REVIEW OF CRITICAL POINTS, CONVEXITY AND INTEGRALS

KATHERINE DONOGHUE & ROB KUSNER

1. Gradient and Chain Rule. First a reminder about the gradient ∇f of a smooth function f. In $\mathbb{R}^n$, the *gradient operator* or **grad** has the form

$$\nabla = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \partial_{x_1} \\ \vdots \\ \partial_{x_n} \end{bmatrix}$$

Consider the special case of a function f on $\mathbb{R}^2$, where $\nabla f = \begin{bmatrix} f_x \\ f_y \end{bmatrix}$. For any (x_*, y_*) in the domain, $\nabla f(x_*, y_*)$ gives a vector which points in the direction of the greatest increase whose "tail" is based at the point (x_*, y_*). As the basepoint (x_*, y_*) varies, so does $\nabla f(x_*, y_*)$, giving a "vectorfield" ∇f on $\mathbb{R}^2$. (More about vectorfields later.)

One important property of gradient is that ∇f is perpendicular to the level sets of f.

Proof. A proof will be shown in $\mathbb{R}^3$, but can be easily generalized to $\mathbb{R}^n$.

Consider a level surface $\{(x, y, z) : f(x, y, z) = c\}$ of f on $\mathbb{R}^3$, and take a point $P_* = (x_*, y_*, z_*)$ that lies on this level surface, that is, $f(x_*, y_*, z_*) = c$. Any curve $\vec{p}(t) = (x(t), y(t), z(t))$ on the level surface with $\vec{p}(t_*) = P_* = (x_*, y_*, z_*)$ defines a tangent vector $\vec{p}\,'(t_*) = \langle x'(t_*), y'(t_*), z'(t_*)\rangle$ to the level surface of f at P_*.

On the other hand, composing $\vec{p}$ with f yields a constant function $g(t) = f(\vec{p}(t)) = f(x(t), y(t), z(t)) = c$, so differentiating $g(t) = c$ with respect to t using the chain rule gives the equation

$$\frac{dg}{dt} = \frac{\partial f}{\partial x}\bigg|_{P_*} \frac{dx}{dt}\bigg|_{t_*} + \frac{\partial f}{\partial y}\bigg|_{P_*} \frac{dy}{dt}\bigg|_{t_*} + \frac{\partial f}{\partial z}\bigg|_{P_*} \frac{dz}{dt}\bigg|_{t_*} = 0$$

Then,

$$\left\langle \frac{\partial f}{\partial x}\bigg|_{P_*}, \frac{\partial f}{\partial y}\bigg|_{P_*}, \frac{\partial f}{\partial x}\bigg|_{P_*} \right\rangle \cdot \left\langle \frac{dx}{dt}\bigg|_{t_*}, \frac{dy}{dt}\bigg|_{t_*}, \frac{dz}{dt}\bigg|_{t_*} \right\rangle = 0 \Longleftrightarrow \nabla f\big|_{P_*} \cdot \vec{p}\,'(t_*) = 0$$

The vanishing dot product means any tangent vector $\vec{p}\,'(t_*)$ to the level surface of f at P_* is perpendicular to $\nabla f\big|_{P_*}$. □

Date: DRAFT 25 September 2018.

2. Directional Derivatives and Chain Rule. If we define the *derivative operator*

$$D := \nabla^T = \begin{bmatrix} \partial_{x_1} & \cdots & \partial_{x_n} \end{bmatrix}$$

then the derivative of function f on $\mathbb{R}^n$ is the $1 \times n$ matrix field

$$D(f) := \begin{bmatrix} f_{x_1} & \cdots & f_{x_n} \end{bmatrix}.$$

For a general vector $\vec{v} = \langle v_1, v_2, \cdots, v_n \rangle$, the derivative of f along $\vec{v}$ is

$$D_{\vec{v}} f = D(f)\vec{v} = \nabla f \cdot \vec{v}.$$

For a function $g(x, y)$ and vector $\vec{v} = \langle v_1, v_2 \rangle$ in $\mathbb{R}^2$,

$$D_{\vec{v}} g = \begin{bmatrix} g_x & g_y \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \nabla g \cdot \vec{v} = g_x v_1 + g_y v_2.$$

Remark. Though some authors assume $\vec{v}$ must be a unit vector, it need not be. The length of the vector affects the value of the directional derivative of f: though $\vec{v_1} = \langle 1, 1 \rangle$ and $\vec{v_2} = \langle 2, 2 \rangle$ point in the same direction, the latter would produce a value of the directional derivative twice as big; in fact, you should check that our version of the directional derivative scales according to $D_{a\vec{v}} f = a D_{\vec{v}} f$ for any real number $a \in \mathbb{R}$.

Suppose instead of a vector, $\vec{v} = \langle v_1, v_2 \rangle$, we are given a path $\vec{p}(t) = \langle x(t), y(t) \rangle$. If we have a function f in $\mathbb{R}^2$, then the height of its graph $z = f(x, y)$ gives a height function $z(t) = f(x(t), y(t))$ whose rate of change

$$z'(t) = D_{\vec{p}'}(f) = \nabla f \cdot \vec{p}'$$

where $\vec{p}' = \begin{bmatrix} \frac{d}{dt}x & \frac{d}{dt}y \end{bmatrix} = \begin{bmatrix} x_t & y_t \end{bmatrix}$.

