3.5 Quadratic Approximation and Convexity/Concavity

Overview: Second derivatives are useful for understanding how the linear approximation varies locally. In one variable calculus, this meant the bending of tangent lines and earlier in Chapter 2 this extended to the bending of space curves and curvature. This section generalizes those results to functions of several variables. The quadratic approximation is derived and used in the classification of critical points for max-min problems (in Section 3.6). The matrix of second derivatives, known as the Hessian, is symmetric when the derivatives are all continuous. The analysis of quadratic functions from Chapter 1 becomes a fundamental tool for describing behavior that is beyond the linear approximation, such as bending (convexity/concavity). While positivity or negativity of second derivatives involving one variable only are valuable information, the possibility of cross terms from the mixed derivatives makes the analysis more intricate.

3.5.1 Quadratic Approximation and the Hessian Matrix

Using second derivatives, a function $f(x, y)$ which is twice continuously differentiable can be approximated by a quadratic function, its Taylor polynomial of order 2. A relatively easy way to see how this gets done is to look at a quadratic function with constants denoted by letters and then look at its second derivatives. Consider the dimension 2 case of $f(x, y) = Ax^2 + Bxy + Cy^2$, then $\frac{\partial f}{\partial x}(x, y) = 2Ax + By$, $\frac{\partial f}{\partial y}(x, y) = Bx + 2Cy$, and therefore $\frac{\partial^2 f}{\partial x^2}(x, y) = 2A$, $\frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{\partial^2 f}{\partial y \partial x}(x, y) = B$, and $\frac{\partial^2 f}{\partial y^2}(x, y) = 2C$. This suggests a general quadratic approximation term at $(0, 0)$ of the form

$$Q(x, y) = \frac{1}{2}(\frac{\partial^2 g}{\partial x^2}(0, 0)x^2 + (2\frac{\partial^2 g}{\partial x \partial y}(0, 0)xy + \frac{\partial^2 g}{\partial y^2}(0, 0)y^2)$$

which can be seen to match the second derivatives of $g(x, y)$ at the point $(0, 0)$. Note that the mixed second derivative comes twice because of the equality of mixed partial derivatives.

Your turn: do the corresponding guess (and check it) for three variables x, y, z at $(0, 0, 0)$ – there are nine possible terms, but some get grouped in pairs. Start with $f(x, y, z) = Ax^2 + Bxy + Cxz + Dy^2 + Eyz + Fz^2$.

Combined with our earlier discussion of linear approximation, we find the Taylor polynomial of order 2 in several variables near a location (a, b) for a twice continuously differentiable function f:

$$\begin{aligned} P_2(x, y) &= f(a, b) + f_x(a, b) \cdot (x - a) + f_y(a, b) \cdot (y - b) \\ &+ \frac{1}{2}\big(f_{xx}(a, b) \cdot (x - a)^2 + 2f_{xy}(a, b) \cdot (x - a)(y - b) + f_{yy}(a, b)(y - b)^2\big) \end{aligned}$$

Example 3.37 *Find the Taylor polynomial of order 2 for the function* $g(x,y) = x^4 + y^4$ *near the point* $(1,2)$

Solution: We calculate the value of g and its first two derivatives at the point in question, namely: $g(1,2) = 1^4 + 2^4 = 17$, $g_x(1,2) = 4$, $g_y(1,2) = 4 \cdot 8 = 32$,$g_{xx}(1,2) = 12$,$g_{xy}(1,2) = g_{yx}(1,2) = 0$,$g_{yy}(1,2) = 12 \cdot 4 = 48$ and therefore we find $P_2(x,y) = 17 + 4(x-1) + 32(y-2) + 6(x-1)^2 + 24(y-2)^2$, which could also be found by algebraic expansion of $g(x,y) = x^4 + y^4 = (1 + (x-1))^4 + (2 + (y-2))^4$, dropping terms that were degree 3 or higher. Note that there are no cross terms and that is predicted by the lack of cross terms in g.

