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 $(0, 0)$.

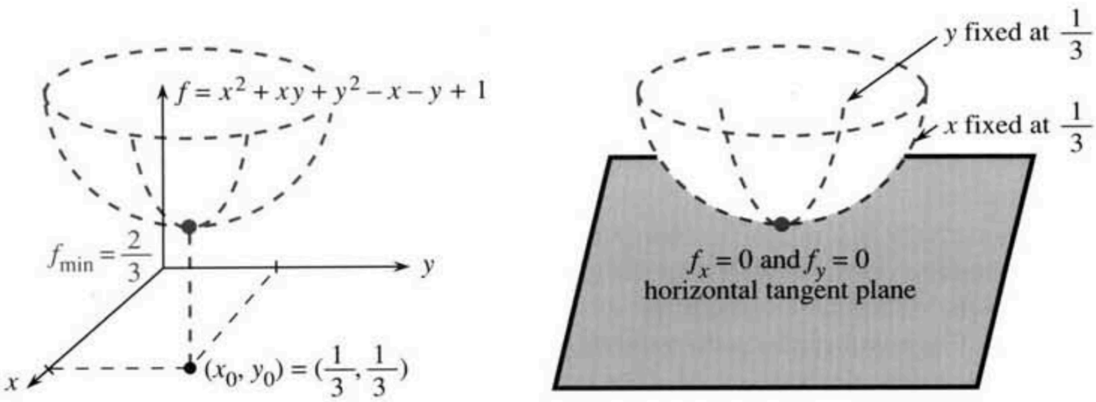


Figure 21: $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$ at the minimum. Quadratic f has linear derivatives.

Exemple 46. Minimize the quadratic $f(x, y) = x^2 + xy + y^2 - x - y + 1$.

Answer 46 To locate the minimum (or maximum), set $f_x = 0$ and $f_y = 0$

$$f_x = 2x + y - 1 = 0 \quad \text{and} \quad f_y = x + 2y - 1 = 0 \tag{8.1}$$

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 $x_0 = 1/3, y_0 = 1/3$ (the stationary point). This is actually a minimum, but to prove that you need to read further.

The constant 1 affects the minimum value $f = 2/3$, 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.

Exemple 47. (Steiner's problem) Find the point that is nearest to three given points.

Answer 47 This example is worth your attention. We are locating an airport dose 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 :

$$\textbf{Minimize} \quad d_1 + d_2 + d_3 \quad \text{or} \quad d_1^2 + d_2^2 + d_3^2 \quad \text{or even} \quad d_1^p + d_2^p + d_3^p \tag{8.2}$$

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

$$f(x, y) = (x - x_1)^2 + (y - y_1)^2 + (x - x_2)^2 + (y - y_2)^2 + (x - x_3)^2 + (y - y_3)^2 \tag{8.3}$$

then

$$\frac{\partial f}{\partial x} = 2(x - x_1 + x - x_2 + x - x_3) = 0 \quad \text{and} \quad \frac{\partial f}{\partial y} = 2(y - y_1 + y - y_2 + y - y_3) = 0$$

Solving $\partial f/\partial x = 0$ gives $x = 1/3(x_1 + x_2 + x_3)$. Then $\partial f/\partial y = 0$ gives $y = 1/3(y_1 + y_2 + y_3)$. The best point is the **centroid of the triangle** (**Figure 22 (left)**). 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.

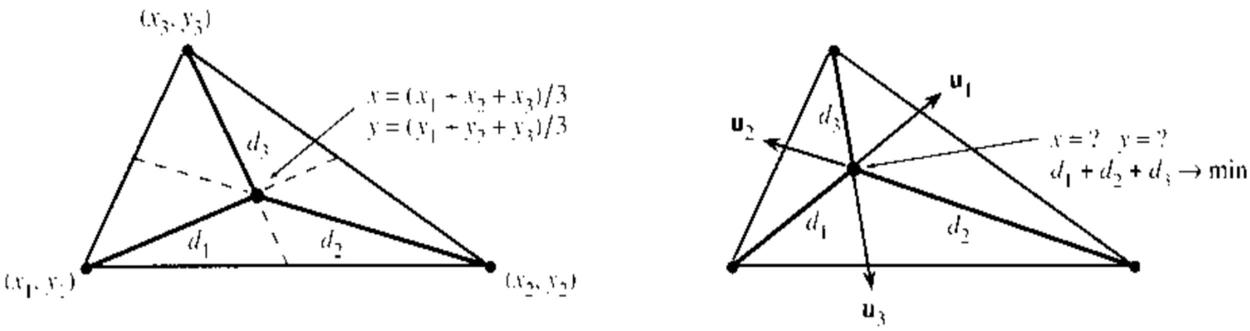


Figure 22: The centroid minimizes $d_1^2 + d_2^2 + d_3^2$. The Steiner point minimizes $d_1 d_2 + d_3$.

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 22 (right)** 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 $2d_1$:

$$\frac{\partial d_1}{\partial x} = \frac{x - x_1}{d_1} \quad \text{and} \quad \frac{\partial d_1}{\partial y} = \frac{y - y_1}{d_1} \quad \text{so} \quad \textbf{grad}\, d_1 = \left(\frac{x - x_1}{d_1}, \frac{y - y_1}{d_1}\right) \tag{8.4}$$

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 the previosu section : Distance increases with slope 1 away from the center. The gradient of d_1 (call it $\boldsymbol{u}_1$) is a unit vector from the center point (x_1, y_1).

Similarly the gradients of d_2 and d_3 are unit vectors $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$. 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

$$\boldsymbol{\nabla} f = \boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = \mathbf{0} \quad \textbf{at the minimum}$$

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 « $\boldsymbol{\nabla} f = \mathbf{0}$ **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 23 (left)**. 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.

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.

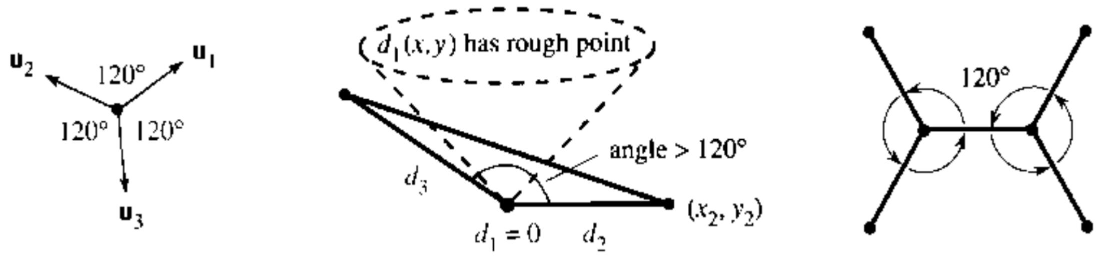


Figure 23: Gradients $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = \boldsymbol{0}$ for 120° angles. Corner wins at wide angle. *Four corners.* In this case two branch points are better-still 120°.