3. Hessian Matrix. Next we consider the second derivative or Hessian matrix $\nabla D(f)$. For f on $\mathbb{R}^2$,

$$\text{Hess}(f) := \nabla D(f) = \begin{bmatrix} \partial_x \\ \partial_y \end{bmatrix} \begin{bmatrix} f_x & f_y \end{bmatrix} = \begin{bmatrix} f_{xx} & f_{yx} \\ f_{xy} & f_{yy} \end{bmatrix}$$

and

$$\text{trace}(\text{Hess}(f)) := f_{xx} + f_{yy}.$$

The Hessian of a function f on $\mathbb{R}^n$ is an $n \times n$ matrix, while ∇f is an $n \times 1$ matrix and $D(f)$ is a $1 \times n$ matrix.

Example. Find Hess(g) and trace(Hess(g)) for $g(x, y) = \sin(x^2 + y^2)$.

Solution: $\text{Hess}(g) = \begin{bmatrix} \partial_x^2(\sin(x^2+y^2)) & \partial_x\partial_y(\sin(x^2+y^2)) \\ \partial_y\partial_x(\sin(x^2+y^2)) & \partial_y^2(\sin(x^2+y^2)) \end{bmatrix}$

$$= \begin{bmatrix} 2\sin(x^2+y^2) + 4x^2\cos(x^2+y^2) & -4xy\sin(x^2+y^2) \\ -4xy\sin(x^2+y^2) & 2\sin(x^2+y^2) + 4y^2\cos(x^2+y^2)) \end{bmatrix}$$

$\text{trace}(\text{Hess}(g)) = \partial_x^2(\sin(x^2+y^2)) + \partial_y^2(\sin(x^2+y^2)) = 4\sin(x^2+y^2) + 4(x^2+y^2)\cos(x^2+y^2))$

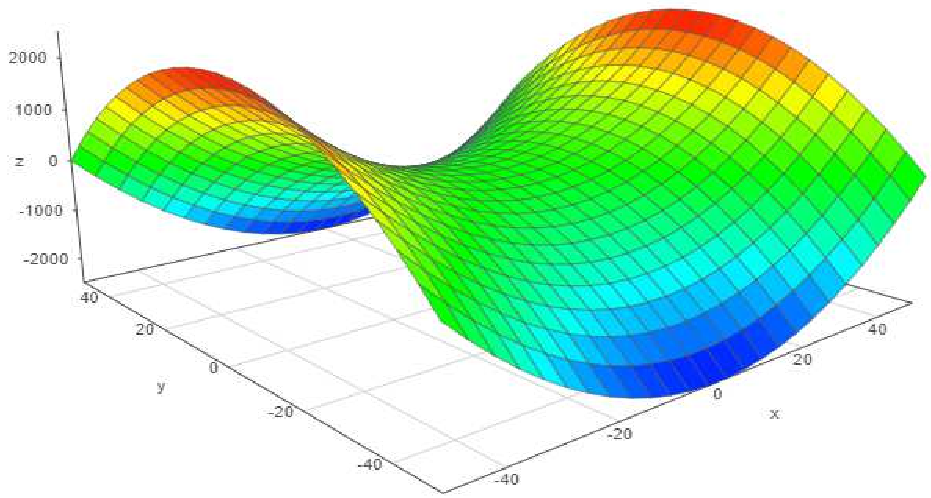

FIGURE 1. $f(x,y) = x^2 - y^2$

Note that $g_{xy} = g_{yx}$ — this is Clairut's Theorem, true for any sufficiently smooth function g — and thus at any point (x_*, y_*) the matrix $\text{Hess}(g) = \text{Hess}(g)^T$ is symmetric. From linear algebra, this means $\text{Hess}(g)$ is diagonalizable with respect to an orthornormal basis of eigenvectors, and it has real eigenvalues.

Challenge: find an insufficiently smooth function whose Hessian is not symmetric!

4. Critical Points and Lagrange Multipliers. Critical points are points that are candidates for being maxima, minima, or saddle points on a surface. As a reminder, saddle points occur when there is a maximum in one direction and a minimum in another direction. A classic example of a saddle is the function $f(x,y) = x^2 - y^2$ seen in Figure 1.

For a function $f(x)$ of one variable, the critical points x_* are where $f'(x_*) = 0$. For a function $f(x,y)$ on $\mathbb{R}^2$, finding the critical points (x_*, y_*) means solving two equations: both $f_x(x_*, y_*) = 0$ and $f_y(x_*, y_*) = 0$ are satisfied; or equivalently, the gradient vectorfield vanishes: $\nabla f = \vec{0}$.

Example. Using the same equation referenced in Sec. 3, $g(x,y) = \sin(x^2 + y^2)$, find all critical points.

Solution: $\nabla g = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} 2x\cos(x^2+y^2) \\ 2y\cos(x^2+y^2) \end{bmatrix}$

$g_x = 0$ when $x = 0$ or $x^2 + y^2 = n\pi/2$, where $n \in 2\mathbb{Z}+1$, the set of all odd numbers.

$g_y = 0$ when $y = 0$ or $x^2 + y^2 = n\pi/2$, where $n \in 2\mathbb{Z}+1$.

Thus the critical points are $(0,0)$ and all points along the infinitely many circles defined by $x^2 + y^2 = n\pi/2$, where $n \in 2\mathbb{Z}+1$ is odd. This periodic behavior can be seen in the graph of $g(x,y)$ given by Figure 2.

