Math 1c — TA: Padraic Bartlett

Recitation 5: The Midterm Review

Week 5 — *Caltech 2012*

You have a midterm! Specifically, you have a midterm that's going to be on the following topics:

- Level curves.
- Limits in $\mathbb{R}^n$.
- Differentiation in $\mathbb{R}^n$: how to calculate the partial derivatives of functions, as well as the higher-order derivatives and mixed partial derivatives of a function.
- The directional derivative.
- The chain rule.
- How to find the tangent plane to a graph or surface at a point.
- How to find a second-order Taylor approximation to a function at a point.
- How to use the derivative to find the critical points of a function.
- How to use the Hessian to classify critical points into local maxima, minima, saddle points, and indeterminate points.
- How to optimize a function given a constraint, via the methods of Lagrange multipliers.

I have examples! Specifically, I have example problems on all of these concepts. This handout is made out of six problems that deal with the concepts listed above, and are all about as hard or somewhat harder than the problems you'll see on the midterm. In theory, understanding the material we've discussed throughout the course should be equivalent to understanding how to solve and talk about the problems in this handout; so, if you feel comfortable with these examples, the midterm should hopefully go fairly well for you.

Example. (Lagrange multipliers; level curves.) Consider the function

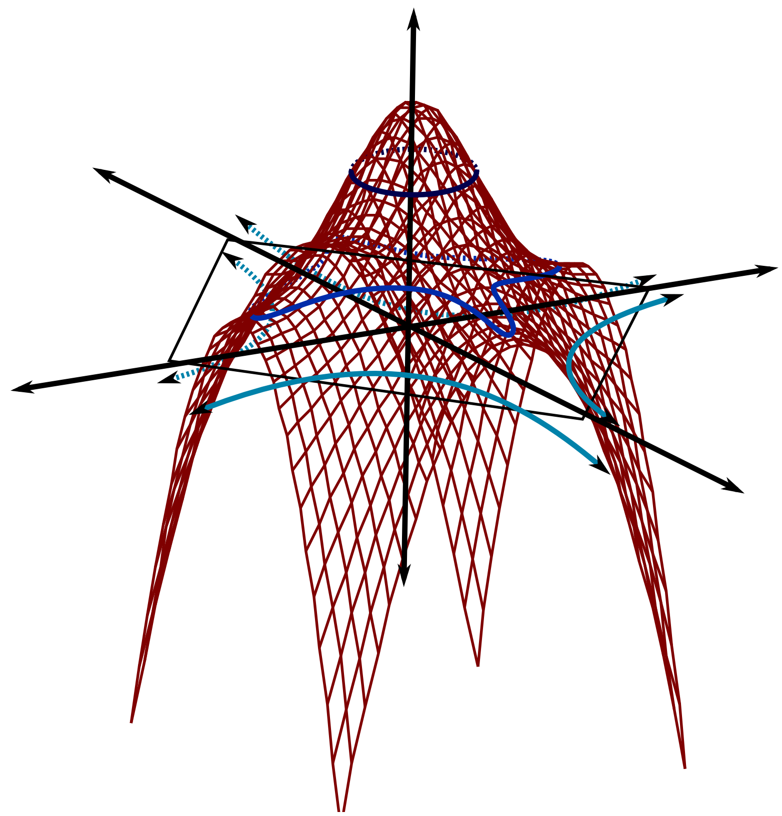
$$g(x,y) = e^{-x^2-y^2} - x^2y^2.$$

(a) Draw several level curves of this function.

(b) Let $f(x,y) = x+y$, and let S be the constraint set given by the level curve $\{(x,y) : g(x,y) = c\}$. For what values of c does $f\big|_S$ have a global maximum? For what values does it fail to have a global maximum: i.e. for what values of c is f unbounded on S?

(c) For $c = \frac{1}{4}$, find the global maximum of f on the above constraint set $S = \{(x,y) : g(x,y) = c\}$.