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.

Remarque 2. 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 **Figure 23 (right)**. The angles at the branch points are 120°. There are at most two branch points (two less than the number of corners).

Remarque 3. For other powers p, the cost is $d_1^p + d_2^p + d_3^p$. The x derivative is

$$\frac{\partial df}{\partial x} = p\, d_1^{p-2}(x - x_1) + p\, d_2^{p-2}(x - x_2) + p\, d_3^{p-2}(x - x_3) \tag{8.5}$$

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 $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$ 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.

8.2 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 $x \geq 0$ or $y \leq 0$ or $x^2 + y^2 \geq 1$. 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.

Exemple 48. Minimize $f(x,y) = x^2 + xy + y^2 - x - y + 1$ in the half plane $x \geq 0$.

Answer 48 The minimum in **Example 46** was $2/3$. It occurred at $x_0 = 1/3, y_0 = 1/3$. *This point is still allowed*. It satisfies the restriction $x \geq 0$. So the minimum is not moved.

Exemple 49. Minimize the same $f(x,y)$ restricted to the lower half-plane $y \leq 0$.

Answer 49 Now the absolute minimum at $(1/3, 1/3)$ is not allowed. There are no rough points. We look for a minimum on the boundary line $y = 0$ in **Figure 24**. Set $y = 0$, so f depends only on x. Then choose the best x :

$$f(x,0) = x^2 + 0 - x - 0 + 1 \quad \text{and} \quad f_x = 2x - 1 = 0 \tag{8.6}$$

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

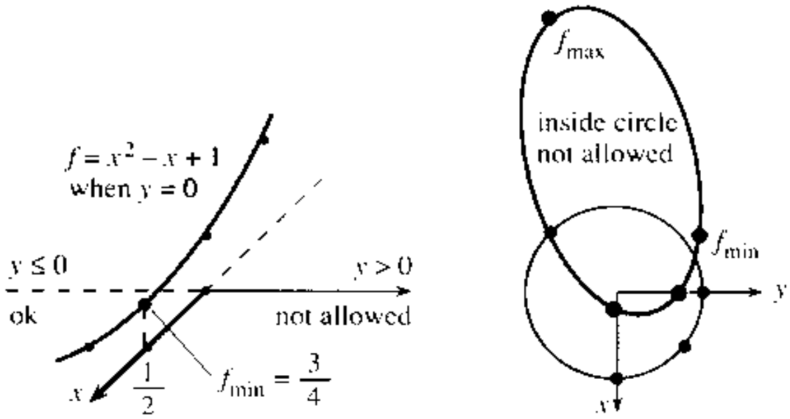


Figure 24: The boundaries $y = 0$ and $x^2 + y^2 = 1$ contain the minimum points.

Exemple 50. Minimize the same .f(x,y) on or outside the circle x2 y2 = 1.

Answer 50 One possibility is $f_x = 0$ and $f_y = 0$. But this is at $(1/3/1/3)$, 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 $f = x^2 + xy + xy + 1$ is a function of the angle t :

$$\begin{aligned} f(t) &= 1 + \cos t \sin t - \cos t - \sin t + 1 \\ \frac{df}{dt} &= \cos^2 t - \sin^2 t + \sin t - \cos t \\ &= (\cos t - \sin t)(\cos t + \sin t - 1) \end{aligned}$$

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. **Exercise 26** sorts them out, and we can find 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.

8.3 Maximum vs. minum 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.

Exemple 51. $f(x, y) = x^2 + 10xy + y^2$ has $f_{xx} = 2$, $f_{xy} = 10$, $f_{yy} = 2$ (minimum or not?)

Answer 51 All second derivatives are positive, but wait and see. The stationary point is $(0, 0)$, where $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are both zero. Our function is the sum of $x^2 + y^2$, which goes upward, and $10\,xy$ which has a saddle. The second derivatives must decide whether $x^2 + y^2$ or $10\,xy$ 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 $10\,xy$.

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**.

Exemple 52. $f(x, y) = x^2 + xy + y^2$ has $f_{xx} = 2, f_{xy} = 1, f_{yy} = 2$ (minimum or not ?)

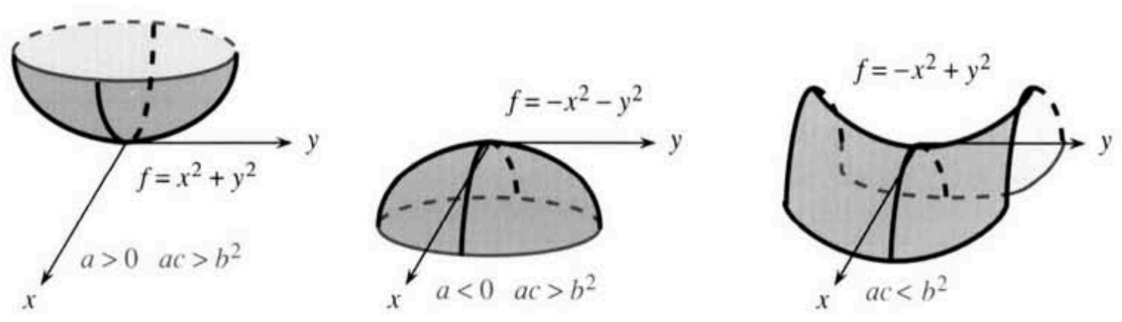


Figure 25: Minimum, maximum, saddle point based on the signs of a and $ac - b^2$.

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 51** : 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 48** moved the stationary point to $(1/3, 1/3)$. The second derivatives are still $2, 1, 2$, and they pass the test for a minimum :

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

- **minimum** if $a > 0$ and $ac > b^2$
- **maximum** if $a < 0$ and $ac > b^2$
- **saddle point** if $ac < b^2$

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

$$ax^2 + 2bxy + cy^2 = a\left(x + \frac{b}{a}\,y\right)^2 + \frac{ac-b^2}{a}\,y^2 \tag{8.7}$$

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 51** had $b = 5$ and a saddle point. **Example 52** had $b = 1$ 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.