Here are plots of the function g in red and its Taylor polynomial of order 2 in blue together, first in a somewhat large window and then in a closer view. Note that up close, the linear approximation dominates and any curving is barely visible.

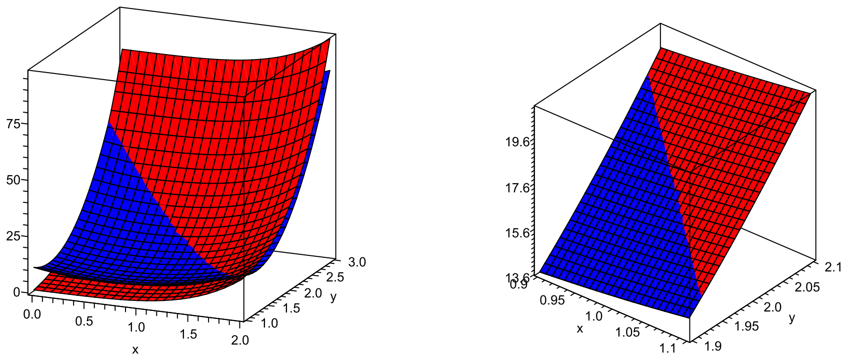

Figure 18: The graphs are of the function $g(x,y) = x^4 + y^4$ (red) and its Taylor polynomial approximation (blue). The second picture is a smaller scale picture near $(1,2)$.

To remove the linear part in visualization, the next pair of plots shows the difference between g and its linear approximation alongside the purely quadratic terms in the Taylor polynomial approximation.

Fans of matrix algebra will recall the matrix discussion of quadratic functions in Chapter 1 and write the pure quadratic terms more compactly in terms of a 2×2 matrix, known as the **Hessian matrix** of the function f. This creates the partial derivatives as entries as follows:

$$H(a,b) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(a,b) & \frac{\partial^2 f}{\partial y \partial x}(a,b) \\ \frac{\partial^2 f}{\partial x \partial y}(a,b) & \frac{\partial^2 f}{\partial y^2}(a,b) \end{pmatrix}.$$

We can then write the purely quadratic term as $\frac{1}{2}(\mathbf{r} - \mathbf{r_0})^T H(a,b)(\mathbf{r} - \mathbf{r_0})$ where $\mathbf{r} =< x, y >$,$\mathbf{r_0} =< a, b >$, and the Hessian matrix is seen to be symmetric when f has continuous second partial derivatives.

There are nine second derivatives if there are three variables, and these form a 3×3 matrix with a similar form for the purely quadratic part.

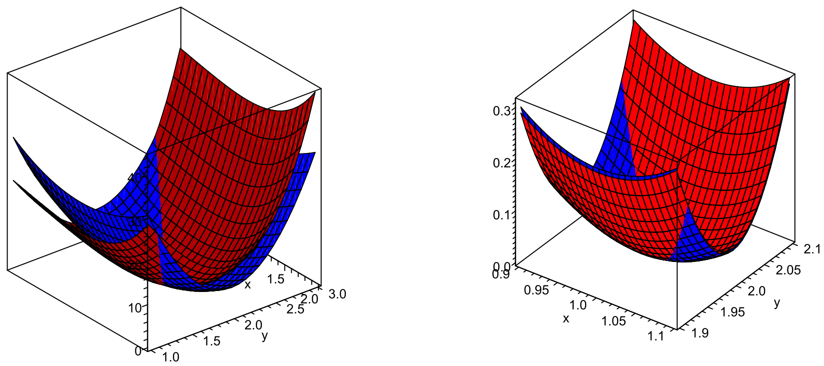

Figure 19: The graphs are of the function $g(x, y) = x^4 + y^4$, less its linear approximation (red) and its purely quadratic term in the Taylor polynomial (blue). The second picture is a smaller scale picture near $(1, 2)$.