Solution. We graph the function in red, along with three level curves in different shades of blue, in the following picture.



Roughly speaking, there are three kinds of level curves for our function:

1. Level curves $g(x, y) = c$ where c is close to 1. There, because we need g to be close to 1, we need to have x and y very small (so that the $e^{-x^2-y^2}$ part is as close to 1 as we can get it, and the $-x^2y^2$ part is not too large.) In particular, this forces us to have a roughly circular shape, as for very small values of (x, y) the x^2y^2 part is insignificant and our function looks roughly like $e^{-x^2-y^2}$, which is roughly $1 - x^2 - y^2$ (via Taylor series) for small values of (x, y).

2. Level curves $g(x, y) = c$ where c is greater than 0, but not by much. For these values of c, we wind up having kind of a "four-armed" shape, with arms stretching out along the x- and y- axes. This is because when one of our coordinates is nearly zero, the other can become much larger (because our function is roughly $e^{-x^2-y^2}$ then), whereas when the coordinates are roughly the same, the dominant term is now the $-x^2y^2$ term, and we need to have both x and y be much smaller.

3. Level curves $g(x, y) = c$ where c is ≤ 0. In these cases, our level curves look like hyperbola-style curves, one in each quadrant. This is because on each axis, our function $g(x, y)$ can never be 0, as the $e^{-x^2-y^2}$-part is always positive and the $-x^2y^2$ part is zero on the axes.

This graphing and subsequent analysis suggests an answer to part (b), as well:

Claim 1 *Our function $f(x,y)$ has a global maximum on the curve $g(x,y)=c$ if and only if $1 \geq c > 0$.*

Proof. If $c > 1$, then there are no points (x,y) such that $g(x,y)=c$, because $e^{-x^2-y^2}$ is bounded above by $e^0 = 1$, while $-x^2y^2$ is bounded above by 0.

So: suppose that $1 \geq c > 0$. Then, if (x,y) are such that $g(x,y)=c$, we know that in particular

$$\begin{aligned} & e^{-x^2-y^2} \geq c \\ \Rightarrow \quad & -x^2 - y^2 \geq \ln(c) \\ \Rightarrow \quad & x^2 + y^2 \leq -\ln(c) \\ \Rightarrow \quad & \sqrt{x^2+y^2} \leq \sqrt{-\ln(c)} \\ \Rightarrow \quad & ||(x,y)|| \leq \sqrt{-\ln(c)}, \end{aligned}$$

i.e. the point (x,y) can be no further than $\sqrt{-\ln(c)}$ from the origin. (Because $1 \leq c > 0$, we know that $-\infty < \ln(c) \leq 0$, and therefore that this is a well-defined finite radius.)

Therefore, the set of points such that $g(x,y)=c$ is bounded. We also know that it is closed, because it is the level curve of a continuous function. Therefore, we know that any continuous function (in particular, f) will attain its global maxima and minima on this set, and do so at the critical points identified by the method of Lagrange multipliers.

Finally, suppose that $c \leq 0$. In this case, our claim is that f does not attain its global maximum on $g(x,y)=c$. To prove this, pick any value of n: we want to find a point (x,y) on our curve such that $f(x,y) > n$.

To do this, we simply use the intermediate value theorem. Pick any n, and choose x such that $-x^2 < c - 1$, and also $x > n$. Then, we know that

$$g(x,0) = e^{-x^2-0} - x^2 \cdot 0 = e^{-x^2} > 0 \geq c,$$

while

$$g(x,1) = e^{-x^2-1} - x^2 \cdot 1 = e^{-x^2} - x^2 < e^{-x^2} - c - 1 < c,$$

because $e^{-x^2} < 1$.

Therefore,because $g(x,0) > c$ and $g(x,1) < c$, by the intermediate value theorem, there is some value of y between 0 and 1 such that $g(x,y)=c$. At this point (x,y), we know that

$$f(x,y) = x + y \geq n + 0 \geq n,$$

which is what we wanted to prove: i.e. we've shown that we can find points on our curve along which $f(x,y)$ is arbitrarily large, and therefore that there is no global maximum.