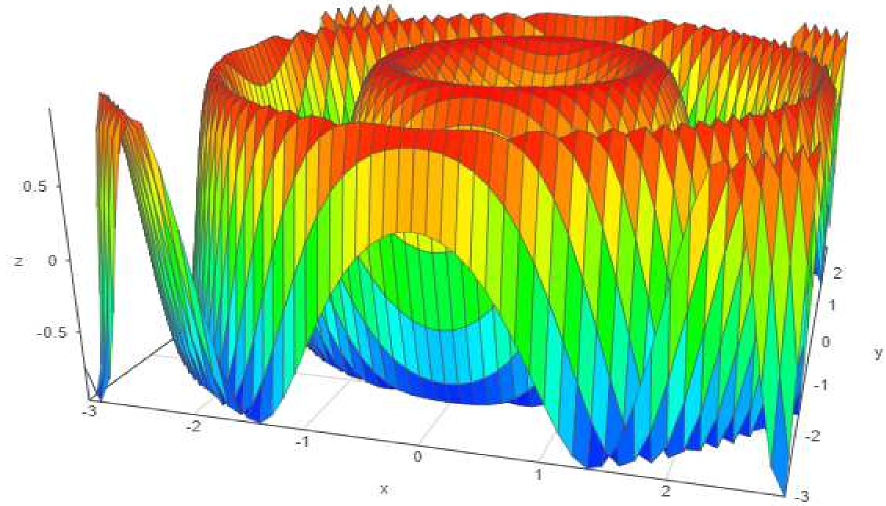

FIGURE 2. $g(x, y) = \sin(x^2 + y^2)$

Remark. Remember that cos, sin, tan and their inverses are periodic functions, that is that $\cos(\pi/2) = \cos(3\pi/2) = \cos(n\pi/2)\, \forall n \in 2\mathbb{Z} + 1$, $\sin(0) = \sin(\pi) = \sin(n\pi)\, \forall n \in \mathbb{Z}$ and so on. This means that functions involving sin, cos and other trig functions can easily have infinitely many critical points and the rules for those critical points must be noted when finding all critical points, similar to how we defined the equations for the circles in the above example.

We have been able to find global maxima and minima using $\nabla f = \vec{0}$, but what if we have a specified domain? For example, what if we want to find the critical points of $f(x, y) = xy$, but only along the curve $3x^2 + y^2 = 6$? The Lagrange multipliers method helps us maximize or minimize functions like $f(x, y, z)$ with a given constraint like $g(x, y, z) = c$.

The method involves solving the following system of equations,

$$\begin{aligned} f_x(x, y, z) &= \lambda g_x(x, y, z) \\ f_y(x, y, z) &= \lambda g_y(x, y, z) \\ f_z(x, y, z) &= \lambda g_z(x, y, z) \\ g(x, y, z) &= c \end{aligned}$$

where λ is a scalar. It is very important to be sure to exhaust all possiblities, which can be a tedious task. A few tips include:

(1) Since λ is an arbitrary scalar, solve for λ in terms of x, y, and z to eliminate λ from the equations
(2) Solve for one variable in terms of the others
(3) Remember that whenever you take a square root, consider both the positive and negative square roots

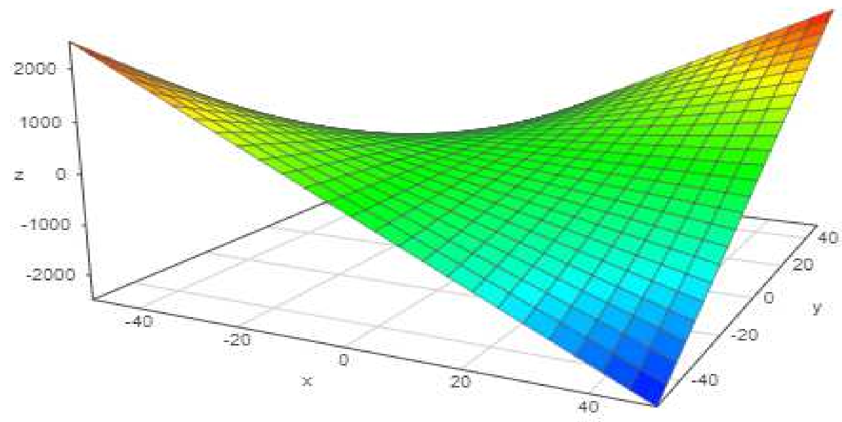

FIGURE 3. $f(x, y) = xy$

(4) Remember that whenever you divide an equation by an expression, you must be sure that the denominator is not zero. You could split the problem into two cases, solving while assuming that the expression is 0 and solving while assuming the expression is not zero

(5) Remember, as notted above, that $\sin, \cos, \tan, \arcsin, \ldots,$ are periodic functions and you must consider all multiples of $\pi, \pi/2$, or any other relevant multiple to ensure you have covered all possible cases (*Hint* : examine different cases of odd or even multiples)

Remark. There's an interesting and important "application" of Lagrange multipliers to the problem of finding critical points for a quadratic function $q(\vec{x})$ on $\mathbb{R}^n$ constrained to the $(n-1)$-sphere of unit vectors. Generically there will be exactly n pairs of critical points $\{\pm\vec{u_1}, \cdots, \pm\vec{u_n}\}$, where each such pair corresponds to a (unit length) eigenvector $\vec{u_i}$ of the symmetric $n \times n$ matrix S defining the given quadratic function via $q(\vec{x}) := \vec{x} \cdot S\vec{x}$. The corresponding eigenvalue $\lambda_i = q(\vec{u_i})$. (Maybe that's why λ is used for both Lagrange multipliers and eigenvalues?!)

Example. Use Lagrange multipliers to find all critical points of $f(x, y) = xy$, seen in Figure 3 on the curve $3x^2 + y^2 = 6$.