As in one variable calculus, the purely quadratic terms tell us about how the graph is bending. In particular, if the quadratic part is positive away from (a, b), the function is **convex** (also known as concave up) and if the quadratic part is negative, the function is **concave down**. We will use this to create a second-derivative test for critical points when we consider max-min problems in the next section.

Reminder: The cross terms like xy or yz are intrinsically indefinite (positive and negative in alternating quadrants!). Quadratic functions may fail to be positive even when the coefficients of the pure squares (the diagonal entries in the Hessian) are positive if the cross terms are too large. Various algebra tests will figure out if that is the situation - completing the squares, finding eigenvalues, or using subdeterminants.

Graphical comment: When there is a non-zero gradient at a point, the zoomed graph will resemble the graph of the linear approximation. Since quadratic terms are more refined (smaller scale closer in), seeing the quadratic behavior graphically requires subtracting off the known linear behavior.

3.5.2 Equality of mixed partial derivatives revisited

The "reverse engineering" of tying coefficients in a general purely quadratic function to the second derivatives of that function forced the equality of the mixed partial derivatives, since they came from terms like $6xy$ or $-2yz$ which clearly have that property. Working locally for a general function, a heuristic argument will show that mixed partial derivatives for twice continuously differentiable functions should be equal. In the next chapter, this will be made rigorous using integration, but for now, the heuristic argument will suffice.

Since continuous functions behave locally like constants, consider the situation where some mixed partial second derivative of a nice function violated Clairaut's theorem. Then say $f_{yz} \approx A$ and $f_{zy} \approx B$ near some point with $a \neq B$. Subtracting these constants would leave terms tending to zero, which we will ignore. Integrating back would create clashing terms of the form Ayz and Byz each claiming to be a good approximation to the same function f near such a point. This is not viable, so we are led to believe heuristically that no such points can exist for twice continuously differentiable functions.

3.5.3 Unit sphere as graph

A revealing example is to consider the upper (or lower) hemisphere of radius 1, namely the function $f(x, y) = \sqrt{1 - x^2 - y^2}$ and find the quadratic approximation at all points. We will compare the "north pole" where $x = y = 0$ with other points and find something surprising.

A direct calculation shows that

$$f_x = -x(1 - x^2 - y^2)^{-\frac{1}{2}}, \ \ f_y = -y(1 - x^2 - y^2)^{-\frac{1}{2}}$$

and therefore using the product rule

$$\begin{aligned}
f_{xx} &= -(1 - x^2 - y^2)^{-\frac{1}{2}} - x^2(1 - x^2 - y^2)^{-\frac{3}{2}} = -\frac{1 - y^2}{(1 - x^2 - y^2)^{\frac{3}{2}}} \\
f_{xy} &= -xy(1 - x^2 - y^2)^{-\frac{3}{2}} \\
f_{yy} &= -(1 - x^2 - y^2)^{-\frac{1}{2}} - y^2(1 - x^2 - y^2)^{-\frac{3}{2}} = -\frac{1 - x^2}{(1 - x^2 - y^2)^{\frac{3}{2}}}
\end{aligned}$$

This leads to the Hessian matrix for f at $(0, 0)$:

$$H(0, 0) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

which is clearly negative definite and the value 1 would seem related to the curvature of the hemisphere.

Now consider a general location (a, b) with $1 - a^2 - b^2 > 0$ so we are in the domain of f. Now the Hessian becomes:

$$H(a, b) = \begin{pmatrix} -\frac{1-b^2}{(1-a^2-b^2)^{\frac{3}{2}}} & -\frac{ab}{(1-a^2-b^2)^{\frac{3}{2}}} \\ -\frac{ab}{(1-a^2-b^2)^{-\frac{3}{2}}} & -\frac{1-a^2}{(1-a^2-b^2)^{\frac{3}{2}}} \end{pmatrix}.$$