Exemple 53. $f(x,y) = e^x - x - \cos y$ 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 $e^x - x - \cos y$:

$$\left(1 + x + \frac{1}{2}x^2 + \dots\right) - x - \left(1 + y + \frac{1}{2}y^2 + \dots\right) = \frac{1}{2}x^2 + \frac{1}{2}y^2 + \text{higher order terms} \tag{8.8}$$

There is a *minimum* at the origin. The quadratic part $\frac{1}{2}x^2 + \frac{1}{2}y^2$ 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 $\frac{1}{2}x^2 + \frac{1}{2}y^2$.

The previous test 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».

Exemple 54. $f(x,y) = e^{xy}$ 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:

$$e^{xy} = 1 + xy + \frac{1}{2}\,x^2y^2 + \dots \tag{8.9}$$

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 $f_{fxx} f_{yy} > (f_{xy})^2$? 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**.

8.4 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). We already know the n-th derivative of $f(x)$ denoted $d^n f/dx^n$, that appear in a single variable expansion, is mutliplied 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 $f(x,y)$:

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

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

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**.

Exemple 55. All derivatives of e^{x+y} equal one at the origin. The Taylor series is

$$e^{x+y} = 1 + x + y + \frac{x^2}{2} + xy + \frac{y^2}{2} + \cdots = \sum\sum \frac{x^n y^m}{n!m!} \tag{8.10}$$

This happens to have $ac = b^2$, the special case that was omitted so far. *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. 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

$$\begin{cases} (g_x)\Delta x + (g_y)\Delta y = -g(x_n, y_n) \\ (h_x)\Delta x + (h_y)\Delta y = -h(x_n, y_n) \end{cases} \quad \text{or} \quad \begin{cases} (f_{xx})\Delta x + (f_{xy})\Delta y = -f_x(x_n, y_n) \\ (f_{xy})\Delta x + (f_{yy})\Delta y = -f_y(x_n, y_n) \end{cases} \tag{8.11}$$

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 (see **Exercise 29**). At the end of each step, compute first derivatives and start again in the new steepest direction.

A Norms

The relative position of a point (say A) to another point (say B) can be represented by a vector $\boldsymbol{v}$. Then, we can associate to this nonzero vector $\boldsymbol{v}$ a positive number $||\boldsymbol{v}||$ called its **norm**. That number measures the « length » of the vector, meaning the distance between A and B. There exist many useful measures of length (many different norms). Every norm for vectors (or functions or matrices...) must share these two properties of the absolute value $|c|$ of a number :

- Multiply $\boldsymbol{v}$ by c (**Resealing**) : $||c \cdot \boldsymbol{v}\| = |c| \cdot \|\boldsymbol{v}\|$
- Add $\boldsymbol{v}$ to $\boldsymbol{w}$ (**Triangle inequality**) : $\|\boldsymbol{v}+\boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$

We start with three special norms, by far the most important. They are the $\boldsymbol{\ell^2}$ norm and $\boldsymbol{\ell^1}$ norm and $\boldsymbol{\ell^\infty}$ norm of the vector $\boldsymbol{v} = (v_1, \ldots, v_n)$. The vector $\boldsymbol{v}$ is in $\mathbb{R}^n$ (real v_i) or in $\mathbb{C}^n$ (complex v_i) :

- $\boldsymbol{\ell^2}$ **norm = Euclidean norm** : $\|\boldsymbol{v}\|_2 = \sqrt{|v_1|^2 + \cdots + |v_n|^2}$
- $\boldsymbol{\ell^1}$ **norm = 1-norm** : $\|\boldsymbol{v}\|_2 = |v_1| + \cdots + |v_n|$
- $\boldsymbol{\ell^\infty}$ **norm = max norm** : $\|\boldsymbol{v}\|_\infty =$ **maximum of** $|v_1|, \ldots, |v_n|$

The vector $\boldsymbol{v} = (1, 1, \ldots, 1)$ has norms $\|\boldsymbol{v}\|_2 = \sqrt{n}$ and $\|\boldsymbol{v}\|_1 = n$ and $\|\boldsymbol{v}\|_\infty = 1$. These three norms are the particular cases $p = 2$ and $p = 1$ and $p = \infty$ of the ℓ^p norm $\|\boldsymbol{v}\|_p = (|v_1|^p + \cdots + |v_n|^p)^{1/p}$. **Figure 26** shows vectors with **norm 1**: $p = 1/2$ is illegal.

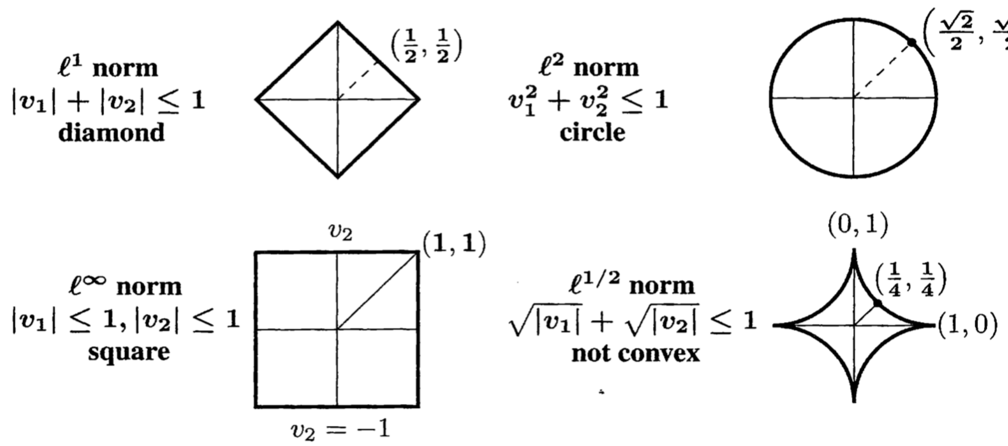


Figure 26: The important vector norms $\|\boldsymbol{v}\|_1$, $\|\boldsymbol{v}\|_2$, $\|\boldsymbol{v}\|_\infty$, and a failure ($p = 0$ fails too).

The failure for $p = 1/2$ is in the triangle inequality : $(1, 0)$ and $(0, 1)$ have norm 1, but their sum $(1, 1)$ has norm $2^{1/p} = 4$. Only $1 < p < \infty$ produce an acceptable norm $\|\boldsymbol{v}\|_\infty$.

