give $\partial f/\partial k$—the sensitivity to changes in 9 and 20.

Fig. 13.22 Perpendicular vector grad f is a combination λ_1 grad $g + \lambda_2$ grad h.

INEQUALITY CONSTRAINTS

In practice, applications involve *inequalities* as well as equations. The constraints might be $g \leqslant k$ and $h \geqslant 0$. The first means: It is not required to use the whole resource k, but you cannot use more. The second means: h measures a quantity that cannot be negative. *At the minimum point, the multipliers must satisfy the same inequalities*: $\lambda_1 \leqslant 0$ and $\lambda_2 \geqslant 0$. There are inequalities on the λ's when there are inequalities in the constraints.

Brief reasoning: With $g \leqslant k$ the minimum can be *on or inside* the constraint curve. Inside the curve, where $g < k$, we are free to move in all directions. The constraint is not really constraining. This brings back $f_x = 0$ and $f_y = 0$ and $\lambda = 0$—an ordinary minimum. On the curve, where $g = k$ constrains the minimum from going lower, we have $\lambda < 0$. We don't know in advance which to expect.

For 100 constraints $g_i \leqslant k_i$, there are 100 λ's. Some λ's are zero (when $g_i < k_i$) and some are nonzero (when $g_i = k_i$). It is those 2^{100} possibilities that make optimization interesting. In *linear programming* with two variables, the constraints are $x \geqslant 0$, $y \geqslant 0$:

EXAMPLE 4   Minimize $f = 5x + 6y$ with $g = x + y = 4$ and $h = x \geqslant 0$ and $H = y \geqslant 0$.

The constraint $g = 4$ is an equation, h and H yield inequalities. Each has its own Lagrange multiplier—and the inequalities require $\lambda_2 \geqslant 0$ and $\lambda_3 \geqslant 0$. The derivatives of f, g, h, H are no problem to compute:

$$\frac{\partial f}{\partial x} = \lambda_1 \frac{\partial g}{\partial x} + \lambda_2 \frac{\partial h}{\partial x} + \lambda_3 \frac{\partial H}{\partial x} \quad \text{yields} \quad 5 = \lambda_1 + \lambda_2$$

$$\frac{\partial f}{\partial y} = \lambda_1 \frac{\partial g}{\partial y} + \lambda_2 \frac{\partial h}{\partial y} + \lambda_3 \frac{\partial H}{\partial y} \quad \text{yields} \quad 6 = \lambda_1 + \lambda_3. \tag{7}$$

Those equations make λ_3 larger than λ_2. Therefore $\lambda_3 > 0$, which means that the constraint on H must be an equation. (Inequality for the multiplier means equality for the constraint.) In other words $H = y = 0$. Then $x + y = 4$ leads to $x = 4$. The solution is at $(x_{\min}, y_{\min}) = (4, 0)$, where $f_{\min} = 20$.

At this minimum, $h = x = 4$ is above zero. The multiplier for the constraint $h \geqslant 0$ must be $\lambda_2 = 0$. Then the first equation gives $\lambda_1 = 5$. As always, the multiplier measures sensitivity. When $g = 4$ is increased by Δk, the cost $f_{\min} = 20$ is increased by $5\Delta k$. In economics $\lambda_1 = 5$ is called a *shadow price*—it is the cost of increasing the constraint.

Behind this example is a nice problem in geometry. The constraint curve $x + y = 4$ is a line. The inequalities $x \geqslant 0$ and $y \geqslant 0$ leave a piece of that line—from P to Q in Figure 13.23. The level curves $f = 5x + 6y = c$ move out as c increases, until they touch the line. *The first touching point is $Q = (4, 0)$, which is the solution. It is always an endpoint*—or a corner of the triangle PQR. It gives the smallest cost $f_{\min}$, which is $c = 20$.
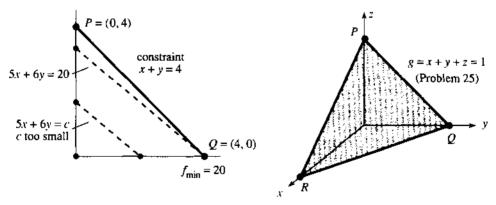

Fig. 13.23   Linear programming: f and g are linear, inequalities cut off x and y.

