Quadratic Approximations

We are already familiar with linear approximations of multivariable functions. These essentially amount to equations of tangent planes. We also know these linear approximations by the name "differentials". The next natural step is to consider higher order approximations. We will study what these look like and in particular focus on second order approximations – that is – quadratic approximations.

Single Variable:

First, let's quickly review single variable theory. A linear approximation of a function $f(x)$ at $x = a$ is given by $L(x) = f(a) + f'(a) \cdot (x - a)$. This linear approximation is the line which best approximates $f(x)$ at $x = a$ in the sense that $f(x)$ and $L(x)$ both pass through the point $(a, f(a))$ and both have the same slope at $x = a$: $f'(a) = L'(a)$. Another way to state this is: $L(x)$ is the first order Taylor polynomial of $f(x)$ at $x = a$. Recall from calculus II (or I) that the second order Taylor polynomial for $f(x)$ at $x = a$ is...

$$Q(x) = f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2} \cdot (x - a)^2$$

This is the best quadratic approximation of $f(x)$ in the sense that $Q(x)$ is a quadratic polynomial, $f(a) = Q(a)$, $f'(a) = Q'(a)$, and $f''(a) = Q''(a)$. Finally, recall that the n^{th}-order Taylor polynomial of $f(x)$ at $x = a$ is given by...

$$P_n(x) = f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2} \cdot (x - a)^2 + \frac{f'''(a)}{3!} \cdot (x - a)^3 + \cdots + \frac{f^{(n)}(a)}{n!} \cdot (x - a)^n$$

This is the best n^{th}-order polynomial approximation of $f(x)$ in sense that $P_n(x)$ is an n^{th}-order polynomial, $f(a) = P_n(a)$, $f'(a) = P_n'(a)$, ..., and $f^{(n)}(a) = P_n^{(n)}(a)$.

Multivariable Linear Approximations:

Our approximations and Taylor polynomials for multivariable functions will be *best* approximations in the same way our single variable approximations were best – our approximations will have a matching function value and matching partial derivatives at the base point. For a function of two variables $f(x, y)$ at the point $(x, y) = (a, b)$ this looks like...

$$L(x, y) = f(a, b) + f_x(a, b) \cdot (x - a) + f_y(a, b) \cdot (y - b)$$

In three variables we have...

$$L(x, y, z) = f(a, b, c) + f_x(a, b, c) \cdot (x - a) + f_y(a, b, c) \cdot (y - b) + f_z(a, b, c) \cdot (z - c)$$

And finally in n variables the linearization of $f(x_1, \ldots, x_n)$ at $(x_1, \ldots, x_n) = (a_1, \ldots, a_n)$ is...

$$L(x_1, \ldots, x_n) = f(a_1, \ldots, a_n) + f_{x_1}(a_1, \ldots, a_n) \cdot (x_1 - a_1) + f_{x_2}(a_1, \ldots, a_n) \cdot (x_2 - a_2) + \cdots + f_{x_n}(a_1, \ldots, a_n) \cdot (x_n - a_n)$$

By introducing the *gradient* we can state this much more concisely. The gradient of a function $f(x_1, \ldots, x_n) = f(\vec{x})$ is a vector whose entries consist of the first partials of f: $\nabla f(\vec{x}) = \langle f_{x_1}(\vec{x}), f_{x_2}(\vec{x}), \ldots, f_{x_n}(\vec{x}) \rangle$ In particular, for $f(x, y, z)$ we have $\nabla f = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle$. Using this notation the linearization of $f(\vec{x})$ at $\vec{x} = \vec{a}$ is...

$$L(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \bullet (\vec{x} - \vec{a})$$

where the dot between $\nabla f(\vec{a})$ and $\vec{x} - \vec{a}$ is the familiar dot product.

These linearizations are first order Taylor polynomials for multivariable functions. Next, we consider quadratic approximations (second order Taylor polynomials).

Example: Let's find the linearization of $f(x, y) = x^2y + x^3y^2 + y^3$ at $(x, y) = (-1, 1)$.

We need to compute the first partials of f. $f_x(x, y) = 2xy + 3x^2y^2$ and $f_y(x, y) = x^2 + 2x^3y + 3y^2$. Next, we need to compute the value of f and its partials at the point $(-1, 1)$ this yields $f(-1, 1) = (-1)^2(1) + (-1)^3(1^2) + 1^3 = 1$, $f_x(-1, 1) = 2(-1)(1) + 3(-1)^2(1^2) = 1$, and $f_y(-1, 1) = (-1)^2 + 2(-1)^3(1) + 3(1^2) = 2$. Thus the linearization of $f(x, y)$ at $(-1, 1)$ is $L(x, y) = 1 + 1 \cdot (x - (-1)) + 2 \cdot (y - 1)$.