Finally, with this theoretical discussion out of the way, we can turn to the calculational part of (c), which asks us to find the global maximum of our function f on the constraint set $g(x,y) = \frac{1}{4}$. First, note that by our above discussion, we know that a global maximum does exist, because when $1 \geq c > 0$ we've shown that our constraint set is closed and bounded. Furthermore, to find this maximum, it suffices to use the method of Lagrange multipliers to find all of the critical points of our function restricted to this curve, and simply select the largest value amongst these critical points. (Again, this is because $g(x, y) = c$ is closed and bounded, which means that our global maximum must occur a critical point.)

So: we calculate. We are looking for any points (x, y) such that either

- $\nabla(f)$ or $\nabla(g)$ are 0,
- $\nabla(f)$ or $\nabla(g)$ are undefined, or
- there is some constant λ such that $\nabla(f) = \lambda\nabla(g)$.

Because

$$\nabla(f)(x, y) = (1, 1)\,,$$

we can immediately see that $\nabla(f)$ is never undefined or zero.

Similarly, because

$$\nabla(g) = \left(-2xe^{-x^2-y^2} - 2xy^2, -2ye^{-x^2-y^2} - 2yx^2\right),$$

we can see that this is zero if and only if

$$\begin{aligned} &0 = -2xe^{-x^2-y^2} - 2xy^2 \\ \Leftrightarrow &0 = -2x\left(e^{-x^2-y^2} + y^2\right) \\ \Leftrightarrow &0 = x, \text{ because } e^{-x^2-y^2} + y^2 \text{ is strictly positive,} \end{aligned}$$

and

$$\begin{aligned} &0 = -2ye^{-x^2-y^2} - 2yx^2 \\ \Leftrightarrow &0 = -2y\left(e^{-x^2-y^2} + x^2\right) \\ \Leftrightarrow &0 = y, \text{ because } e^{-x^2-y^2} + x^2 \text{ is strictly positive.} \end{aligned}$$

So $\nabla(g)$ is always defined and is only zero at $(0, 0)$, which is not a point on our curve $g(x, y) = \frac{1}{4}$. Therefore, the only points we're concerned with are ones at which $\nabla(f) = \lambda\nabla(g)$; i.e. points such that

$$\begin{aligned} &\nabla(f) = (1, 1) = \lambda\nabla(g) = \lambda\left(-2xe^{-x^2-y^2} - 2xy^2, -2ye^{-x^2-y^2} - 2yx^2\right) \\ \Leftrightarrow &-2xe^{-x^2-y^2} - 2xy^2 = -2ye^{-x^2-y^2} - 2yx^2, \end{aligned}$$

because the above equation is equivalent to forcing both the left and right coordinates of $\nabla(g)$ to equal the same quantity (namely, $\frac{1}{\lambda}$.)

Solving, we can see that this is equivalent to

$$\begin{aligned} &0 = 2xe^{-x^2-y^2} + 2xy^2 - 2ye^{-x^2-y^2} - 2yx^2 \\ \Leftrightarrow &2(x - y)e^{-x^2-y^2} - 2xy(x - y) = 0. \end{aligned}$$

If $x - y = 0$, i.e. $x = y$, this equation holds. Otherwise, we can divide through by $2(x-y)$, and get

$$e^{-x^2-y^2} = xy.$$

Plugging this into our constraint equation $g(x,y) = \frac{1}{4}$ gives us

$$e^{-x^2-y^2} - (xy)^2 = \frac{1}{4} \Rightarrow (xy) - (xy)^2 = \frac{1}{4} \Rightarrow xy = \frac{1}{2},$$

by thinking of "xy" as one term and using the quadratic formula. But, if we think about what this means for the equation $e^{-x^2-y^2} = xy$, and specifically use $y = \frac{1}{2x}$, we have

$$\frac{1}{2} = xy = e^{-x^2-y^2} = e^{-x^2-\frac{1}{4x^2}}.$$

This is impossible! In specific, by taking a single-variable derivative, you can easily see that the largest value of $-x^2 - \frac{1}{4x^2}$ happens at $x = \frac{1}{\sqrt{2}}$, at which this is -1. This means that the largest that $e^{-x^2} - \frac{1}{4x^2}$ gets is $e^{-1} = \frac{1}{e}$, which is smaller than $\frac{1}{2}$.