This is seen to still be negative definite (the graph lies under the tangent plane!), as some algebra will tell, but the numbers are now more complicated. For example, when moving along the x-axis, with $b = 0$, the Hessian is now diagonal but at $(a, 0)$ the entries are

$$H(a, 0) = \begin{pmatrix} -\frac{1}{(1-a^2)^{\frac{3}{2}}} & 0 \\ 0 & -\frac{1-a^2}{(1-a^2)^{\frac{3}{2}}} \end{pmatrix}.$$

This means that the Hessian by itself will no longer reflect the curvature of the sphere, which ought to be the same at all locations! Instead the curvature of the hemisphere, written as a graph, will use a more complicated blend of first and second derivatives. Geometers compensate by using the tangent plane to reorient everything, making all locations on the surface of the sphere (or any other surface) refer to the tangent plane and its normal when trying to describe curvature of surfaces. They also have multiple notions of curvature. Note that the difficulties our formula has as $a^2 + b^2 \to 1^-$ reflects the difficulties with using x and y as coordinates near the equator of the hemisphere.

3.5.4 Higher order Taylor polynomials

There is no inherent difficulty in approximating more accurately by including cubic terms and higher as long as those derivatives are continuous. The algebraic form gets more cumbersome as the number of derivatives increases. For example, the purely cubic terms near the point $(0,0)$ become

$$\begin{aligned} T_3(x,y) &= \frac{1}{3!}(f_{xxx}(0,0)x^3 + 3f_{xxy}(0,0)x^2y \\ &+ 3f_{xyy}(0,0)xy^2 + f_{yyy}(0,0)y^3), \end{aligned}$$

which involves $2^3 = 8$ terms, while in three variables there are $3^3 = 27$ possible terms, which of course are grouped together using equality of mixed partial derivatives.

Example 3.38 *(Extension of previous problem) Find the Taylor polynomial of order 3 for the function* $g(x,y) = x^4 + y^4$ *near the point* $(1,2)$.

Solution: *Having found the order 2 polynomial, adding the purely cubic terms yields the order 3 polynomial. We continue the previous calculation, finding after dividing by* $3! = 6$*:* $g_{xxx}(1,2) = 24$*,*$g_{xxy}(1,2) = g_{xyy}(1,2) = 0$*,*$g_{yyy}(1,2) = 24 \cdot 2 = 48$ *and therefore we find* $P_3(x,y) = 17 + 4(x-1) + 32(y-2) + 6(x-1)^2 + 24(y-2)^2 + 4(x-1)^3 + 8*(y-2)^3$*, which could also be found by algebraic expansion of* $g(x,y) = x^4 + y^4 = (1 + (x-1))^4 + (2 + (y-2))^4$*, dropping terms that were degree 4.*

The terms involving total degree k will number n^k in n variables, with lots of subgrouping for mixed partial derivatives. These involve summing all the k-fold products and the corresponding multidimensional array with k dimensions is called a symmetric k-tensor.

An alternative form uses the chain rule and Taylor's theorem in one variable as follows: If the line segment from (a,b) to (x,y) including the endpoints lies inside an open set S where the function f is n times continuously differentiable, then the function g defined by writing $f(x,y) - f(a,b) = f(a+t(x-a), b+t(y-b))|_{t=0}^{t=1} = g(1) - g(0)$ where $g(t) = f(a+t(x-a), b+t(y-b))$ is also n times continuously differentiable. Then using Taylor's theorem for g, we get an approximation for f in powers of $x-a$ and $y-b$ and if f is $n+1$ times continuously differentiable, we get a remainder formula too! This is pursued slightly in the exercises below and the interested student can find more information in the references.

3.5.5 EXERCISES

Writing: Explaining, Reacting, Questioning

1. If a purely quadratic function $Q(x, y) = Ax^2 + Bxy + Cy^2$ has maximum value 2 and minimum value 1 on the unit circle: $x^2 + y^2 = 1$, what are its maximum and minimum values on the circle of radius 4? Explain your answer.