Answer: $L(x, y) = 1 + (x + 1) + 2(y - 1)$ which is $L(x, y) = x + 2y$.

This means that the plane tangent to the surface $z = x^2y + x^3y^2 + y^3$ at $(x, y, z) = (-1, 1, 1)$ is $z = x + 2y$.

Example: Let's find the linearization of $f(x,y,z) = x\sin(y) + e^{yz^2} + xyz$ at $(x,y,z) = (3,0,-2)$.

We need to compute the first partials of f. $f_x(x,y,z) = \sin(y) + yz$, $f_y(x,y,z) = x\cos(y) + z^2e^{yz^2} + xz$, and $f_z(x,y,z) = 2zye^{yz^2} + xy$. Next, we need to compute the value of f and its partials at the point $(3,0,-2)$ this yields $f(3,0,-2) = 3\sin(0) + e^{0(-2)^2} + 3(0)(-2) = e^0 = 1$, $f_x(3,0,-2) = \sin(0) + 0(-2) = 0$, $f_y(3,0,-2) = 3\cos(0) + (-2)^2e^{0(-2)^2} + 3(-2) = 3 + 4e^0 - 6 = 1$, and $f_z(3,0,-2) = 2(-2)(0)e^{0(-2)^2} + 3(0) = 0$. Thus the linearization of $f(x,y)$ at $(3,0,-2)$ is $L(x,y,z) = 1 + 0\cdot(x-3) + 1\cdot(y-0) + 0\cdot(z-(-2))$.

Answer: $L(x,y,z) = 1 + y$.

Notice that $\nabla f = \langle f_x, f_y, f_z\rangle = \langle \sin(y) + yz, x\cos(y) + z^2e^{yz^2} + xz, 2zye^{yz^2} + xy\rangle$ so that $\nabla f(3,0,-2) = \langle 0,1,0\rangle$. Thus another way to write the linearization is $L(x,y,z) = 1 + \langle 0,1,0\rangle \bullet \langle x-3, y-0, z-(-2)\rangle$.

Multivariable Quadratic Approximations:

A second order Taylor polynomial should match our function's value, its first partials, and its second partials. For a function of two variables $f(x,y)$ at the point $(x,y) = (a,b)$ this looks like…

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

Under the (mild) assumption of continuous second paritals, Clairaut's theorem applies and mixed partials are equal. In this case we have…

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

In three variables this looks like (after suppressing the point of evaluation)…

$$\begin{aligned} Q(x,y,z) \;=\;& f + f_x\cdot(x-a) + f_y\cdot(y-b) + f_z\cdot(z-c) + \frac{1}{2}f_{xx}\cdot(x-a)^2 + \frac{1}{2}f_{yy}\cdot(y-b)^2 + \frac{1}{2}f_{zz}\cdot(z-c)^2 + \\ & \frac{1}{2}f_{xy}\cdot(x-a)(y-b) + \frac{1}{2}f_{yx}\cdot(x-a)(y-b) + \frac{1}{2}f_{xz}\cdot(x-a)(z-c) + \frac{1}{2}f_{zx}\cdot(x-a)(z-c) + \\ & \frac{1}{2}f_{yz}\cdot(y-b)(z-c) + \frac{1}{2}f_{zy}\cdot(y-b)(z-c) \end{aligned}$$

The formula for a quadratic approximation of a 3 variable function is bad enough. The n variable expression will be an even bigger mess unless we resort to matrix notation. If you are unfamiliar with matrix multiplication, please read on, but don't worry about the details.