Therefore, the only points at which $\nabla(f) = \lambda\nabla(g)$ are those at which $x = y$. Plugging this into our constraint $g(x,y) = \frac{1}{4}$ yields

$$e^{-2x^2} - x^4 = \frac{1}{4}$$
$$\Rightarrow x \equiv \pm.65.$$

The function $f(x,y) = x + y$ is equal to 1.3 at the point $(.65, .65)$ and is equal to -1.3 at $(-.65, -.65)$. Therefore, by our discussion earlier about how f must attain its global minima and maxima at the critical points discovered by the Lagrange multiplier process, we can safely conclude that $(.65, .65)$ is roughly the point at which $f(x,y)$ attains its global maxima, which is roughly 1.3.

Example. (Limits.) Either determine the following limits (with proof,) or show that they do not exist:

(a) $$\lim_{(x,y,z)\to(0,0,0)} \frac{||(x,y,z)||}{||(x,y,z)|| + \ln(|x+y+z|+1)}$$

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

Solution. First, recall from recitation 1 how we typically go about showing that limits either exist or do not exist.

- **Showing that a limit exists:** Often, the easiest approach to showing that a limit exists is to bound it or break it into smaller, single-variable limits, and use your Math 1a knowledge to deal with these limits. I.e. you can break things like $x + y + z$ into three individual single-variable limits, which are easy to calculate, and you can bound things like $\sin(xyz)x$ above by $|x|$.

  If this fails, **and you're quite convinced that the limit exists**, then you should try the $\epsilon - \delta$ definition of limits. Refer to the recitation 1 notes for a blueprint that will safely guide you though a $\epsilon - \delta$ proof.

- **Proving discontinuity/that a limit DNE**: One of the easiest ways to show that a limit $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ does not exist is to find two paths that **both go through a**, but such that $f(\mathbf{x})$ has different limits on each path as they approach **a**. Good paths are usually the x- or y- axes, or lines where$x = y$, or lines where $x = y^2$, though these are not the only lines to consider; often, the right paths to consider will become obvious only when you're actually looking at the function in question.

Now that we remember how to deal with limits, we tackle the two examples in this problem:

$$\text{(a)} \qquad \lim_{(x,y,z)\to(0,0,0)} \frac{||(x,y,z)||}{||(x,y,z)|| + \ln(|x+y+z|+1)}.$$

Looking at this function, it seems likely that the limit will not exist. Along certain paths, it seems like the natural log term in the denominator will grow at a similar rate to $||(x,y,z)||$, because natural log has a linear term in its Taylor series and therefore grows kind-of linearly near 1, which is what $||(x,y,z)||$ is also growing like. However, along other paths, we can probably make the natural log term be 0, and therefore not let it influence the limit.

This intuition turns out to be true! In particular, if we look at our function along the line $x = y = z$, we have that it's $\frac{\sqrt{3x^2}}{\sqrt{3x^2}+\ln(9|x|+1)}$, and therefore that

$$\begin{aligned}\lim_{x\to 0^+} \frac{|x|\sqrt{3}}{x\sqrt{3}+\ln(9|x|+1)} &\overset{\text{L'H}}{=} \lim_{x\to 0^+} \frac{\sqrt{3}}{\sqrt{3}+\frac{9}{9|x|+1}} \\ &= \frac{\sqrt{3}}{\sqrt{3}+9}.\end{aligned}$$

However, if we look at our function along the line $x = -y, z = 0$, we have that it's just $\frac{||(x,y,z)||}{||(x,y,z)||+\ln(1)} = \frac{||(x,y,z)||}{||(x,y,z)||+0} = 1$, and therefore that our limit is 1. These are different values; therefore, our function has no limit at 0.