We must solve the following system:

$$\begin{aligned} y &= \lambda 6x \\ x &= \lambda 2y \\ 3x^2 + y^2 &= 6 \end{aligned}$$

While $(0, 0)$ solves the first two equations, it does not solve the third. Trying tip (1) on the first equation gives $\lambda = y/6x$. Suppose we use $x = 0$, then based on the third constraint equation, $y = \pm\sqrt{6}$, but this would not satisfy the first two equations. Thus $x = 0$ is not part of any critical point and we can use $\lambda = y/6x$ without any issues. Plugging $\lambda = y/6x$ into the second equation gives $x = y^2/3x$ which then gives $3x^2 = y^2$. This fits nicely into our constraint equation yielding $2y^2 = 6$ or $6x^2 = 6$ where both yield the same results. First taking $2y^2 = 6$ gives $y = \pm\sqrt{3}$. Plugging this into the original constraint equation gives $x = \pm 1$. Now we have 4 possible critical points, $(-1, -\sqrt{3}), (-1, \sqrt{3}), (1, -\sqrt{3})$, and $(1, \sqrt{3})$. Solving for λ in the second equation yields the exact same results. Knowing that $3x^2+y^2 = 6$ is the equation for an ellipse along with looking at the graph of $f(x, y)$, whose saddle-like nature is given in Figure 3, only 4 critical points are expected. So, all of the critical points along the curve are $(-1, -\sqrt{3}), (-1, \sqrt{3}), (1, -\sqrt{3})$, and $(1, \sqrt{3})$.

Now that we know how to find all critical points, we must identify these as maxima, minima, saddle points or otherwise.

If you are only tasked with finding global maxima and minima of a function, then you only need to plug your critical points into the original function and find which points give the largest and smallest function value.

If you are tasked with finding and identifying *all* critical points, the Hessian is used in a similar way as the second derivative test is for functions on $\mathbb{R}$.

For a function $f(x, y)$ and a critical point (x_*, y_*),

$$\begin{array}{lll} \text{if } \det(\text{Hess}(f(x_*, y_*)) > 0 \text{ and } h_{xx}(x_*, y_*), h_{yy}(x_*, y_*) < 0, & \text{then } (x_*, y_*) \text{ is a local maximum} \\ \text{if } \det(\text{Hess}(f(x_*, y_*)) > 0 \text{ and } h_{xx}(x_*, y_*), h_{yy}(x_*, y_*) > 0, & \text{then } (x_*, y_*) \text{ is a local minimum} \\ \text{if } \det(\text{Hess}(f(x_*, y_*)) < 0, & \text{then } (x_*, y_*) \text{ is a saddle point} \end{array}$$

In another way, if the Hessian is positive definite at (x_*, y_*), that is if the Hessian has positive eigenvalues, then (x_*, y_*) is a local minimum. If the Hessian is negative definite at (x_*, y_*), the Hessian has negative eigenvalues, then (x_*, y_*) is a local maximum. If the Hessian has both positive and negative eigenvalues at (x_*, y_*), then (x_*, y_*) is a saddle point. The only other possibility is the zero determinant. If this occurs at (x_*, y_*), then the Hessian is degenerate and the test is inconclusive.

Example. Let $f(x, y) = x^4 - 8x^2 + y^4 - 18y^2$.

(1) Find the critical points of $f(x, y)$
(2) Identify the critical points as maxima, minima, saddle points or degenerate
(3) Find the global minima of $f(x, y)$
(4) Does $f(x, y)$ have a global maximum? Justify your answer

Solution: (1) First to find the critical points. $\nabla f = \begin{bmatrix} 4x^3 - 16x \\ 4y^3 - 36y \end{bmatrix}$

Factoring f_x and f_y gives $f_x = 4x(x^2 - 4)$ and $f_y = 4y(y^2 - 9)$. From there, we get 9 critical points: $(-2, -3), (-2, 0), (-2, 3), (0, -3), (0, 0), (0, 3), (2, -3), (2, 0),$ and $(2, 3)$.

(2) $\text{Hess}(f) = \begin{bmatrix} 12x^2 - 16 & 0 \\ 0 & 12y^2 - 36 \end{bmatrix}$

Now to test each critical point:

$$(-2,-3): \begin{bmatrix} 32 & 0 \\ 0 & 72 \end{bmatrix}, \det(\text{Hess}) > 0 \text{ and } f_{xx} > 0, \text{ so } (-2,-3) \text{ is a local minimum}$$

$$(-2,0) \quad : \begin{bmatrix} 32 & 0 \\ 0 & -36 \end{bmatrix}, \det(\text{Hess}) < 0, \text{ so } (-2,0) \text{ is a saddle point}$$

$$(-2,3) \quad : \begin{bmatrix} 32 & 0 \\ 0 & 72 \end{bmatrix}, \det(\text{Hess}) > 0 \text{ and } f_{xx} > 0, \text{ so } (-2,3) \text{ is a local minimum}$$

$$(0,-3) \quad : \begin{bmatrix} -16 & 0 \\ 0 & 72 \end{bmatrix}, \det(\text{Hess}) < 0, \text{ so } (0,-3) \text{ is a saddle point}$$

$$(0,0) \quad : \begin{bmatrix} -16 & 0 \\ 0 & -36 \end{bmatrix}, \det(\text{Hess}) > 0 \text{ and } f_{xx} < 0, \text{ so } (0,0) \text{ is a local maximum}$$

$$(0,3) \quad : \begin{bmatrix} -16 & 0 \\ 0 & 72 \end{bmatrix}, \det(\text{Hess}) < 0, \text{ so } (0,3) \text{ is a saddle point}$$

$$(2,-3) \quad : \begin{bmatrix} 32 & 0 \\ 0 & 72 \end{bmatrix}, \det(\text{Hess}) > 0 \text{ and } f_{xx} > 0, \text{ so } (2,-3) \text{ is a local minimum}$$

$$(2,0) \quad : \begin{bmatrix} 32 & 0 \\ 0 & -36 \end{bmatrix}, \det(\text{Hess}) < 0, \text{ so } (2,0) \text{ is a saddle point}$$

$$(2,3) \quad : \begin{bmatrix} 32 & 0 \\ 0 & 72 \end{bmatrix}, \det(\text{Hess}) > 0 \text{ and } f_{xx} > 0, \text{ so } (2,3) \text{ is a local minimum}$$

(3) Using the local minima found in (2), we will plug them back into the original equation to see which are global minima. Since $f(x,y)$ has only even coefficients, $f(-2,-3) = f(-2,3) = f(2,-3) = f(2,3) = 16-32+81-182$, and thus all the local minima are the global minima.