Consider a function of n variables: $f(\vec{x})$. Let's define an $n\times n$ matrix whose (i,j)-entry is $f_{x_ix_j}$. This matrix is called the *Hessian* matrix of f. Here are the Hessians for $h(x,y)$, $g(x,y,z)$ and $f(\vec{x}) = f(x_1,\dots,x_n)$:

$$H_h = \begin{bmatrix} h_{xx} & h_{xy} \\ h_{yx} & h_{yy} \end{bmatrix} \qquad H_g = \begin{bmatrix} g_{xx} & g_{xy} & g_{xz} \\ g_{yx} & g_{yy} & g_{yz} \\ g_{zx} & g_{zy} & g_{zz} \end{bmatrix} \qquad H_f = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & \cdots & f_{x_nx_n} \end{bmatrix}$$

If $\vec{x} = \langle x_1, x_2, \dots, x_n\rangle$, $\vec{x}^T$ is the *transpose* of $\vec{x}$ (transpose basically means "turn rows into columns"), and H_f is the Hessian matrix for f, then we can restate the quadratic approximation of $f(\vec{x})$ at $\vec{x} = \vec{a}$ as follows:

$$Q(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \bullet (\vec{x}-\vec{a}) + \frac{1}{2}(\vec{x}-\vec{a})\, H_f\, (\vec{x}-\vec{a})^T$$

where the last term is computed using matrix multiplication.

One quick note about the Hessian. Remember Clairaut's theorem says that if our second partials are continuous, then the mixed partials $f_{x_ix_j}$ and $f_{x_jx_i}$ are equal. This means that in the first matrix $h_{xy} = h_{yx}$, in the second matrix $g_{xy} = g_{yx}$, $g_{yz} = g_{zy}$, and $g_{xz} = g_{zx}$, and in the last matrix $f_{x_ix_j} = f_{x_jx_i}$ for all i's and j's. Matrices with this property are called *symmetric*. Another way to put this is: If f has continuous second partials, then $H_f = H_f^T$ (remember transpose means turn row i in column i) so in other words the i^{th} row of H_f has the same entries as the i^{th} column of H_f.

Beyond giving us a nice way to write quadratic approximations of multivariable functions, the Hessian matrix helps when we need to determine extremal behavior. Recall that an extreme point (a minimum or maximum) for a function of one variable must occur at a point where the derivative is either zero or does not exist. Functions of more than one variable work the same way. Extreme points can only occur at points where (all of) the first partials are zero or do not exist.

In Calculus I we learned the "second derivative test" which told us that a critial point with negative second derivative (concave down) is a maximum and a critical point with a positive second derivative (concave up) is a minimum. We will soon learn a "second dervative test" for functions of two variables, which relies on the following general result: If all of the eigenvalues of the Hessian matrix are negative at a critical point, then that point is a maximum. Also, if all of the eigenvalues of the Hessian matrix are positive at a critical point, then that point is a minimum. And if the Hessian matrix has a mix of positive and negative values, we have a *saddle point* (we will talk about these later).

If you don't know what an eigenvalue is, that's ok. They are special values associated with a square matrix. Our second derivative test (for functions of two variables) will determine what type of eigenvalues we have (positive or negative) without actually computing them. However, to have a second derivative test for functions of more than two variables we would have to learn how to compute eigenvalues (which we won't do in this class).

Example: Let's find the quadratic approximation of $f(x,y) = x^2y + x^2y^3 + 2x$ at $(x,y) = (2,-1)$.

We need to compute the first and second partials and then plug in $(x,y) = (2,-1)$. Maybe we should make a table to keep track of all of this stuff.

$f(x,y) = x^2y + x^2y^3 + 2x$		$f(2,-1) = -4$	
$f_x(x,y) = 2xy + 2xy^3 + 2$	$f_x(2,-1) = -6$	$f_y(x,y) = x^2 + 3x^2y^2$	$f_y(2,-1) = 16$
$f_{xx}(x,y) = 2y + 2y^3$	$f_{xx}(2,-1) = -4$	$f_{xy}(x,y) = 2x + 6xy^2$	$f_{xy}(2,-1) = 16$
$f_{yx}(x,y) = 2x + 6xy^2$	$f_{yx}(2,-1) = 16$	$f_{yy}(x,y) = 6x^2y$	$f_{yy}(2,-1) = -24$

Putting this together we get:

$$\begin{aligned} Q(x,y) &= -4 + (-6)(x-2) + 16(y-(-1)) + \frac{1}{2}(-4)(x-2)^2 + 16(x-2)(y-(-1)) + \frac{1}{2}(-24)(y-(-1))^2 \\ &= -4 - 6(x-2) + 16(y+1) - 2(x-2)^2 + 16(x-2)(y+1) - 12(y+1)^2 \end{aligned}$$

In our slick vector/matrix notation this looks like...

$$Q(x,y) = -4 + \langle -6, 16 \rangle \bullet \langle x-2, y+1 \rangle + \frac{1}{2} \begin{bmatrix} x-2 & y+1 \end{bmatrix} \begin{bmatrix} -4 & 16 \\ 16 & -24 \end{bmatrix} \begin{bmatrix} x-2 \\ y+1 \end{bmatrix}$$

Notice that since f has continuous second partials, its Hessian is symmetric: $H_f = \begin{bmatrix} 2y+2y^3 & 2x+6xy^2 \\ 2x+6xy^2 & 6x^2y \end{bmatrix}$

Example: Let's find the quadratic approximation of $f(x,y,z) = xyz + ye^{\sin(z)} - 3z^2$ at $(x,y,z) = (2,1,0)$.