13.7   EXERCISES

Read-through questions

A restriction $g(x, y) = k$ is called a __a__. The minimizing equations for $f(x, y)$ subject to $g = k$ are __b__. The number λ is the Lagrange __c__. Geometrically, grad f is __d__ to grad g at the minimum. That is because the __e__ curve $f =$

$f_{\min}$ is __f__ to the constraint curve $g = k$. The number λ turns out to be the derivative of __g__ with respect to __h__. The Lagrange function is $L =$ __i__ and the three equations for x, y, λ are __j__ and __k__ and __l__.

To minimize $f = x^2 - y$ subject to $g = x - y = 0$, the three equations for x, y, λ are __m__. The solution is __n__. In this example the curve $f(x, y) = f_{min} =$ __o__ is a __p__ which is __q__ to the line $g = 0$ at (x_{min}, y_{min}).

With two constraints $g(x, y, z) = k_1$ and $h(x, y, z) = k_2$ there are __r__ multipliers. The five unknowns are __s__. The five equations are __t__. The level surface $f = f_{min}$ is __u__ to the curve where $g = k_1$ and $h = k_2$. Then grad f is __v__ to this curve, and so are grad g and __w__. Thus __x__ is a combination of grad g and __y__. With nine variables and six constraints, there will be __z__ multipliers and eventually __A__ equations. If a constraint is an __B__ $g \leqslant k$, then its multiplier must satisfy $\lambda \leqslant 0$ at a minimum.

1 Example 1 minimized $f = x^2 + y^2$ subject to $2x + y = k$. Solve the constraint equation for $y = k - 2x$, substitute into f, and minimize this function of x. The minimum is at $(x, y) =$ _____, where $f =$ _____.

Note: This direct approach *reduces* to one unknown x. Lagrange *increases* to x, y, λ. But Lagrange is better when the first step of solving for y is difficult or impossible.

Minimize and maximize $f(x, y)$ in 2–6. Find x, y, and λ.

2 $f = x^2 y$ with $g = x^2 + y^2 = 1$

3 $f = x + y$ with $g = \dfrac{1}{x} + \dfrac{1}{y} = 1$

4 $f = 3x + y$ with $g = x^2 + 9y^2 = 1$

5 $f = x^2 + y^2$ with $g = x^6 + y^6 = 2$.

6 $f = x + y$ with $g = x^{1/3} y^{2/3} = k$. With $x =$ capital and $y =$ labor, g is a Cobb-Douglas function in economics. Draw two of its level curves.

7 Find the point on the circle $x^2 + y^2 = 13$ where $f = 2x - 3y$ is a maximum. Explain the answer.

8 Maximize $ax + by + cz$ subject to $x^2 + y^2 + z^2 = k^2$. Write your answer as the Schwarz inequality for dot products: $(a, b, c) \cdot (x, y, z) \leqslant$ _____ k.

9 Find the plane $z = ax + by + c$ that best fits the points $(x, y, z) = (0, 0, 1), (1, 0, 0), (1, 1, 2), (0, 1, 2)$. The answer a, b, c minimizes the sum of $(z - ax - by - c)^2$ at the four points.

10 The base of a triangle is the top of a rectangle (5 sides, combined area = 1). What dimensions minimize the distance around?

11 Draw the hyperbola $xy = -1$ touching the circle $g = x^2 + y^2 = 2$. The minimum of $f = xy$ on the circle is reached at the points _____. The equations $f_x = \lambda g_x$ and $f_y = \lambda g_y$ are satisfied at those points with $\lambda =$ _____.