The Minimum of $\|\boldsymbol{v}\|_p$ on the line $a_1 v_1 + a_2 v_2 = 1$

Which point on a diagonal line like $3v_1 + 4v_2 = 1$ is closest to $(0, 0)$? The answer (and the meaning of « closest ») will depend on the norm. This is another way to see important differences between $\boldsymbol{\ell^1}$ and $\boldsymbol{\ell^2}$ and $\boldsymbol{\ell^\infty}$. We will see a first example of a very special feature : **Minimization in $\boldsymbol{\ell^1}$ produces sparse solutions**.

To see the closest point to $(0,0)$, expand the $\boldsymbol{\ell^1}$ diamond and $\boldsymbol{\ell^2}$ circle and $\boldsymbol{\ell^2}$ square until they touch the diagonal line. For each p, that touching point $\boldsymbol{v}^*$ will solve our optimization problem :

$$\boxed{\text{Minimize } \|\boldsymbol{v}\|_p, \text{ among vectors } (v_1, v_2) \text{ on the line } 3v_1 + 4v_2 = 1}$$

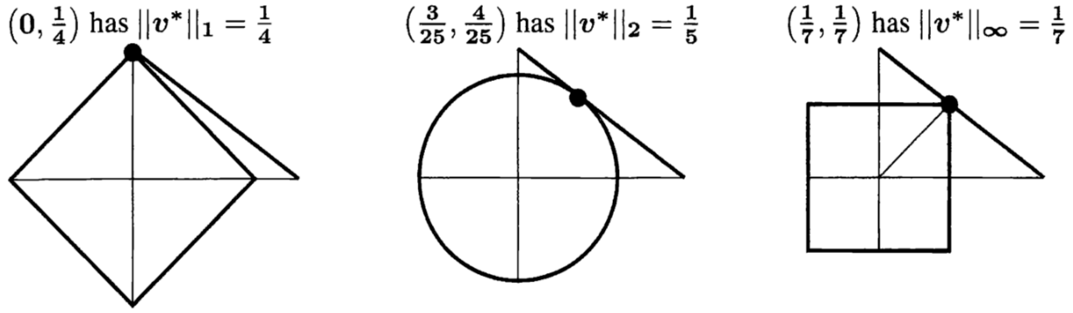


Figure 27: The solutions $\boldsymbol{v}^*$ to the $\boldsymbol{\ell^1}$ and $\boldsymbol{\ell^2}$ and $\boldsymbol{\ell^\infty}$ minimizations. The first is **sparse**.

The first figure displays a highly important property of the minimizing solution to the $\boldsymbol{\ell^1}$ problem : **That solution $\boldsymbol{v}^*$ has zero components**. The vector $\boldsymbol{v}^*$ is «sparse». This is because **a diamond touches a line at a sharp point**. The line (or hyperplane in high dimensions) contains the vectors that solve the m constraints $\boldsymbol{Av} = \boldsymbol{b}$. The surface of the diamond contains vectors with the same $\boldsymbol{\ell^1}$ norm. The diamond expands to meet the line at a corner of the diamond !

The essential point is that **the solutions to those problems are sparse**. They have few nonzero components, and those components have meaning. By contrast the least squares solution (using $\boldsymbol{\ell^2}$) has many small and non-interesting components. By squaring, those components become very small and hardly affect the $\boldsymbol{\ell^2}$ distance.

One final observation : **The «$\boldsymbol{\ell^0}$ norm»** of a vector $\boldsymbol{v}$ counts the number of nonzero components. But this is not a true norm. The points with $\|\boldsymbol{v}\|_0 = 1$ lie on the x axis or y axis, one nonzero component only. The figure for $p = 1/2$ on the previous page becomes even more extreme, just a cross or a skeleton along the two axes.

Of course this skeleton is not at all convex. The «zero norm» violates the fundamental requirement that $||2\boldsymbol{v}|| = 2||\boldsymbol{v}||$. In fact $||2\boldsymbol{v}||_0 = \|\boldsymbol{v}\|_0 =$ number of non zeros in $\boldsymbol{v}$.

The wonderful observation is that we can find the sparsest solution to $\boldsymbol{Av} = \boldsymbol{b}$ *by using the $\boldsymbol{\ell^1}$ norm*. We have «convexified» that $\boldsymbol{\ell^0}$ skeleton along the two axes. We filled in the skeleton, and the result is the $\boldsymbol{\ell^1}$ diamond.

Inner Products and Angles

The $\boldsymbol{\ell^2}$ norm has a special place. When we write $||\boldsymbol{v}||$ with no subscript, this is the norm we mean. It connects to the ordinary geometry of inner products[2] $(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{v}^\mathsf{T}\boldsymbol{w}$ and angles θ between vectors:

(a) Inner product = length squared : $\boldsymbol{v} \cdot \boldsymbol{v} = \boldsymbol{v}^\mathsf{T}\boldsymbol{v} = \|\boldsymbol{v}\|^2$

(b) Angle θ between vectors $\boldsymbol{v}$ and $\boldsymbol{w}$: $\boldsymbol{v}^\mathsf{T}\boldsymbol{v} = \|\boldsymbol{v}\|\,\|\boldsymbol{w}\| \cos\theta$

[2]Here, the vector $\boldsymbol{v}^\mathsf{T}$ is the transposed vector of $\boldsymbol{v}$.

Then $\boldsymbol{v}$ is orthogonal to $\boldsymbol{w}$ when $\theta = 90°$ and $\cos\theta = 0$ and $\boldsymbol{v}^\mathsf{T}\boldsymbol{w} = 0$.

Those two connections lead to the most important inequalities in mathematics :

$\boxed{\textbf{Cauchy-Schwarz : } |\boldsymbol{v}^\mathsf{T}\boldsymbol{w}| \le \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|}$ $\quad$ $\boxed{\textbf{Triangle Inequality : } \|\boldsymbol{v}+\boldsymbol{w}\| \le \|\boldsymbol{v}\| + \|\boldsymbol{w}\|}$

We can connect the Cauchy-Schwarz equality to the equation for the inner product with the cosine: $|\cos\theta| \le 1$ means that $|\boldsymbol{v}^\mathsf{T}\boldsymbol{w}| \le \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$. And this in turn leads to the triangle inequality, connecting the sides $\boldsymbol{v}$, $\boldsymbol{w}$, and $\boldsymbol{v}+\boldsymbol{w}$ of an ordinary triangle in n dimensions :

(1) Equality : $\|\boldsymbol{v}+\boldsymbol{w}\|^2 = (\boldsymbol{v}+\boldsymbol{w}, \boldsymbol{v}+\boldsymbol{w}) = \boldsymbol{v}^\mathsf{T}\boldsymbol{v} + \boldsymbol{v}^\mathsf{T}\boldsymbol{w} + \boldsymbol{w}^\mathsf{T}\boldsymbol{v} + \boldsymbol{w}^\mathsf{T}\boldsymbol{w}$

(2) Inequality : $\|\boldsymbol{v}+\boldsymbol{w}\|^2 \le \|\boldsymbol{v}\|^2 + 2\,\|\boldsymbol{v}\|\,\|\boldsymbol{w}\| + \|\boldsymbol{w}\|^2 = (\|\boldsymbol{v}\| + \|\boldsymbol{w}\|)^2$

This confirms our intuition : Any side length in a triangle is less than the sum of the other two side lengths : $\|\boldsymbol{v}+\boldsymbol{w}\| \le \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$. Equality in the $\boldsymbol{\ell^2}$ norm is only possible when the triangle is totally flat and all angles have $|\cos\theta| = 1$.