Conversely, for

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

it seems like we should have a limit at 0: we know that for for small values, sin is roughly its input, and therefore because the numerator will roughly be a higher-order polynomial than the denominator, as we go to 0 our limit should be 0.

To make this rigorous, we use the squeeze theorem. Because $|\sin(a)| \leq |a|$ for any input a, we have in specific that

$$\left|\frac{\sin(x^4+y^4)}{x^2+y^2}\right| \leq \frac{x^4+y^4}{x^2+y^2}.$$

To make this easier to deal with, observe that we can bound the numerator above by $2 \cdot \max(x^4, y^4)$ and the denominator below by $\max(x^2, y^2)$, because all of the quantities involved are positive. Doing this gives us that

$$\frac{x^4 + y^4}{x^2 + y^2} \leq \frac{2 \cdot \max(x^4, y^4)}{\max(x^2, y^2)} = 2 \cdot \max(x^2, y^2).$$

Therefore, because $\lim_{(x,y)\to(0,0)} 2 \cdot \max(x^2, y^2) = 0$, the squeeze theorem tells us that

$$\text{(b)} \lim_{(x,y)\to(0,0)} \frac{\sin(x^4 + y^4)}{x^2 + y^2} = 0$$

as well.

Example. (Tangent planes.) Let S be the surface in $\mathbb{R}^3$ formed by the collection of all points (x, y, z) such that $e^{xyz} = e$. Find the tangent plane to S at $(1, 1, 1)$.

Solution. One way to attack this problem is to apply natural logs to both sides, which lets us write S as the collection of all points (x, y, z) such that $xyz = 1$; i.e. all points $x, y \neq 0$ such that $z = \frac{1}{xy}$. In other words, we can write S as the graph of the function $f(x, y) = \frac{1}{xy}$. We know that the gradient of $f(x, y)$ is just

$$\left(-\frac{y}{(xy)^2}, -\frac{x}{(xy)^2}\right),$$

which at 1 is just $(-1, -1)$. Therefore, using the formula from class for describing the tangent plane of surfaces of the form $f(x, y) = z$, we have that the tangent plane to our surface at $(1, 1, 1)$ is just

$$\begin{aligned}
(z - 1) &= \nabla(f)\Big|_{(1,1,1)} \cdot (x - 1, y - 1) = (-1, -1) \cdot (x - 1, y - 1) \\
\Rightarrow & z - 1 + x - 1 + y - 1 = 0.
\end{aligned}$$

Alternately, we also discussed a second formula in class for finding tangent planes to surfaces of the form $g(x, y, z) = C$, at some point (a, b, c). Specifically, we observed that the gradient of g at the point (a, b, c) was orthogonal to the tangent plane to our surface at this point: in other words, that we could define our tangent plane as just the set of all vectors orthogonal to the gradient of g through this point. As a formula, this was

$$\begin{aligned}
& 0 = \nabla(g)\Big|_{(1,1,1)} \cdot (x - 1, y - 1, z - 1) \\
\Leftrightarrow & 0 = \left(yze^{xyz}, xze^{xyz}, xye^{xyz}\right)\Big|_{(1,1,1)} \cdot (x - 1, y - 1, z - 1) \\
\Leftrightarrow & 0 = (1, 1, 1) \cdot (x - 1, y - 1, z - 1) \\
\Leftrightarrow & 0 = z - 1 + x - 1 + y - 1.
\end{aligned}$$

Reassuringly, we get the same answer no matter which method we pick.

Example. (Chain rule.) Let $g : \mathbb{R}^4 \to \mathbb{R}$ be defined by the equation $(w, x, y, z) = (wz - yx)$, and $h_\lambda : \mathbb{R}^2 \to \mathbb{R}^4$ be defined by the equation $h_\lambda(a, b) = (a, \lambda a, b, \lambda b)$.