(4) Since $f(x,y)$ has strictly even coefficients, this means the function grows unbounded as $x \to \pm\infty$ and $y \to \pm\infty$, thus $(0,0)$ is only local maxima.

Lastly, the Hessian can also be used to test the convexity of a function. Instead of testing individual points found using $\nabla f = 0$, like when identifying critical points, you will be finding where on the graph $\det(\text{Hess}(f)) > 0$ and where $\det(\text{Hess}(f)) < 0$.

Example. Let $f(x,y) = x^4 - 8x^2 + y^4 - 18y^2$. Identify where the graph of f is convex-up, convex-down (concave) or otherwise.

Solution: $\text{Hess}(f) = \begin{bmatrix} 12x^2 - 16 & 0 \\ 0 & 12y^2 - 36 \end{bmatrix}$

We are tasked with finding where $\det(\text{Hess}(f)) < 0$ and where $\det(\text{Hess}(f)) > 0$.

$\det(\text{Hess}(f)) = 48(3x^2-4)(y^2-3) = 0$ when $y = \pm\sqrt{3}$ or $x = \pm\sqrt{\frac{4}{3}}$ which means we have 9 cases to check:

(1) $y > +\sqrt{3}$ and $x > +\sqrt{\frac{4}{3}}$
(2) $y > +\sqrt{3}$ and $+\sqrt{\frac{4}{3}} > x > -\sqrt{\frac{4}{3}}$
(3) $y > +\sqrt{3}$ and $x < -\sqrt{\frac{4}{3}}$
(4) $+\sqrt{3} > y > -\sqrt{3}$ and $x > +\sqrt{\frac{4}{3}}$
(5) $+\sqrt{3} > y > -\sqrt{3}$ and $+\sqrt{\frac{4}{3}} > x > -\sqrt{\frac{4}{3}}$
(6) $+\sqrt{3} > y > -\sqrt{3}$ and $x < -\sqrt{\frac{4}{3}}$
(7) $y < -\sqrt{3}$ and $x > +\sqrt{\frac{4}{3}}$
(8) $y < -\sqrt{3}$ and $+\sqrt{\frac{4}{3}} > x > -\sqrt{\frac{4}{3}}$
(9) $y < -\sqrt{3}$ and $x < -\sqrt{\frac{4}{3}}$

One benefit of our function is that since both x and y are squared, the sign on the number does not matter, only the magnitude does. Thus, we only have 4 cases:

(1) $y > \sqrt{3}$ and $x > \sqrt{\frac{4}{3}}$
(2) $y > \sqrt{3}$ and $\sqrt{\frac{4}{3}} > x > 0$
(3) $\sqrt{3} > y > 0$ and $x > \sqrt{\frac{4}{3}}$
(4) $\sqrt{3} > y > 0$ and $\sqrt{\frac{4}{3}} > x > 0$

(1) Since $\det(\text{Hess}(f)) > 0$ and f_{xx} and f_{yy} are both positive, the graph is convex-up.

(2) Since $\det(\text{Hess}(f)) < 0$, the graph is saddle-like.

(3) Since $\det(\text{Hess}(f)) < 0$, the graph is saddle-like.

(4) Since $\det(\text{Hess}(f)) > 0$ and f_{xx} and f_{yy} are both negative, the graph is convex-down (concave).

5. Problem Set I.

The following problems concern the functions

$$h(x,y) := xy,\ e(x,y) := (\sin x)(\sin y),\ f(x,y) := \sin(xy) \text{ and } p(x,y) := x^5 - y^4$$

on $\mathbb{R}^2$ (or on the specified subdomains). We have explored h and p in class, whose graphs resemble a saddle and the weird pen-holder we passed around in class; the graph of e should remind you of the egg-crate we also passed around, and you are welcome to draw or describe the graph of f for extra credit!

Problem 5.1. *Suppose you're hiking along the graph of the function h above the path $(x(t), y(t)) := (t^2, t^3)$ in $\mathbb{R}^2$. At what rate is your height $z(t) = h(x(t), y(t))$ changing when $t = 0$? At what rate is it changing when $t = 1$? (You can do this various ways, but make sure one of the ways uses the chain rule and directional derivatives!)*

Problem 5.2. *For the functions e and f above, compute their various first and second partial derivatives (as functions of (x, y)), as well as their gradient vector fields and Hessian matrix fields (whose entries are also functions of (x, y) – we did this in class for the functions h and p). Find all the critical points for both e and f, and use their Hessians there to analyze whether these critical points are (local) minima, maxima, saddles or otherwise.*

Problem 5.3. *We saw (in class) that the only critical point of the function p is at $(0, 0)$ (it's degenerate – its Hessian is the zero matrix there), but we can also use its Hessian to find where the graph of p is convex-up, convex-down (concave) or otherwise – do that!*

Problem 5.4. *Find all the global maximum and global minimum* points *and* values *of the function h above restricted to the* unit square $Q^2 := [0, 1] \times [0, 1]$. *Do the same for the* unit disk $D^2 = \{(x, y) \mid x^2 + y^2 \leq 1\}$.