Inner Products and S-Norms

A final question about vector norms. Is $\boldsymbol{\ell^2}$ the only norm connected to inner products (dot products) and to angles ? There are no dot products for $\boldsymbol{\ell^1}$ and $\boldsymbol{\ell^\infty}$. But we can find other inner products that match other norms. For instance of we consider any symmetric positive definite matrix $\boldsymbol{S}$ then

- $\|\boldsymbol{v}\|_{\boldsymbol{S}}^2$ gives a norm for $\boldsymbol{v}$ in $\mathbb{R}$ (called the $\boldsymbol{S}$-norm)
- $(\boldsymbol{v}, \boldsymbol{w})_{\boldsymbol{S}} = \boldsymbol{v}^\mathsf{T}\boldsymbol{S}\boldsymbol{w}$ gives the $\boldsymbol{S}$-inner product for $\boldsymbol{v}$, $\boldsymbol{w}$ in $\mathbb{R}$

The inner product $(\boldsymbol{v}, \boldsymbol{w})_{\boldsymbol{S}}$ agrees with $\|\boldsymbol{v}\|_{\boldsymbol{S}}^2$. We have angles from **(b)**. We have inequalities from **(2)**. The proof is in **(2)** when every norm includes the matrix $\boldsymbol{S}$. We know that every positive definite matrix $\boldsymbol{S}$ can be factored into $\boldsymbol{A}^\mathsf{T}\boldsymbol{A}$. Then the $\boldsymbol{S}$-norm and $\boldsymbol{S}$-inner product for $\boldsymbol{v}$ and $\boldsymbol{w}$ are exactly the standard $\boldsymbol{\ell^2}$ norm and the standard inner product for $\boldsymbol{A}\boldsymbol{v}$ and $\boldsymbol{A}\boldsymbol{w}$. Therefore we can write

$$(\boldsymbol{v}, \boldsymbol{w})_{\boldsymbol{S}} = \boldsymbol{v}^\mathsf{T}\boldsymbol{S}\boldsymbol{w} = (\boldsymbol{S}\boldsymbol{v})^\mathsf{T}(\boldsymbol{A}\boldsymbol{w}) \quad \text{because} \quad \boldsymbol{S} = \boldsymbol{A}^\mathsf{T}\boldsymbol{A} \tag{A.1}$$

This is not an impressive idea but it is convenient. The matrices $\boldsymbol{S}$ and $\boldsymbol{A}$ are « weighting » the vectors and their lengths. Then weighted least squares is just ordinary least squares in this weighted norm.

Einstein needed a new definition of length and distance in 4-dimensional space-time. Lorentz proposed this one, which Einstein accepted (c = speed of light) :

$$\boldsymbol{v} = (x, y, z, t) \qquad \text{with} \qquad \|\boldsymbol{v}\| = x^2 + y^2 + z^2 - c^2t^2$$

Is this a true norm in $\mathbb{R}^4$?

Norms and Inner Products of Functions

A function $f(x)$ is a « **vector in function space** » . That simple idea gives linear algebra an importance that goes beyond n-dimensional space $\mathbb{R}^n$. All the intuition associated with linearity carries us from finite dimensions onward to infinite dimensions. The fundamental requirement for a vector space is to allow *linear combinations* $c\,f + d\,g$ of vectors $\boldsymbol{v}$ and $\boldsymbol{w}$. This idea extends directly to linear combinations $c\,f + d\,g$ of functions f and g.

It is exactly with norms that new questions arise in infinite dimensions. Think about the particular vectors $\boldsymbol{v}_n = (1, 1/2, \dots (1/2)^n, 0, 0, \dots)$ in the usual $\boldsymbol{\ell^2}$ norm. Those vectors come closer together since $\|\boldsymbol{v}_n - \boldsymbol{v}_M\|$ as $n \to \infty$ and $N \to \infty$. For a vector space to be « complete » , every converging sequence $\boldsymbol{v}_n$ must have a limit $\boldsymbol{v}_\infty$ in the space : $\|\boldsymbol{v}_n - \boldsymbol{v}_\infty\| \to 0$.

(a) The space of infinite vectors $\boldsymbol{v} = (v_1, \dots, v_N, 0, 0, \dots)$ ending in all zeros is ***not complete***.

(b) The space of vectors with $\|\boldsymbol{v}\|^2 = |v_1|^2 + |v_2|^2 + \cdots < +\infty$ is ***complete***. A vector like $\boldsymbol{v}_\infty = (1, 1/2, 1/4, 1/8, \dots)$ is included in this space but not in **1**. It doesn't end in zeros.

Two famous names are associated with complete infinite-dimensional vector spaces :

- A ***Banach space*** is a complete vector space with a norm $\|\boldsymbol{v}\|$ satisfying rules (a) and (b).
- A ***Hilbert space*** is a Banach space that also has an inner product with $(\boldsymbol{v}, \boldsymbol{v})$ equal to $\|\boldsymbol{v}\|^2$.

Those spaces are infinite-dimensional when the vectors have infinitely many components :

- $\boldsymbol{\ell^1}$ is a Banach space with norm $\|\boldsymbol{v}\| = |v_1| + |v_2| + \dots$
- $\boldsymbol{\ell^2}$ is a Hilbert space because it has an inner product $(\boldsymbol{v}, \boldsymbol{w}) = v_1 w_1 + v_2 w_2 + \dots$
- $\boldsymbol{\ell^\infty}$ is a Banach space with norm $\|\boldsymbol{v}\|_\infty =$ supremum of the numbers $|v_1|, |v_2|, \dots$

Our special interest is in function spaces. The vectors can be functions $f(x)$ for $0 < x < 1$.