12 Find the maximum of $f = xy$ on the circle $g = x^2 + y^2 = 2$ by solving $f_x = \lambda g_x$ and $f_y = \lambda g_y$ and substituting x and y into f. Draw the level curve $f = f_{max}$ that touches the circle.

13 Draw the level curves of $f = x^2 + y^2$ with a closed curve C across them to represent $g(x, y) = k$. Mark a point where C crosses a level curve. Why is that point not a minimum of f on C? Mark a point where C is *tangent* to a level curve. Is that the minimum of f on C?

14 On the circle $g = x^2 + y^2 = 1$, Example 5 of 13.6 minimized $f = xy - x - y$. (a) Set up the three Lagrange equations for x, y, λ. (b) The first two equations give $x = y =$ _____. (c) There is another solution for the special value $\lambda = -\frac{1}{2}$, when the equations become _____. This is easy to miss but it gives $f_{min} = -1$ at the point _____.

Problems 15–18 develop the theory of Lagrange multipliers.

15 (*Sensitivity*) Certainly $L = f - \lambda(g - k)$ has $\partial L/\partial k = \lambda$. Since $L = f_{min}$ and $g = k$ at the minimum point, this seems to prove the key formula $df_{min}/dk = \lambda$. But x_{min}, y_{min}, λ, and f_{min} all change with k. We need the *total* derivative of $L(x, y, \lambda, k)$:

$$\frac{dL}{dk} = \frac{\partial L}{\partial x}\frac{dx}{dk} + \frac{\partial L}{\partial y}\frac{dy}{dk} + \frac{\partial L}{\partial \lambda}\frac{d\lambda}{dk} + \frac{\partial L}{\partial k}\frac{dk}{dk}.$$

Equation (1) at the minimum point should now yield the sensitivity formula $df_{min}/dk = \lambda$.

16 (Theory behind λ) When $g(x, y) = k$ is solved for y, it gives a curve $y = R(x)$. Then minimizing $f(x, y)$ along this curve yields

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dR}{dx} = 0, \quad \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}\frac{dR}{dx} = 0.$$

Those come from the _____ rule: $df/dx = 0$ at the minimum and $dg/dx = 0$ along the curve because $g =$ _____. Multiplying the second equation by $\lambda = (\partial f/\partial y)/(\partial g/\partial y)$ and subtracting from the first gives _____ $= 0$. Also $\partial f/\partial y = \lambda \partial g/\partial y$. These are the equations (1) for x, y, λ.

17 (*Example of failure*) $\lambda = f_y/g_y$ breaks down if $g_y = 0$ at the minimum point.

(a) $g = x^2 - y^3 = 0$ does not allow negative y because _____.

(b) When $g = 0$ the minimum of $f = x^2 + y$ is at the point _____.

(c) At that point $f_y = \lambda g_y$ becomes _____ which is impossible.

(d) Draw the pointed curve $g = 0$ to see why it is not tangent to a level curve of f.

18 (*No maximum*) Find a point on the line $g = x + y = 1$ where $f(x, y) = 2x + y$ is greater than 100 (or 1000). Write out grad $f = \lambda$ grad g to see that there is no solution.

19 Find the minimum of $f = x^2 + 2y^2 + z^2$ if (x, y, z) is restricted to the planes $g = x + y + z = 0$ and $h = x - z = 1$.

20 (a) Find by Lagrange multipliers the volume $V = xyz$ of the largest box with sides adding up to $x + y + z = k$. (b) Check that $\lambda = dV_{max}/dk$. (c) United Airlines accepts baggage with $x + y + z = 108"$. If it changes to $111"$, approximately how much (by λ) and exactly how much does V_{max} increase?

21 The planes $x = 0$ and $y = 0$ intersect in the line $x = y = 0$, which is the z axis. Write down a vector perpendicular to the plane $x = 0$ and a vector perpendicular to the plane $y = 0$. Find λ_1 times the first vector plus λ_2 times the second. This combination is perpendicular to the line _____.