6. Basic Integrals. This section will be a review of double and triple integrals from Calculus 2. As you should remember, double and triple integrals are not much more difficult than a single integral. They are simply computing two or three integrals in a row rather than one. In some important special cases, they turn out to be products of single integrals, just like the simple way we might compute the area of a rectangle or volume of a box.

Example. For the *unit square* $Q^2 := [0, 1] \times [0, 1] = \{(x, y) \mid 0 \leq x,\ y \leq 1\}$, find $\iint\limits_{Q^2} f(x, y)\,\mathrm{d}x\,\mathrm{d}y$ for $f(x, y) = x^4 - 8x^2 + y^4 - 18y^2$

Solution: $\iint\limits_{Q^2} x^4 - 8x^2 + y^4 - 18y^2\,\mathrm{d}x\,\mathrm{d}y = \int_{y=0}^{y=1} \left[\int_{x=0}^{x=1} x^4 - 8x^2 + y^4 - 18y^2\,\mathrm{d}x\right]\mathrm{d}y$

$= \int_{y=0}^{y=1} \left[\frac{1}{5}x^5 - \frac{8}{3}x^3 + xy^4 - 18xy^2\Big|_{x=0}^{x=1}\right]\mathrm{d}y = \int_{y=0}^{y=1} \left[\frac{1}{5} - \frac{8}{3} + y^4 - 18y^2\right]\mathrm{d}y$

$= \int_{y=0}^{y=1} y^4 - 18y^2 - \frac{37}{15}\mathrm{d}y = \frac{1}{5}y^5 - 6y^3 - \frac{37}{15}y\Big|_{y=0}^{y=1} = \frac{1}{5} - 6 - \frac{37}{15} = \boxed{-\frac{124}{15}}$

In certain cases, *Fubini's Theorem* can help simplify complicated, iterated integrals; it holds much more generally (not just for rectangular boxes).

Fubini's Theorem: If $f(x, y)$ is continuous over the region, R, defined by $a \leq x \leq b$ and $c \leq y \leq d$, then

$$\iint_R f(x, y)\,\mathrm{d}A = \int_a^b \int_c^d f(x, y)\,\mathrm{d}y\,\mathrm{d}x = \int_c^d \int_a^b f(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

This simply means that the order of integration does not matter. Let's take the same example from above and reverse the integration order:

Example. For the *unit square* $Q^2 := [0, 1] \times [0, 1] = \{(x, y) \mid 0 \leq x,\ y \leq 1\}$, find $\iint\limits_{Q^2} f(x, y)\,\mathrm{d}y\,\mathrm{d}x$ for $f(x, y) = x^4 - 8x^2 + y^4 - 18y^2$

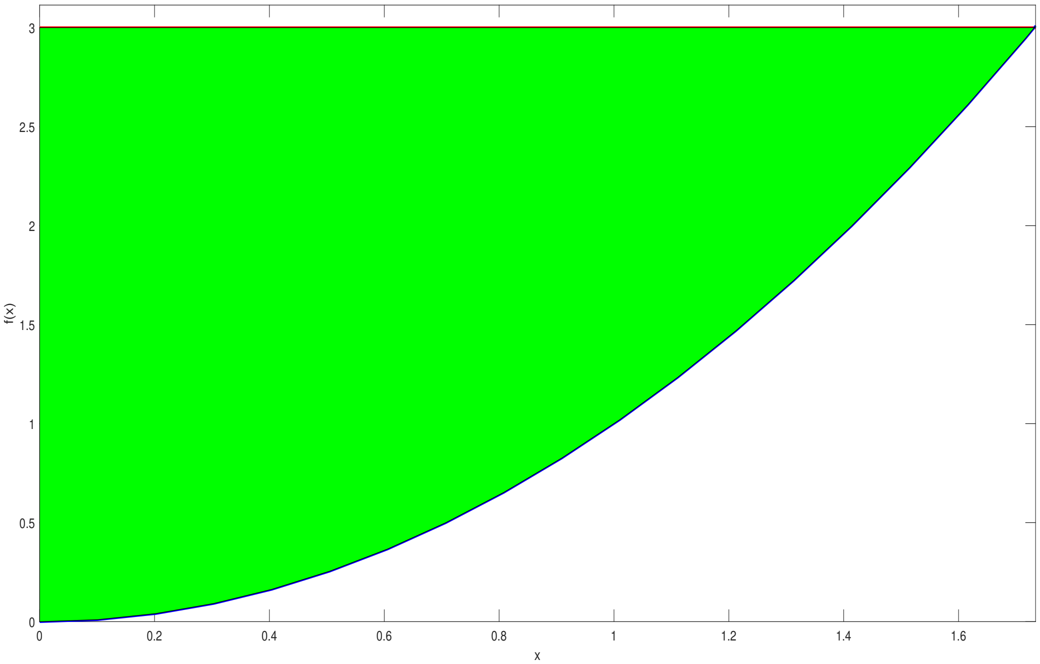

FIGURE 4. The green shaded region in the first quadrant between $\{y = x^2\}$ and $\{y = 3\}$.