- $\boldsymbol{L^1}[0, 1]$ is a Banach space with $\|\boldsymbol{f}\| = \int_0^1 |f(x)|\, dx$
- $\boldsymbol{L^2}[0, 1]$ is a Hilbert space with (f, g) = fo f (x) g (x) dx and 11 If 113 = fo I f (x)I2dx
- $\boldsymbol{L^\infty}[0, 1]$ is a Banach space with $\|\boldsymbol{f}\|_\infty =$ supremum of $|f(x)|$.

Notice the parallel between sums of components in $\boldsymbol{\ell^1}$ and integrals of functions in $\boldsymbol{L^1}$. Similarly for sums of squares in $\boldsymbol{\ell^2}$ and integrals of $|f(x)|^2$ in $\boldsymbol{L^2}$. *Add or integrate.*

Smoothness of Functions

There are many types of function spaces. This part of mathematics is « **functional analysis** » . Often a function space brings together all functions with a specific level of smoothness. An outstanding example is the space $\boldsymbol{C}[0, 1]$ containing all **continuous functions** :

$$f \text{ belongs to } \boldsymbol{C}[0,1] \text{ and } \|\boldsymbol{f}\|_{\boldsymbol{C}} = \max |f(x)| \text{ if } f(x) \text{ is continuous for all } 0 < x < 1.$$

The max norm in the function space $\boldsymbol{C}$ is like the $\boldsymbol{\ell^\infty}$ norm for vectors. We can increase the level of smoothness to $\boldsymbol{C^1}[0, 1]$ or $\boldsymbol{C^2}[0, 1]$. Then the first derivative or second derivative must also be continuous. These are Banach spaces but not Hilbert spaces. Their norms do not come from inner products, compare the next two formula :

$$\|\boldsymbol{f}\|_{\boldsymbol{C^1}} = \|\boldsymbol{f}\|_{\boldsymbol{C}} + \left\| \frac{d\boldsymbol{f}}{dx} \right\|_{\boldsymbol{C}} \qquad \|\boldsymbol{f}\|_{\boldsymbol{C^2}} = \|\boldsymbol{f}\|_{\boldsymbol{C}} + \left\| \frac{d^2\boldsymbol{f}}{dx^2} \right\|_{\boldsymbol{C}} \tag{A.2}$$

If we want a Hilbert space $\boldsymbol{H^1}$ then we build on the usual $\boldsymbol{L^2}$ space (which is $\boldsymbol{H^0}$) :

$$\|\boldsymbol{f}\|_{\boldsymbol{H^1}} = \|\boldsymbol{f}\|^2 + \left\| \frac{d^2\boldsymbol{f}}{dx^2} \right\|^2 \quad \text{and} \quad (\boldsymbol{f}, \boldsymbol{g})_{\boldsymbol{H^1}} = \int_0^1 f(x)g(x)\, dx + \int_0^1 \frac{df}{dx}\frac{dg}{dx}\, dx \tag{A.3}$$

We bring this wild excursion in function space to an end with three examples.

(a) The infinite vector $\boldsymbol{v} = (1, 1/2, 1/3, 1/4, \dots)$ is in $\boldsymbol{\ell^2}$ and $\boldsymbol{\ell^\infty}$. But it is not in $\boldsymbol{\ell^1}$. The sum of its components is infinite.

(b) A step function is in $\boldsymbol{L^1}$ and $\boldsymbol{L^2}$ and $\boldsymbol{L^\infty}$, but not in $\boldsymbol{C}$. The function has a jump.

(c) The ramp function $\max(0, x)$ is in $\boldsymbol{C}$ and $\boldsymbol{H^1}$ but not in $\boldsymbol{C^1}$. The slope has a jump.

B Exercises

B.1 Surfaces and Level Curves

Exercice 1.

(1) Draw the surface $z = f(x, y)$ for these four functions :

$$f_1 = \sqrt{4 - x^2 - y^2} \qquad f_2 = 2 - \sqrt{x^2 + y^2} \qquad f_3 = 2 - \frac{1}{2}(x^2 + y^2) \qquad f_4 = 1 + e^{-x^2-y^2}$$

(2) The level curves of all four functions are ______. They enclose the maximum at ______. Draw the four curves $f(x, y) = 1$ and rank them by increasing radius.

(3) Set $y = 0$ and compute the x derivative of each function at $x = 2$. Which mountain is flattest and which is steepest at that point?

(4) Set $y = 1$ and compute the x derivative of each function at $x = 1$.

Exercice 2.

(1) The graph of $w = F(x, y, z)$ is ______-dimensional surface in $xyzw$ space. Its level sets $F(x, y, z) = c$ are ______ dimensional surfaces in xyz space. For $w = x - 2y + z$ those level sets are ______. For $w = x^2 + y^2 + z^2$ those level sets are ______.

(2) The level sets of $F = x^2 + y^2 + qz^2$ look like footballs when q is ______, like basketballs when q is ______, and like frisbees when q is ______.

(3) **(a)** The level curves of $f(x, y) = \sin(x - y)$ are ______.

(b) The level curves of $g(x, y) = \sin(x^2 - y^2)$ are ______.

(c) The level curves of $h(x, y) = \sin(x - y^2)$ are ______.

B.2 Limit and Continuity

Exercice 3. For the following exercises, evaluate the limits at the indicated values of x and y. If the limit does not exist, state this and explain why the limit does not exist.

(1) $\lim_{(x,y)\to(0,0)} \frac{4x^2+10y^2+4}{4x^2-10y^2+6}$

(2) $\lim_{(x,y)\to(0,1)} \frac{y^2 \sin x}{x}$

(3) $\lim_{(x,y)\to(\pi/4,1)} \frac{y \tan x}{y+1}$

(4) $\lim_{(x,y)\to(2,5)} \left(\frac{1}{x} - \frac{5}{y}\right)$

(5) $\lim_{(x,y)\to(4,4)} e^{-x^2-y^2}$

(6) $\lim_{(x,y)\to(1,2)} \left(x^2y^3 - x^3y^2 + 3x + 2y\right)$

(7) $\lim_{(x,y)\to(0,0)} \frac{xy+1}{x^2+y^2+1}$

(8) $\lim_{(x,y)\to(0,0)} \ln(x^2 + y^2)$

Exercice 4. For the following cases, use algebraic techniques to evaluate the limit.

(1) $\lim_{(x,y)\to(0,0)} \frac{x^4-4y^4}{x^2+2y^2}$ **(2)** $\lim_{(x,y)\to(0,0)} \frac{x^2-xy}{\sqrt{x}-\sqrt{y}}$ **(3)** $\lim_{(x,y,z)\to(0,0,0)} \frac{x^2-y^2-z^2}{x^2+y^2-z^2}$

Exercice 5. For the following cases, evaluate the limit of the function by determining the value the function approaches along the indicated paths. If the limit does not exist, explain why not.

