Critical Points

A point P is a **critical point** of f if grad $f(P) = O$. Equivalently, all the partial derivatives $D_i f$ are 0 at P.

Example. Find the critical points of $f(x, y) = e^{-(x^2 + y^2)}$. We take partial derivatives and set them to 0 to find the critical points.

As in the single variable case, we can have a variety of behaviors at a critical point; we do not necessarily have a local minimum or local maximum.

Let f be defined on an open set U. A point P is called a **local maximum** of f if, in some neighborhood N of P, we have

$$f(X) \leq f(P)$$

for all $X \in N$.

The concept of local minimum is defined similarly.

Theorem. Let f be a differentiable function on U. Let P be a local maximum. Then P is a critical point of f.

The proof of this amounts to reducing it to a one variable problem. If H is a nonzero vector, and t is small enough, then $P + tH \in U$. Moreover, if t is small enough, $P + tH$ will land in the neighborhood mentioned in the definition, so that

$$f(P + tH) \leq f(P)$$

for all t in an interval of the form $(-\delta, \delta)$, $\delta > 0$. So $g(t) = f(P + tH)$ has a local maximum at $t = 0$. Thus $g'(t) = 0$. By the chain rule,

$$\text{grad } f(P) \cdot H = 0.$$

This is true for all H, so we must have grad $f(P) = 0$. ∎

A similar argument shows that local minima are also critical points of f.

Boundary, Interior, etc.

An **open ball** of radius $r > 0$ in $\mathbb{R}^n$ centered at P is defined to be the set of all points X such that $\|X - P\| < r$.

A **closed ball** is similarly defined except $\|X - P\| \leq r$ (rather than strict inequality).

A subset $U \subseteq \mathbb{R}^n$ is **open** if at every point $P \in U$, there is a ball of some radius around P contained entirely in U.

Note that open balls are indeed open sets.

An **interior point** P of a set $S \subseteq \mathbb{R}^n$ is one such there exists a ball of some radius around P contained entirely in S. Thus one could rephrase the definition of openness as each point being an interior point.

A point P (not necessarily in S) is called a **boundary point** of S if every open ball around P contains both a point in S and a point not in S.

A set is **closed** if it contains all of its boundary points. We note that a closed ball is indeed closed.

A set is **bounded** if one can fit the set inside a ball. Equivalently, S is bounded if there is some $b > 0$ such that $\|X\| \le b$ for all $X \in S$.

Theorem. Let f be a continuous function defined on a closed and bounded set S. Then f has a maximum and a minimum on S.

The proof of this requires things out of the scope of this course.

One can see via examples that a continuous f can fail to achieve extreme values if S is not closed and bounded. As an easy example, consider $f : (0,1) \to \mathbb{R}$, $f(x) = x$.

In a general situation for some f on a closed and bounded region S, we find the maxima and minima by looking at the critical points in the interior of S. One must also look at the values of f on the boundary, however.

Example. Find the maxima and minima of $f(x,y) = x^3 + xy$ on the square with vertices $(\pm 1, \pm 1)$. We do this by finding the critical points on the interior, and then also testing the boundary points by parametrizing the four sides that make up the square and plugging those parametrizations into f.

Example. Find the maximum of the function

$$f(x,y) = x^2 e^{-x^4 - y^2}.$$

This function goes to 0 as the distance from the origin goes to ∞. This implies that the function indeed achieves a maximum (as a result of the above theorem).

Lagrange Multipliers

Now suppose we are trying to find minima and maxima of a function $f : \mathbb{R}^n \to \mathbb{R}$, but have the added constraint that we can only plug in points lying on the curve/surface $g(X) = 0$ where g is some other function.

Theorem. Let f and g be differentiable functions on U with continuous partial derivatives. Let S be the set of points X where $g(X) = 0$ and $\operatorname{grad} g(X) \ne O$. Suppose that P is an extremum of f on S. Then there is some number λ such that

$$\operatorname{grad} f(P) = \lambda \operatorname{grad} g(P).$$

To get a sense of why this is true, let $X(t)$ be a differentiable curve passing through P, say $X(t_0) = P$. Then $f(X(t))$ has a maximum or minimum at t_0. So

$$\frac{d}{dt} f(X(t)) \bigg|_{t_0} = 0.$$

But the chain rule implies

$$0 = \frac{d}{dt} f(X(t)) \bigg|_{t_0} = \operatorname{grad} f(P) \cdot X'(t_0).$$

So $\operatorname{grad} f(P)$ is perpendicular to every curve on the surface passing through P. In other words, $\operatorname{grad} f(P)$ and $\operatorname{grad} g(P)$ are both perpendicular to the surface $g(X) = 0$ at the point P. So one is a multiple of the other.

This theorem gives us a way of finding the extrema of f subject to a constraint. The

2

theorem says that if you have an extreme value, this equation involving the gradients must hold. So one can solve this equation to get a list of points that are candidates for the extreme values and then figure out the max and min by inspection. In other words, the theorem says, "If there are any extreme values, this is where they would have to be."

Example. Find the maximum of the function $f(x, y,) = x + y$ subject to the constraint $x^2 + y^2 = 1$. We have $\operatorname{grad} f = (1, 1)$ and $\operatorname{grad} g = (2x, 2y)$. So suppose (x_0, y_0) is an extreme value. The equation $\operatorname{grad} f(x_0, y_0) = \lambda g(x_0, y_0)$ yields the two equations

$$1 = 2\lambda x_0, \ 1 = 2\lambda y_0.$$

So x_0 and y_0 are nonzero. Hence $\lambda = 1/(2x_0) = 1/(2y_0)$, meaning $x_0 = y_0$. This point must also satisfy $x_0^2 + y_0^2 = 1$, so that gives the possibilities

$$x_0 = \pm\frac{1}{\sqrt{2}}, y_0 = \pm\frac{1}{\sqrt{2}}.$$

So the two solutions to the system of equations are $(-1/\sqrt{2}, -1/\sqrt{2})$ and $(1/\sqrt{2}, 1/\sqrt{2})$. The maximum obviously corresponds to the latter, where f has the value $2/\sqrt{2}$.