Solution: $\iint\limits_{Q^2} x^4 - 8x^2 + y^4 - 18y^2 \,\mathrm{d}x\,\mathrm{d}y = \int_{x=0}^{x=1} \left[\int_{y=0}^{y=1} x^4 - 8x^2 + y^4 - 18y^2 \,\mathrm{d}y \right] \mathrm{d}x$

$$= \int_{x=0}^{x=1} \left[x^4 y - 8x^2 y + \tfrac{1}{5}y^5 - 6y^3 \Big|_{y=0}^{y=1} \right] \mathrm{d}x = \int_{x=0}^{x=1} \left[x^4 - 8x^2 + \tfrac{1}{5} - 6 \right] \mathrm{d}x$$

$$= \int_{x=0}^{x=1} x^4 - 8x^2 + \tfrac{29}{5} \,\mathrm{d}x = \tfrac{1}{5}x^5 - \tfrac{8}{3}x^3 - \tfrac{29}{5}x \Big|_{x=0}^{x=1} = \tfrac{1}{5} - \tfrac{8}{3} - \tfrac{29}{5} = \boxed{-\frac{124}{15}}$$

Being able to change the order of integration can help simplify problems. For example, suppose instead of integrating over $Q^2 = [0,1] \times [0,1] = \{(x,y) \mid 0 \le x,\ y \le 1\}$, we had to integrate over a region $P := [0,1] \times \left[\frac{1}{\sqrt{2}}, \frac{9}{4}\right] = \{(x,y) \mid 0 \le x \le 1,\ \frac{1}{\sqrt{2}} \le y \le \frac{9}{4}\}$. It would probably be simpler to integrate with respect to x first instead of dealing with the messy fractions of the y bound before the end.

Fubini's Theorem also holds for domains which are not rectangular, but here one needs to use more care in expressing the limits of integration in an iterated integral — indeed, changing the order can turn a tricky integral into child's play, or turn an intractible integral into one that's doable. (Of course, it can go the other way as well if one is careless!)

For example, suppose we look at a region P in the first quadrant between the graphs $\{y = x^2\}$ and $\{y = 3\}$, the shaded area in Figure 4. The region can be defined in two different but equivalent ways: $P = \{(x,y) \mid 0 \le x \le \sqrt{y},\ 0 \le y \le 3\}$ or $P = \{(x,y) \mid x^2 \le y \le 3,\ 0 \le x \le \sqrt{3}\}$. If we integrate a function $f(x,y)$ over P with respect to x first, we use the first representation of the region P, since the interval of x integration depends on y

(and because the limits of the x-integral cannot have the x variable left in them):

$$\int_0^3 \int_0^{\sqrt{y}} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

Similarly, the second representation of the region P is used if we integrate with respect to y first:

$$\int_0^{\sqrt{3}} \int_{x^2}^{3} f(x,y)\,\mathrm{d}y\mathrm{d}x.$$

And for an example that shows how changing the order can simplify things, consider this integral over the first quadrant of the unit disk

$$\iint_{Q^2 \cap D^2} \sqrt{x}y\,\mathrm{d}x\mathrm{d}y.$$

In one order, this is

$$\int_0^1 \int_0^{\sqrt{1-y^2}} \sqrt{x}y\,\mathrm{d}x\mathrm{d}y = \frac{2}{3}\int_0^1 (1-y^2)^{\frac{3}{2}} y\,\mathrm{d}y,$$

which requires a substitution like $u = 1 - y^2$ to integrate. But in the other order, it becomes

$$\int_0^1 \int_0^{\sqrt{1-x^2}} \sqrt{x}y\,\mathrm{d}y\mathrm{d}x = \int_0^1 \sqrt{x}\frac{y^2}{2}\bigg|_{y=0}^{y=\sqrt{1-x^2}} \mathrm{d}x = \frac{1}{2}\int_0^1 \sqrt{x}(1-x^2)\,\mathrm{d}x = \frac{1}{2}\int_0^1 x^{\frac{1}{2}} - x^{\frac{5}{2}}\,\mathrm{d}x,$$

which is easy. Of course, in more variables, there are more ways to change the order of integration, so there's an "art" to doing this.

Here's a useful trick for computing multiple integrals over a rectangular domain $[a,b]\times[c,d]$ of functions $f(x,y) = g(x)h(y)$ which are products of functions of a single variable:

$$\int_a^b \int_c^d f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \Big(\int_a^b g(x)\,\mathrm{d}x\Big)\Big(\int_c^d h(y)\,\mathrm{d}y\Big).$$

In other words, the multiple integral becomes a product of single integrals. This generalizes to any number of variables, and also works for more general product domains $R = P_1 \times \cdots \times P_k$ in many variables.

7. Volume, Average, and Center of Mass. We are interested in measuring the volume $\mathrm{vol}(R) = \iiint_R 1\,\mathrm{d}x\mathrm{d}y\mathrm{d}z$ of a region $R \subset \mathbb{R}^3$, computing related quantities like the *average*

$$\bar{f} := \frac{1}{\mathrm{vol}(R)} \iiint_R f(x,y,z)\,\mathrm{d}x\ \mathrm{d}y\mathrm{d}z$$

of a function $f : R \to \mathbb{R}$, especially the coordinate funtions whose averages combine into the *center of mass* $(\bar{x}, \bar{y}, \bar{z})$ of R. (We should really call it the *center of volume*, but since nobody does, neither will we!)

Note that a constant function f has average $\bar{f} = f$, and more generally, if $a \le f \le b$ then $a \le \bar{f} \le b$. This implies that if our region is contained in a box $R \subset [a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]$, then its center of mass is also inside the box

$$a_1 \le \bar{x} \le b_1, a_2 \le \bar{y} \le b_2, a_3 \le \bar{z} \le b_3,$$

which is a useful estimate to know if you are actually computing these.