(1) $\lim_{(x,y)\to(0,0)} \frac{xy+y^3}{x^2+y^2}$

(a) Along the x-axis ($y = 0$) **(b)** Along the y-axis ($x = 0$) **(c)** Along the path $y = 2x$

(2) Evaluate $\lim_{(x,y)\to(0,0)} \frac{xy+y^3}{x^2+y^2}$ using the results of the previous question.

(3) $\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^4+y^2}$

(a) Along the x-axis ($y = 0$) **(b)** Along the y-axis ($x = 0$) **(c)** Along the path $y = x^2$

(4) Evaluate $\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^4+y^2}$ using the results of the previous question.

Exercice 6. Discuss the continuity of the following functions. Find the largest region in the xy-plane in which the following functions are continuous.

(1) $f(x, y) = \ln(x + y)$

(2) $f(x, y) = \frac{1}{xy}$

Exercice 7.

(1) Determine the region of the xy-plane in which the function $g(x, y) = \arctan(\frac{xy^2}{x+y})$ is continuous.

(2) At what point in space is $g(x, y, z) = x^2 + y^2 - 2z^2$ continuous ?

B.3 Partial Derivatives

Exercice 8. Find $\partial f/\partial x$ and $\partial f/\partial y$ for the following functions.

(1) $3x - y + x^2y^2$

(2) $\sin(3x - y) + y$

(3) $x^3y^2 - x^2 - e^y$

(4) xe^{x+4}

(5) $(x + y)/(x - y)$

(6) $1/\sqrt{x^2 + y^2}$

(7) $(x^2 + y^2)^{-1}$

(8) $\ln(x + 2y)$

(9) $\ln\sqrt{x^2 + y^2}$

(10) y^x

(11) $\tan^{-1}(y/x)$

(12) $\ln(xy)$

Exercice 9. Compute $f_{xx}, f_{xy} = f_{yx}$ and f_{yy} for the following functions.

(1) $x^2 + 3xy + 2y^2$

(2) $(x + 3y)$

(3) $(x + iy)^3$

(4) e^{ax+by}

(5) $1/\sqrt{x^2 + y^2}$

(6) $(x + y)^n$

(7) $\cos ax \sin by$

(8) $1/(x + iy)$

Exercice 10.

(1) Show that $t^{-1/2}e^{-x^2/4t}$ satisfies the **heat equation** $f_t = f_{xx}$. This $f(x, t)$ is the temperature at position x and time t due to a point source of heat at $x = 0$, $t = 0$.

(2) The equation for heat flow in the xy plane is $f_t = f_{xx} + f_{yy}$. Show that $f(x, y, t) = e^{-2t} \sin x \sin y$ is a solution. What exponent α in $f = e^{\alpha} \sin 2x \sin 3y$ gives a solution?

(3) Find solutions $f(x, y) = e^{\alpha} \sin mx \cos ny$ of the heat equation $f_t = f_{xx} + f_{yy}$. Show that $t^{-1}e^{-x^2/4t}e^{-y^2/4t}$ is also a solution.

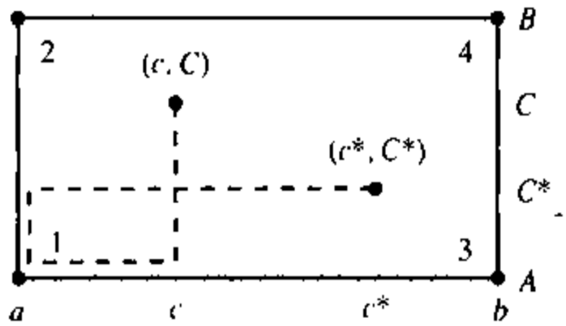
(4) The basic **wave equation** is $f_{tt} = f_{xx}$. Verify that $f(x,t) = \sin(x+t)$ and $f(x,t) = \sin(x-t)$ are solutions. Draw both graphs at $t = \pi/4$. Which wave moved to the left and which moved to the right ?

Exercice 11. The proof of $f_{xy} = f_{yx}$ studies $f(x,y)$ in a small rectangle. The top-bottom difference is $g(x) = f(x,B) - f(x,A)$. The difference at the corners $1,2,3,4$ is :

$$\begin{aligned} Q &= [f_4 - f_3] - [f_2 - f_1] \\ &= g(b) - g(a) \quad \text{(definition of } g) \\ &= (b-a)g_x(c) \quad \text{(Mean Value Theorem)} \\ &= (b-a)[f_x(c,B) - f_x(c,A)] \ \text{(compute } g_x) \\ &= (b-a)(B-A)f_{xy}(c,C) \quad \text{(MVT again)} \end{aligned}$$

(1) The right-left difference is $h(y) = f(b,y) - f(a,y)$. The same Q is $h(B) - h(A)$. Change the steps to reach $Q = (B-A)(b-a)f_y(c^*,C^*)$.

(2) The two forms of Q make f_{xy} at (c,C) equal to f_{yx} at (c^*,C^*). Shrink the rectangle toward (a,A). What assumption yields $f_{xy} = f_{yx}$ at that typical point ?



Exercice 12.

(1) Definition of continuity : $f(x,y)$ is continuous at (a,b) if $f(a,b)$ is defined and $f(x,y)$ approaches the limit ______ as (x,y) approaches (a,b). Construct a function that is not continuous at $(1,2)$.

(2) Show that $x^2y/(x^4+y^2) \to 0$ along every straight line $y = mx$ to the origin. But traveling down the parabola $y = x^2$, the ratio equals ______.

(3) Can you define $f(0,0)$ so that $f(x,y)$ is continuous at $(0,0)$?

(a) $f = |x| + |y-1|$ **(b)** $f = (1+x)^y$ **(c)** $f = x^{1+y}$

B.4 Tangent Planes and Linear Approximations

Exercice 13. Find the tangent plane and the normal vector at P.

(1) $z = \sqrt{x^2+y^2}$, $P(0,1,1)$

(2) $x + y + z = 17$, $P(3,4,10)$

(3) $z = x/y$, $P(6,3,2)$

(4) $z = e^{x+2y}$, $P(0,0,1)$

(5) $x^2+y^2+z^2=6$, $P(1,2,1)$

(6) $x^2+y^2+2z^2=7$, $P(1,2,1)$

(7) $z=x^y$, $P(1,1,1)$

(8) $V=\pi r^2 h$, $P(2,2,8\pi)$

Exercice 14.