Taylor's Theorem Revisited

Recall that for single variable smooth function f, Taylor's theorem gives us

$$f(a + h) = f(a) + f'(a)h + \frac{f''(a)}{2}h^2 + R_3,$$

where $R_3 = \frac{f^{(3)}(c)}{3!}h^3$ for some number c between a and $a + h$.

Let's see how to carry this over to two-variable functions. Let $P = (x_0, y_0)$ and $H = (h, k)$, and assume that $P \in U$ for some open set U on which f is defined with continuous partial derivatives up to order 3. We want to obtain a similar expansion to the one above. Define

$$g(t) = f(P + tH), \ t \in [0, 1].$$

Assume that $P + tH \in U$ for all $t \in [0, 1]$. We note that $g(0) = f(P)$, $g(1) = f(P + H)$. Then $g(t)$ is differentiable (by chain rule), and we can apply the single-variable Taylor's theorem:

$$g(1) = g(0) + g'(0) + \frac{g''(0)}{2} + R_3.$$

Here, we're using a value of $h = 1$. So our remainder R_3 has the form

$$R_3 = \frac{1}{3!}g^{(3)}(\tau),$$

where τ lies between 0 and 1.

Now we have

$$g'(t) = \operatorname{grad} f(P + tH) \cdot H = D_1 f(P + tH)h + D_2 f(P + tH)k,$$

which comes from the chain rule and then writing things out in coordinates. For $t = 0$, this

3

yields
$$g'(0) = D_1 f(P)h + D_2 f(P)k.$$

Let's rewrite $g'(t)$ as

$$g'(t) = f_1(P + tH), \ f_1 = hD_1 f(P + tH) + kD_2 f(P + tH).$$

We can then apply chain rule once more as

$$g''(t) = hD_1 f_1(P + tH) + kD_2 f_1(P + tH) = \left(h^2 D_1^2 f + 2hk D_1 D_2 f + k^2 D_2^2 f \right) \Big|_{P+tH}.$$

Putting it all together and substituting $t = 0$, we get

$$f(P + H) = f(P) + D_1 f(P)h + D_2 f(P)k$$
$$+ \frac{1}{2} \left[D_1^2 f(P)h^2 + 2D_1 D_2 f(P)hk + D_2^2 f(P)k^2 \right]$$
$$+ R^3.$$

If we write $P = (x_0, y_0)$ and use a slightly different notation, we get

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = f_x h + f_y k$$
$$+ \frac{1}{2} \left[f_{xx} h^2 + 2f_{xy} hk + f_{yy} k^2 \right]$$
$$+ R_3$$

where all the derivatives are understood to be evaluated at P.

Notice that at a critical point (i.e. where $f_x = f_y = 0$), the first order terms drop off, and so the behavior of the function is essentially determined by the term

$$Q(h, k) := ah^2 + 2bhk + ck^2$$

where $a = f_{xx}(P)$, $b = f_{xy}(P)$, $c = f_{yy}(P)$. Thus examining this **quadratic form** Q will reveal the local min/max behavior of the function at the point P. Note that if a, b, c are all 0, we're in a tough spot, because then we would have to look at the cubic terms in Taylor's formula, which is no fun. So we assume at least one of a, b, c is nonzero.

Quadratic Forms

As above let
$$Q(x, y) = ax^2 + 2bxy + cy^2.$$

A quadratic form of the type above is said to be **definite** if it attains values of one sign only (either all positive or all negative), except at $(0, 0)$, where Q is of course 0. $x^2 + y^2$ is an easy example.

$Q(x, y)$ is **indefinite** if it can assume values of both signs. xy is such an example.

$Q(x, y)$ is **semi-definite** if it vanishes for values other than $(x, y) = (0, 0)$ but otherwise only takes values of one sign. $(x - y)^2$ is an example of this.

Now, if $a = c = 0$, our form is $2bxy$ which is indefinite. So suppose one of a or c is nonzero. Suppose for concreteness that $a \neq 0$. Then

$$Q(x,y) = a \left[\left(x + \frac{b}{a}y \right)^2 + \frac{ca - b^2}{a^2}y^2 \right].$$

If $ca - b^2 > 0$, our form is definite (why?). If $ca - b^2 < 0$, it is indefinite. If $ca - b^2 = 0$ it is semidefinite.

If $a > 0$, then the definite and semidefinite cases yield a local min for $Q(x,y)$. The indefinite case yields a saddle point. If $a < 0$, then we have the same except replace the word max with max.

The semidefinite case is much harder to analyze. For instance, suppose a function has the Taylor expansion $f(x_0, y_0) = (h - k)^2 + k^3$. Although the corresponding quadratic form has a local min, the x^3 term complicates things. In general, the semidefinite case requires one to look at the higher order terms, so we put this case aside for now.

Example. $Q(x,y) = 3x^2 - 4xy - 4y^2$. We may complete the square and write this as

$$Q(x,y) = 3 \left[\left(x - \frac{2}{3}y \right)^2 - \frac{16}{9}y^2 \right].$$

So Q is indefinie.

Example. $Q(x,y) = -3x^2 + 5xy - 7y^2$. We can complete the square to get

$$Q(x,y) = -3 \left[\left(x - \frac{5}{6}y \right)^2 + \left(-\frac{25}{36} + \frac{7}{3} \right)y^2 \right].$$

Thus Q has a local max at the origin.