(a) Calculate the derivative of $g \circ h_\lambda$ using the chain rule.

(b) Geometrically, explain why your answer in (a) is "obvious," in some sense.

Solution. So, we know that both g and h_λ are continuous functions on all of their domains; therefore, we know that their composition is continuous everywhere. Therefore, we know that the total derivative of $g \circ h_\lambda$ is just given by the partial derivatives of $g \circ h_\lambda$: i.e. $T(g \circ h_\lambda) = D(g \circ h_\lambda)$. Therefore, we can use the chain rule:

$$\begin{aligned}
D(g \circ h_\lambda)\Big|_{(a,b)} &= D(g)\Big|_{h_\lambda(a,b)} \cdot D(h_\lambda)\Big|_{(a,b)} \\
&= \begin{bmatrix} z & -y & -x & w \end{bmatrix}\Big|_{h_\lambda(a,b)} \cdot \begin{bmatrix} 1 & 0 \\ \lambda & 0 \\ 0 & 1 \\ 0 & \lambda \end{bmatrix} \\
&= \begin{bmatrix} \lambda b & -b & -\lambda a & a \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ \lambda & 0 \\ 0 & 1 \\ 0 & \lambda \end{bmatrix} \\
&= [\lambda b - \lambda b, \lambda a - \lambda a] \\
&= [0, 0].
\end{aligned}$$

Notice that this is geometrically somewhat obvious because g is just the determinant of the matrix $\begin{pmatrix} w & x \\ y & z \end{pmatrix}$, while the function h_λ just outputs the rank-1 matrix $\begin{pmatrix} a & b \\ \lambda a & \lambda b \end{pmatrix}$. Because the determinant of a rank 1 matrix is 0, we have that $g \circ h_\lambda$ is identically 0, and therefore also has derivative 0.

Example. (Taylor series; directional derivatives.) Let $g(x, y) = \sin(xy)$.

(a) Calculate the directional derivative of $g(x, y)$ at $(1, 2)$ in the direction $(3, 4)$.

(b) Calculate the second-order Taylor approximation of $g(x, y)$ at $(0, 0)$.

Solution. Because the gradient of g is just

$$\nabla(g) = (y\cos(xy), x\cos(xy)),$$

we know that the directional derivative at $(1, 2)$ in the direction $(3, 4)$ is just given to us by the dot product of $\nabla(g)(1, 2)$ with the **unit-length vector in the direction** $(3, 4)$, given by $\frac{1}{||(3,4)||} \cdot (3, 4) = \frac{1}{\sqrt{9+16}}(3, 4) = \left(\frac{3}{5}, \frac{4}{5}\right)$:

$$\nabla(g)(1, 2) \cdot \left(\frac{3}{5}, \frac{4}{5}\right) = (2\cos(1), \cos(2)) \cdot \left(\frac{3}{5}, \frac{4}{5}\right) = \frac{6\cos(1) + 4\cos(2)}{5}.$$

To calculate the Taylor approximation of g at $(0,0)$, we just need to construct the following function:

$$T_2(g)\big|_{(0,0)}(h_1,h_2) = g(0,0) + \nabla(g)\big|_{(0,0)} \cdot (x,y) + H(g)\big|_{(0,0)}(x,y).$$