We need to compute the first and second partials and then plug in $(x,y,z) = (2,1,0)$. Let's make a table.

$f(x,y,z) = xyz + ye^z - 3z^2$			$f(2,1,0) = 1$		
$f_x = yz$	$f_x(2,1,0) = 0$	$f_y = xz + e^z$	$f_y(2,1,0) = 1$	$f_z = xy + ye^z - 6z$	$f_y(2,1,0) = 3$
$f_{xx} = 0$	$f_{xx}(2,1,0) = 0$	$f_{yy} = 0$	$f_{yy}(2,1,0) = 0$	$f_{zz} = ye^z - 6$	$f_{zz}(2,1,0) = -5$
$f_{xy} = f_{yx} = z$	$f_{xy}(2,1,0) = 0$	$f_{yz} = f_{zy} = x + e^z$	$f_{yz}(2,1,0) = 3$	$f_{xz} = f_{zx} = y$	$f_{xz}(2,1,0) = 1$

Putting this together we get:

$$\begin{aligned} Q(x,y) &= 1 + 0(x-2) + 1(y-1) + 3(z-0) + \frac{1}{2}(0)(x-2)^2 + \frac{1}{2}(0)(y-1)^2 + \frac{1}{2}(-5)(z-0)^2 + \\ &\quad 0(x-2)(y-1) + 3(y-1)(z-0) + 1(x-2)(z-0) \\ &= 1 + (y-1) + 3z - \frac{5}{2}z^2 + 3(y-1)z + (x-2)z \end{aligned}$$

In our slick vector/matrix notation this looks like...

$$Q(x,y) = 1 + \langle 0,1,3 \rangle \bullet \langle x-2, y-1, z \rangle + \frac{1}{2} \begin{bmatrix} x-2 & y-1 & z \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 3 \\ 1 & 3 & -5 \end{bmatrix} \begin{bmatrix} x-2 \\ y-1 \\ z \end{bmatrix}$$

Again, since f has continuous second partials, its Hessian is symmetric: $H_f = \begin{bmatrix} 0 & z & y \\ z & 0 & x+e^z \\ y & x+e^z & ye^z-6 \end{bmatrix}$

A Bit About Taylor Polynomials:

The formulas for multivariate Taylor polynomials are quite complicated – or so they seem. If we approach Taylor polynomials from a operator viewpoint, the multivariate formulas are easy to explain.

Recall that the MacLaurin series for e^x is $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{k!}$. Some textbooks use this series to *define* what we mean by e^x! It is quite surprising how many places the MacLaurin series for e^x shows up. For example, if one defines $e^A = \displaystyle\sum_{k=0}^{\infty} \frac{A^k}{k!}$ for a square matrix A, you get the *matrix exponential* which is computed when one wants to solve linear systems of differential equations.

We define a formal series of derivative operators: $e^{(x-a)\frac{d}{dx}\big|_{x=a}} = \displaystyle\sum_{k=0}^{\infty} \frac{(x-a)^k}{k!} \left.\frac{d^k}{dx^k}\right|_{x=a}$. If we apply this to some function $f(x)$, we get... the Taylor series of $f(x)$ based at $x=a$:

$$e^{(x-a)\frac{d}{dx}\big|_{x=a}} f(x) = \sum_{k=0}^{\infty} \frac{(x-a)^k}{k!} \left.\frac{d^k f}{dx^k}\right|_{x=a} = f(a) + f'(a)\cdot(x-a) + \frac{1}{2}f''(a)\cdot(x-a)^2 + \frac{1}{3!}f'''(a)\cdot(x-a)^3 + \cdots$$

For a function of 2 variables, $f(x,y)$, we use: [*Note:* I will drop the evaluation bars. All partials are evaluated at $(x,y)=(a,b)$.]