How might we compute these integrals? If R is *simple* in the z direction, meaning it can be described as the region be between the lower graph $\{z = g(x, y)\}$ and upper graph $\{z = h(x, y)\}$ of functions over a domain $P \subset \mathbb{R}^2$, then we can integrate first in the z direction to get

$$\operatorname{vol}(R) = \iint_P \int_{g(x,y)}^{h(x,y)} 1 \, \mathrm{d}z\mathrm{d}x\mathrm{d}y = \iint_P z\Big|_{g(x,y)}^{h(x,y)} \mathrm{d}x\mathrm{d}y = \iint_P h(x, y) - g(x, y) \, \mathrm{d}x\mathrm{d}y.$$

(This expresses the geometrically obvious fact that if we slice a "potato" R into very skinny vertical "fries" of height approximately $h(x, y) - g(x, y)$ at the point $(x, y) \in P$, then the total volume of the potato is the sum of the volumes of the fries, which becomes the integral of $h(x, y) - g(x, y)$ over the horizontal slice P in the limit of infinitely skinny fries!)

In many examples, we take $g(x, y) = 0$ and the region R is simply between a domain $P = \operatorname{graph}(0)$ in the (x, y)-plane $\mathbb{R}^2$ and $\operatorname{graph}(h) = \{z = h(x, y)\}$, which we will assume for the rest of the discussion. To compute the integral of a function f over R which does *not* depend in z, we can again perform the z integral first, and observe that

$$\iiint_R f \, \mathrm{d}z\mathrm{d}x\mathrm{d}y = \iint_P f(x, y)h(x, y) \, \mathrm{d}x\mathrm{d}y.$$

In other words, the triple-integral of z-independent f over R becomes a double-integral weighted by the height h of the vertical slices. For example, if we take the special case $f = x$, which arises when computing the center of mass, then

$$\iiint_R \mathrm{d}z\mathrm{d}x\mathrm{d}y = \iint_P xh(x, y) \, \mathrm{d}x\mathrm{d}y.$$

On the other hand, if the function f depends on z, then we need to return to the triple integral, and see what performing the z integration first does to f. For example, in the special case $f = z$, which also arises when computing the center of mass, we need to integrate the *squared* height over the horizontal slice P:

$$\iiint_R z \, \mathrm{d}z\mathrm{d}x\mathrm{d}y = \iint_P \int_0^{h(x,y)} \mathrm{d}z\mathrm{d}x\mathrm{d}y = \iint_P \frac{z^2}{2}\Big|_0^{h(x,y)} \mathrm{d}x\mathrm{d}y = \frac{1}{2} \iint_P h^2(x, y) \, \mathrm{d}x\mathrm{d}y.$$

Both of these observations will be useful in the following problems.

8. Problem Set II. In the problems below, we continue using the familiar functions

$$h(x, y) := xy, \quad e(x, y) := (\sin x)(\sin y), \quad \text{and} \quad p(x, y) := x^5 - y^4,$$

and the familiar subdomains of the (x, y)-plane $\mathbb{R}^2$:

- the *unit square* $Q^2 := [0, 1] \times [0, 1] = \{(x, y) \mid 0 \le x, \; y \le 1\}$,
- the *unit disk* $D^2 := \{(x, y) \mid x^2 + y^2 \le 1\}$, and
- the first quadrant of the unit disk $Q^2 \cap D^2 = \{(x, y) \mid 0 \le x, \, y, \, x^2 + y^2 \le 1\}$.

We will explore the 3-dimensional regions R between their graphs in $\mathbb{R}^3$ and various subdomains P of $\mathbb{R}^2$ which we view as the plane $\{z = 0\} \subset \mathbb{R}^3$.

For *each* problem:

- Express the volume of R as an integral over P, and evaluate it as an iterated integral. [Hint: for some problems we may need to figure out P, or make substitutions, or use polar coordinates...!]
- Find the center of mass $(\bar{x}, \bar{y}, \bar{z})$ for the region R, where $\bar{f}$ means the *average* of the function f over the region R. [Hint: symmetry helps check $\bar{x}$ or $\bar{y}$ without computation of integrals; $\bar{z}$ is trickier to compute: reduce it to (a multiple of) the integral over P of the *square* of the function whose graph defines R!]

Problem 8.1. *Let $R \subset \mathbb{R}^3$ be the region between $P = Q^2 = [0,1] \times [0,1] \subset \mathbf{R}^2$ and* $\mathrm{graph}(h)$*.*

[Hint: both h and $P = Q^2$ are symmetric with respect to the line $\{x = y\}$!]

Problem 8.2. *Let $R \subset \mathbb{R}^3$ be the region between $P = Q^2 \cap D^2$ and* $\mathrm{graph}(h)$*.*

[Hint: consider polar coordinates as well as symmetry with respect to $\{x = y\}$!]

Problem 8.3. *Let $R \subset \mathbb{R}^3$ be the region between $P = \pi Q^2 = [0,\pi] \times [0,\pi] \subset \mathbf{R}^2$ (the unit square rescaled by π) and* $\mathrm{graph}(e)$*.*

[Hint: symmetry with respect to $\{x = y\}$ as well as $\{x = \frac{\pi}{2}\}$ and $\{y = \frac{\pi}{2}\}$; the integral $\int \sin^2 t\, dt$ over a half-period is easy to compute!]

Problem 8.4. *Let $R \subset \mathbb{R}^3$ be the region between* $\mathrm{graph}(p)$ *and $P = \{(x,y) |\ p(x,y) \geq 0\} \cap Q^2$, the subdomain of the unit square where p is non-negative.*

[Hint: first find the curved part of the boundary of P where $p(x,y) = 0$; this gives limits of integration depending on the order one does the iterated integral; one order is much easier than the other, but Fubini's Theorem certifies they are equal!]

Mathematics & Statistics, University of Massachusetts, Amherst, MA 01003, USA
Email address: profkusner@gmail.com, kdonoghue@umass.edu