To do this, simply note that the Hessian $H(g)$ of g is just

$$\begin{aligned}
H(g)\big|_{(0,0)}(h_1,h_2) &= \frac{1}{2}\left[h_1,h_2\right]\left[\begin{array}{cc} -y^2\sin(xy) & \cos(xy)-xy\sin(xy) \\ \cos(xy)-xy\sin(xy) & -x^2\sin(xy)\end{array}\right]\Bigg|_{(0,0)} \left[\begin{array}{c} h_1 \\ h_2 \end{array}\right] \\
&= \frac{1}{2}\left[h_1,h_2\right]\left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right] \cdot \left[\begin{array}{c} h_1 \\ h_2 \end{array}\right] \\
&= \frac{1}{2}\left[h_1,h_2\right]\left[\begin{array}{c} h_2 \\ h_1 \end{array}\right] \\
&= \frac{1}{2}(h_1h_2 + h_1h_2) \\
&= h_1h_2,
\end{aligned}$$

and therefore that

$$\begin{aligned}
T_2(g)\big|_{(0,0)}(h_1,h_2) &= g(0,0) + \nabla(g)\big|_{(0,0)} \cdot (x,y) + H(g)\big|_{(0,0)}(x,y) \\
&= \sin(0) + (0\cos(0), 0\sin(0)) \cdot (x,y) + xy \\
&= xy.
\end{aligned}$$

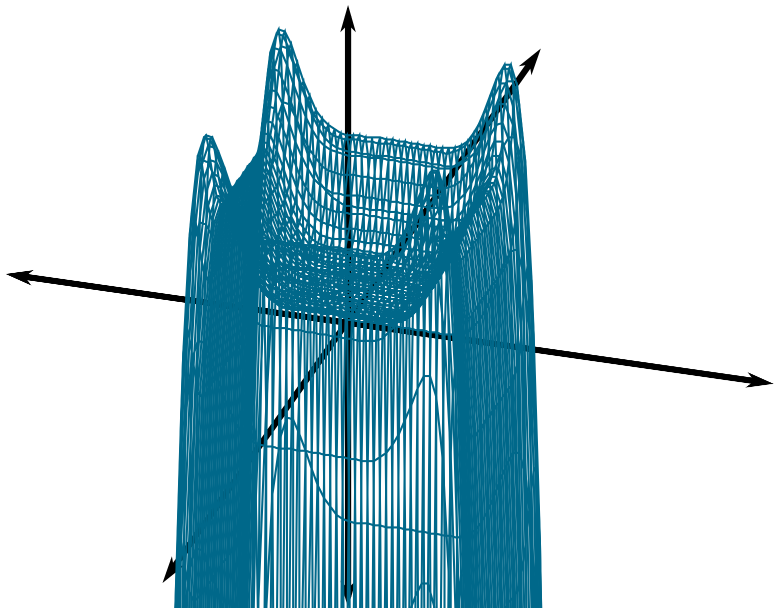
Therefore, the second-order approximation to $\sin(xy)$ at the origin is just $T_2(x,y) = xy$.

Example. (Using derivatives to study local extrema.) Let

$$f(x,y) = -(x^8+y^8) + 4(x^6+y^6) - 4(x^4+y^4).$$

Find all of the critical points of f, and classify them as local maxima, minima, or saddle points. Determine whether f has either a global maxima or minima, and if so identify these points.

Solution. First, we graph this function to get an idea of what's going on:



Roughly speaking, it looks like we have four global maxima, at least four saddle points between these maxima,and probably a bunch of weird things going on in the interior part of our function which are hard to determine from our picture.

So: picture aside, our task here is pretty immediate:

1. First, we want to calculate $\nabla(f)$, and find all of the points where it is either undefined or 0. These are our **critical points**.

2. We then want to calculate $H(f)$, the Hessian of f, for each critical point. If the Hessian is **positive-definite**[1], then we know that this point is a local minimum; if it is **negative-definite**, then it's a local maximum; if it **takes on both positive and negative values**, it's a saddle point; and if it's **identically 0**, we have no idea what's going on, and will need to explore its behavior using other methods.