$$e^{(x-a)\frac{\partial}{\partial x}+(y-b)\frac{\partial}{\partial y}} f(x,y) = f(a,b) + \left((x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right)\Big[f(x,y)\Big] + \frac{1}{2}\left((x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right)^2 \Big[f(x,y)\Big] + \cdots$$

Let's examine how to expand the quadratic term.

$$\begin{aligned}
&\left((x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right)^2 \Big[f(x,y)\Big] \\
&= \left((x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right)\left((x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right)\Big[f(x,y)\Big] \\
&= (x-a)^2\frac{\partial^2}{\partial x^2}\Big[f(x,y)\Big] + (x-a)(y-b)\frac{\partial}{\partial x}\frac{\partial}{\partial y}\Big[f(x,y)\Big] + (x-a)(y-b)\frac{\partial}{\partial y}\frac{\partial}{\partial x}\Big[f(x,y)\Big] + (y-b)^2\frac{\partial^2}{\partial y^2}\Big[f(x,y)\Big] \\
&= (x-a)^2 f_{xx}(a,b) + (x-a)(y-b)f_{yx}(a,b) + (x-a)(y-b)f_{xy}(a,b) + (y-b)^2 f_{yy}(a,b)
\end{aligned}$$

If we assume that the second partials are continuous, this becomes...

$$= (x-a)^2 f_{xx}(a,b) + 2(x-a)(y-b)f_{xy}(a,b) + (y-b)^2 f_{yy}(a,b)$$

Assuming continuity of third order partials the next few terms of the Taylor series are

$$\frac{1}{3!}\left((x-a)^3 f_{xxx}(a,b) + 3(x-a)^2(y-b)f_{xxy}(a,b) + 3(x-a)(y-b)^2 f_{xyy}(a,b) + (y-b)^3 f_{yyy}(a,b)\right)$$

Because we are expanding a binomial of operators, we see binomial coefficients starting to appear.

If we let $\nabla = \left\langle \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n} \right\rangle$ then $(\vec{x}-\vec{a}) \bullet \nabla = (x_1-a_1)\frac{\partial}{\partial x_1} + (x_2-a_2)\frac{\partial}{\partial x_2} + \cdots + (x_n-a_n)\frac{\partial}{\partial x_n}$.

The Taylor series for $f(\vec{x})$ centered at $\vec{x}=\vec{a}$ is $e^{(\vec{x}-\vec{a})\bullet\nabla}[f(\vec{x})]$.

Exercises:

[See the accompanying Maple worksheet for answers.]

For problems 1 – 11, find the linearization and quadratic approximation of f at the given point. Then write the Hessian matrix of the function.

1. $f(x,y) = e^{x+2y}$ with $(a,b) = (0,0)$
2. $f(x,y) = x^2 + xy^2$ with $(a,b) = (2,3)$
3. $f(x,y) = x^4y^2 + xy - 2y$ with $(a,b) = (-1,0)$
4. $f(x,y) = 2x^2 + 3y^4 + e^{xy^2}$ with $(a,b) = (2,0)$
5. $f(x,y) = \sin(xy)$ with $(a,b) = (3,\pi)$
6. $f(x,y) = \ln(x^2+y)$ with $(a,b) = (1,0)$
7. $f(x,y) = \cos(xy)$ with $(a,b) = (\sqrt{\pi},\sqrt{\pi})$
8. $f(x,y,z) = xyz$ with $(a,b,c) = (1,2,3)$
9. $f(x,y,z) = x^4 + y^4 + z^4$ with $(a,b,c) = (-1,0,1)$
10. $f(x,y,z) = x^2z + xy^2 + y^3z^2 + e^{xy+z}$ with $(a,b,c) = (2,1,-2)$
11. $f(w,x,y,z) = w + x^2 + y^3 + z^4 + w^2\cos(xy^2)$ with $(a,b,c,d) = (0,0,0,0)$
12. Let $f(x,y) = Ax + By + C$ with linear and quadratic approximations (based at $(x,y)=(a,b)$) called $L(x,y)$ and $Q(x,y)$. Show $f(x,y) = L(x,y) = Q(x,y)$.
    What should I expect if $f(x,y) = Ax^2 + By^2 + Cxy + Dx + Ey + F$?

Assume continuity of all partial derivatives (so Clairaut's theorem applies).

13. Write down the 3rd-order terms for the Taylor polynomial expansion of $f(x,y,z)$ at $(x,y,z)=(a,b,c)$.
14. Write down the 4th-order terms for the Taylor polynomial expansion of $f(x,y)$ at $(x,y)=(a,b)$.