2. Cal Clueless believes that a term like x^2, with a larger exponent, is always larger than x. Write a persuasive few sentences that remind him how higher powers like x^2y or xy^4 or xyz behave near $(0, 0)$ or $(0, 0, 0)$ compared to terms like x or y.

3. Ivana Calculator does not like matrices. She worked out the quadratic polynomial approximation to $g(x, y) = g(x, y) = 4x^3y + 6x^2\,y^2$ near the point $(1, 2)$ by multiplying out, substituting $x = 1+(x-1)$ and $y = 2+(y-2)$ in each term. Redo her calculation by first observing that products of functions of x and y can be done individually as polynomial approximations in one variable and then multiplying; and second by at least using Taylor polynomials in each of the one variable terms, such as x^3 or y^2.

4. Find the gradient of a general purely quadratic function of two variables: $f(x, y) = Ax^2 + Bxy + Cy^2$. Describe this vector in terms of the Hessian matrix of f. Now extend your result to 3-D. Formulate your result in vector-matrix language.

5. Suppose a purely quadratic function of x, y, $Q(x, y) = Ax^2 + Bxy + Cy^2$ takes on positive and negative values for $x^2 + y^2 = 1$. What can you say about the set where $Q(x, y) = 0$? Why? Make sure you take advantage of the fact that $Q(-x, -y) = Q(x, y)$ as part of your answer.

6. (Continuing previous problem): If the Hessian matrix has diagonal entries of both signs (positive and negative), what can you conclude about the local behavior of the function? If the quadratic function takes on positive and negative values

7. In one variable calculus, inflection points are defined as locations where f'' changes sign. What might be a good definition for functions of two variables, remembering that functions like $x^2 - y^2$ can be neither concave nor convex? Should point become curve? Why or why not? Create an interesting function that has an inflection point according to your definition and describe what is happening nearby.

8. Using the notion of Taylor polynomial of order 2, find the Taylor polynomial expansion of $(x + y + z)^2$ in powers of x, y and z. Then generalize your result to $(x + y + z)^3$ and finally to $(x + y + z)^n$ for any positive integer n. For this last part you might want to use trinomial coefficient notation to

represent the number of ways to split a number n into three parts, usually written $\binom{n}{j;k;l} = \frac{n!}{j!k!l!}$ for splitting into parts with $j+k+l=n$.

9. Consider a twice continuously differentiable function $f(x,y)$ defined on a rectangular domain $[-2,2]\times[-1,1]$ with $(0,0)$ at the center. Using a Taylor polynomial approximation to f centered at $(0,0)$, what are the predicted values of f at the four corners? **Challenge:** Using the predicted values at the corners as data, which values of the derivatives of f at $(0,0)$ can you determine? Explain your answer.

10. Redo the example in the text of finding the quadratic approximation to the polynomial function $g(x,y) = x^4 + y^4$ near the point $(1,2)$ by using the chain rule on $G(t) = g(1+t(x-1), 2+t(y-2))$ which runs from values $g(1,2)$ to $g(x,y)$ as t runs from 0 to 1. **Challenge:** What does the remainder form look like, using your one variable calculus skills?

11. Without running numbers, how would you expect to find the linear approximation near a point in terms of values? If you did second difference tables in your one-variable class, generalize them to functions of two or three variables; if you never did such tables, think about the one variable case and look up the idea, which says that differences of squares grow linearly (like the derivative) and differences of differences therefore have a more or less constant value emerging (constant for purely quadratic functions). That works fine for the diagonal parts of the Hessian. How will you deal with cross terms? Write a brief paragraph that describes the numerical data and its handling.

Calculational Exercises For problems 1-12, find the Hessian matrix and the quadratic approximation to the function at the indicated point:

1. $f(x,y) = x^2 + 6xy + 4y^2$ near $(0,0)$.
2. $g(x,y) = \frac{1}{xy}$ near $(2,4)$.
3.
4.
5.
6.
7.
8.
9.
10.