So: by calculating, we can see that

$$D(f) = (-8x^7 + 24x^5 - 16x^3, -8y^7 + 24y^4 - 16y^3),$$

[1]We say that the Hessian is positive-definite if the associated matrix $\begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{bmatrix}$ of second partial derivatives is positive-definite: i.e. it has n eigenvalues and they're all positive. Negative-definite is similar, except we ask that all of the eigenvalues exist and are negative; if you're in neither case, but your matrix is not identically 0, then you're a saddle point.

and therefore that this is equal to 0 whenever

$$\begin{aligned}
&0 = -8x^7 + 24x^5 - 16x^3\\
\Leftrightarrow &x = 0, \text{ or}\\
&0 = -8x^4 + 24x^2 - 16\\
\Leftrightarrow &0 = (x^2-2)(x^2-1)\\
\Leftrightarrow &x = \pm\sqrt{2}, \pm 1,
\end{aligned}$$

and

$$\begin{aligned}
&0 = -8y^7 + 24y^5 - 16y^3\\
\Leftrightarrow &y = 0, \pm\sqrt{2}, \pm 1.
\end{aligned}$$

So we have twenty-five critical points, consisting of five choices of x and five choices of y. To classify these points, we look at the matrix of second-order-partials formed in the Hessian:

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{bmatrix} = \begin{bmatrix} -56x^6 + 120x^4 - 48x^2 & 0 \\ 0 & -56y^6 + 120y^4 - 48y^2 \end{bmatrix}.$$

When $x = \pm 1$, the polynomial $-56x^6+120x^4-48x^2$ is 16, which is positive; when $x = \pm\sqrt{2}$, this polynomial is -64, which is negative; finally, when $x = 0$ this polynomial is 0. Therefore, at the points

$$(\pm 1, \pm 1)$$

the Hessian is positive-definite, and therefore our function has a local minimum, while at the points

$$(\pm\sqrt{2}, \pm\sqrt{2})$$

the Hessian is negative-definite, and therefore our function has a local maximum, while at

$$(\pm\sqrt{2}, \pm 1), (\pm\sqrt{2}, 0), (\pm 1, \pm\sqrt{2}), (\pm 1, 0), (0, \pm\sqrt{2}), (0, \pm 1)$$

the Hessian is neither identically 0 nor positive- or negative-definite, and therefore our function has a saddle point.

This leaves just the point $(0, 0)$, at which the Hessian is identically 0 and therefore useless to us. There, we need to analyze how small changes in our function

$$f(x, y) = -(x^8 + y^8) + 4(x^6 + y^6) - 2(x^4 + y^4)$$

change its values.

If we wanted to argue that this function was a saddle point, we'd just have to find two paths leaving 0, one along which our function increased and another along which our function decreased. However, looking at the graph of the function, this actually seems like it's not the case. Rather, for very small values of x, y, we know that $x^4 \gg x^6, x^8$ and $y^4 \gg y^6, y^8$; therefore, very very close to the origin, our function is roughly just $-2(x^4+y^4)$, which is a upside-down parabola with a maximum at the origin. Therefore, we can see that this point is actually a local maximum, because (using our approximation) at all values very close to the origin that are not the origin, our function is roughly $-2(x^4+y^4)$ and therefore quite decidedly < 0, its value at the origin.

So we've classified all of our points into local maxima, minima, or saddle points. We now just need to decide whether any of them are global maxima or minima.

To do this: first, notice that as $(x, y) \to \infty$ along any path, our function goes to $-\infty$; this is because for sufficiently large values of x or y, the x^8, y^8 terms dominate our polynomial. Therefore, our function does not have a global minimum. More interestingly, this also tells us that our function **does** have a global maximum: because the values of our function go off to $-\infty$ as (x, y) go to infinity in every direction and our function is continuous and well-defined everywhere, it has nowhere to "go off to $+\infty$": given any sufficiently large cutoff radius R, we know that all of the values of f on points (x, y) further than R from the origin is as incredibly negative as we'd want, while f inside of R is a continuous function on a closed and bounded set, and therefore attains its global maximum.

Therefore, in particular we know that f attains its global maximum at one of our local maxima $(0, 0)$ or $(\pm\sqrt{2}, \pm\sqrt{2})$! Because f is the same value (namely, 0) at each of these points, each of these points is a place at which our function attains its global maxima, which is 0.