(1) The normal $\boldsymbol{N}$ to the surface $F(x,y,z)=0$ has components F_x, F_y, F_z. The *normal line* has $x=x_0+F_x t$, $y=y_0+F_y t$, $z=$ ______. For the surface $xyz-24=0$, find the tangent plane and normal line at $(4,2,3)$.

(2) For the sphere $x^2+y^2+z^2=9$ find the equation of the tangent plane through $(2,1,2)$. Also find the equation of the normal line and show that it goes through $(0,0,0)$.

(3) For $w=xy$ near (x_0,y_0), the linear approximation is $dw=$ ______. This looks like the ______ rule for derivatives. The difference between $\Delta w = xy - x_0 y_0$ and this approximation is ______.

Exercice 15.

(1) If the supplier reduces s, **Figure 15** shows that P decreases and Q ______.

  (a) Find $P_s=50/3$ and $P_t=1/3$ in the economics equation (5.21) by solving the equations above it for Q_s, and Q_t.

  (b) What is the linear approximation to Q around $s=0.4$, $t=10$, $P=30$, $Q=50$?

(2) Solve the equations $P=-0.2\,Q+40$ and $P=s\,Q+t$ for P and Q. Then find $\partial P/\partial s$ and $\partial P/\partial t$ explicitly. At the same s,t,P,Q check $50/3$ and $1/3$.

(3) If the « supply = demand » equation (5.20) changes to $P=s\,Q+t=Q+50$, find P_s and P_t at $s=1$, $t=10$.

B.5 Directional Derivatives and Gradients

Exercice 16. Compute $\boldsymbol{\nabla} f$, then $D_{\boldsymbol{u}} f = (\boldsymbol{\nabla} f)\cdot \boldsymbol{u}$, then $D_{\boldsymbol{u}} f$ at P.

(1) $f(x,y)=x^2-y^2$ with $\boldsymbol{u}=(\sqrt{3}/2,1/2)$ and $P(1,0)$

(2) $f(x,y)=3x+4y+7$ with $\boldsymbol{u}=(3/5,4/5)$ and $P(0,\pi/2)$

(3) $f(x,y)=e^x\cos y$ with $\boldsymbol{u}=(0,1)$ and $P(0,\pi/2)$

(4) $f(x,y)=y^{10}$ with $\boldsymbol{u}=(0,-1)$ and $P(1,-1)$

(5) $f(x,y)=$ distance to $(0,0)$ to $(0,3)$, with $\boldsymbol{u}=(1,0)$ and $P(1,1)$

Exercice 17. Find $\boldsymbol{\nabla} f=(f_x,f_y,f_z)$ for the following functions from physics.

(1) $1/\sqrt{x^2+y^2+z^2}$ (point source at the origin)

(2) $\ln(x^2+y^2)$ (line source along z axis)

(3) $1/\sqrt{(x-1)^2+y^2+z^2}-1/\sqrt{(x+1)^2+y^2+z^2}$ (dipole)

Exercice 18. Find the direction $\boldsymbol{u}$ in which f increases fastest at $P=(1,2)$, How fast ?

(1) $f(x,y)=ax+by$

(2) $f(x,y)=$ smaller of $2x$ and y

(3) $f(x,y)=e^{x-y}$

(4) $f(x,y)=\sqrt{5-x^2-y^2}$ (careful)

Exercice 19. 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 $|\mathbf{grad}\, D| = 1$. The graph of $D(x, y)$ is a ______ with its vertex at ______.

B.6 The Chain Rule

Exercice 20. Find df/dt from the chain rule in the following cases.

(1) $f = x^2 + y^2$ with $x = t$ and $y = t^2$

(2) $f = \sqrt{x^2 + y^2}$ with $x = t$ and $y = t^2$

(3) $f = xy$ with $x = 1 - \sqrt{t}$ and $y = 1 + \sqrt{t}$

(4) $f = x/y$ with $x = e^t$ and $y = 2e^t$

(5) $f = \ln(x + y)$ with $x = e^t$ and $y = e^t$

(6) $f = x^4$ with $x = t$ and $y = t$

Exercice 21.

(1) 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.

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

(3) Suppose $x_t = t$ and $y_t = 2t$, not constant as in (7.9) and (7.10). For $f(x, y)$ find f_t and f_{tt}. The answer involves $f_x, f_y, f_{xx}, f_{xy}, f_{yy}$.

(4) 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 $1/\cos\theta$?

Exercice 22.

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

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

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

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

Exercice 23.

(1) 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** 44 still multiply to give -1. First find $(\partial P/\partial V)_T$. from $\partial P/\partial V + (\partial F/\partial P)(\partial P/\partial V)_T = 0$.

(2) If we change the problem to four variables related by $F(x, y, z, t) = 0$, what is the corresponding product of four derivatives?

B.7 Maxima, minima and saddle points

Exercice 24. Find all stationary points ($f_x = f_y = 0$) in the following cases. Separate minimum from maximum from saddle point. The test with a, b, c applies with $a = f_{xx}, b = f_{xy}, c = f_{yy}$.

(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) $x^2y^2 - x$

(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) $8xy - x^4 - y^4$

Exercice 25.

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

(2) 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 $xy - 3x^2y - 2xy^2$.

(3) (Straight line fit) 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.

(4) (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.

Exercice 26.

(1) Minimize $f = x^2 + y^2 - x - y$ restricted by

(a) $x \leq 0$ **(b)** $y \geq 1$ **(c)** $x \leq 0$ and $y \geq 1$

(2) Minimize $f = x^2 + y^2 + 2x + 4y$ in the regions

(a) $x \leq 0$ **(b)** $y \geq 1$ **(c)** $x \leq 0$ and $y \geq 1$

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

(4) 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 $x = y$. Compute f at those points and locate the minimum.

Exercice 27. Find, in the following cases, $f_x, f_y, f_{xx}, f_{xy}, f_{yy}$ at the basepoint. Write the quadratic approximation to $f(x,y)$, meaning the Taylor series through second order terms.

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

(2) $f = e^{x+y}$ at $(1,1)$

(3) $f = \sin x \cos y$ at $(0,0)$

(4) $f = x^2 + y^2$ at $(1,-1)$

Exercice 28.

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

(2) 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)** $f = e^x e^y$

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

(4) **(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 ______.

Exercice 29.

(1) 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.

(2) 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 (8.11).

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

(4) 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 $x^2 y = f_{\max}$.