22 Minimize $f = x^2 + y^2 + z^2$ on the plane $ax + by + cz = d$—one constraint and one multiplier. Compare f_{min} with the distance formula $|d|/\sqrt{a^2 + b^2 + c^2}$ in Section 11.2.

23 At the absolute minimum of $f(x, y)$, the derivatives _____ are zero. If this point happens to fall on the curve $g(x, y) = k$ then the equations $f_x = \lambda g_x$ and $f_y = \lambda g_y$ hold with $\lambda = $ _____.

Problems 24–33 allow inequality constraints, optional but good.

24 Find the minimum of $f = 3x + 5y$ with the constraints $g = x + 2y = 4$ and $h = x \geqslant 0$ and $H = y \geqslant 0$, using equations like (7). Which multiplier is zero?

25 Figure 13.23 shows the constraint plane $g = x + y + z = 1$ chopped off by the inequalities $x \geqslant 0$, $y \geqslant 0$, $z \geqslant 0$. What are the three "endpoints" of this triangle? Find the minimum and maximum of $f = 4x - 2y + 5z$ on the triangle, by testing f at the endpoints.

26 With an inequality constraint $g \leqslant k$, the multiplier at the minimum satisfies $\lambda \leqslant 0$. If k is increased, f_{min} goes down (since $\lambda = df_{min}/dk$). *Explain the reasoning*: By increasing k, (more) (fewer) points satisfy the constraints. Therefore (more) (fewer) points are available to minimize f. Therefore f_{min} goes (up) (down).

27 With an inequality constraint $g \leqslant k$, the multiplier at a *maximum* point satisfies $\lambda \geqslant 0$. Change the reasoning in **26**.

28 When the constraint $h \geqslant k$ is a strict inequality $h > k$ at the minimum, the multiplier is $\lambda = 0$. *Explain the reasoning*: For a small increase in k, the same minimizer is still available (since $h > k$ leaves room to move). Therefore f_{min} is (changed)(unchanged), and $\lambda = df_{min}/dk$ is _____.

29 Minimize $f = x^2 + y^2$ subject to the inequality constraint $x + y \leqslant 4$. The minimum is obviously at _____, where f_x and f_y are zero. The multiplier is $\lambda = $ _____. A small change from 4 will leave $f_{min} = $ _____ so the sensitivity df_{min}/dk still equals λ.

30 Minimize $f = x^2 + y^2$ subject to the inequality constraint $x + y \geqslant 4$. Now the minimum is at _____ and the multiplier is $\lambda = $ _____ and $f_{min} = $ _____. A small change to $4 + dk$ changes f_{min} by what multiple of dk?

31 Minimize $f = 5x + 6y$ with $g = x + y = 4$ and $h = x \geqslant 0$ and $H = y \leqslant 0$. Now $\lambda_3 \leqslant 0$ and the sign change destroys Example 4. Show that equation (7) has no solution, and choose x, y to make $5x + 6y < -1000$.

32 Minimize $f = 2x + 3y + 4z$ subject to $g = x + y + z = 1$ and $x, y, z \geqslant 0$. These constraints have multipliers $\lambda_2 \geqslant 0$, $\lambda_3 \geqslant 0$, $\lambda_4 \geqslant 0$. The equations are $2 = \lambda_1 + \lambda_2$, _____, and $4 = \lambda_1 + \lambda_4$. Explain why $\lambda_3 > 0$ and $\lambda_4 > 0$ and $f_{min} = 2$.

33 A wire 40" long is used to enclose *one or two* squares (side x and side y). Maximize the total area $x^2 + y^2$ subject to $x \geqslant 0$, $y \geqslant 0$, $4x + 4y = 40$.

Resource: Calculus Online Textbook
Gilbert Strang