Math 51
Lecture 25 – August 11, 2023

Topic(s): Hessian of a function, quadratic approximation of a function

The role of the second derivative at a point, for multivariable functions $f : \mathbb{R}^n \to \mathbb{R}$, is played by an $n \times n$ matrix called the **Hessian**. For a function $f : \mathbb{R}^n \to \mathbb{R}$, the **Hessian matrix** $(Hf)(\mathbf{a})$ of f at a point $\mathbf{a} \in \mathbb{R}^n$ is defined to be the matrix of second partial derivatives, with ij-entry equal to $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})$:

$$(Hf)(\mathbf{a}) = \begin{bmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) & f_{x_1x_3}(\mathbf{a}) & \cdots & f_{x_1x_n}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) & f_{x_2x_3}(\mathbf{a}) & \cdots & f_{x_2x_n}(\mathbf{a}) \\ f_{x_3x_1}(\mathbf{a}) & f_{x_3x_2}(\mathbf{a}) & f_{x_3x_3}(\mathbf{a}) & \cdots & f_{x_3x_n}(\mathbf{a}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1}(\mathbf{a}) & f_{x_nx_2}(\mathbf{a}) & f_{x_nx_3}(\mathbf{a}) & \cdots & f_{x_nx_n}(\mathbf{a}) \end{bmatrix}$$

The Hessian matrix $(Hf)(\mathbf{a})$ of second partial derivatives is exactly $(D(\nabla f))(\mathbf{a})$ and is *symmetric* due to equality of mixed partials.

Example 1. Compute the Hessian of $F(x, y, z) = x^2yz + \cos(xy) + ze^y$.

We compute $\nabla f = (2xyz - y\sin(xy),\ x^2z - x\sin(xy) + ze^y,\ x^2y + e^y)$ and

$$Hf = \begin{bmatrix} 2y - y^2\cos(xy) & 2xz - \sin(xy) - xy\cos(xy) & 2xy \\ 2xz - \sin(xy) - xy\cos(xy) & -x^2\cos(xy) + ze^y & x^2 + e^y \\ 2xy & x^2 + e^y & 0 \end{bmatrix}.$$

The **quadratic approximation** to $f : \mathbb{R}^n \to \mathbb{R}$ near $\mathbf{a} \in \mathbb{R}^n$ is the expression on the right-hand side of

$$f(\mathbf{a} + \mathbf{h}) \approx \overbrace{f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{h}}^{\text{linear approx.}} + \frac{1}{2}\mathbf{h}^{\intercal}((Hf)(\mathbf{a}))\mathbf{h},$$

analogous to $f(a) + f'(a)h + \frac{f''(a)}{2}h^2$

Second degree Taylor!

for small $\mathbf{h}$. This refines "linear approximation."

Example 2. Let $f(x, y) = \ln(xy - 1)$. Approximate the value of $f(1.2, 1.8)$ via quadratic approximation near $(1, 2)$.

We compute $\nabla f = \begin{bmatrix} \frac{y}{xy-1} \\ \frac{x}{xy-1} \end{bmatrix}$ and $Hf = \begin{bmatrix} -\frac{y^2}{(xy-1)^2} & -\frac{1}{(xy-1)^2} \\ -\frac{1}{(xy-1)^2} & -\frac{x^2}{(xy-1)^2} \end{bmatrix}$. Thus,

$$f(1.2, 1.8) \approx f(1,2) + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0.2 \\ -0.2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0.2 & -0.2 \end{bmatrix}\begin{bmatrix} -4 & -1 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} 0.2 \\ -0.2 \end{bmatrix}$$

$$= 0 + 0.2 + \frac{1}{2}\begin{bmatrix} 0.2 & -0.2 \end{bmatrix}\begin{bmatrix} -0.6 \\ 0 \end{bmatrix} = 0.2 - 0.06 = 0.14.$$

Note $f(1.2, 1.8) = \ln(2.16 - 1) = 0.148420005.$

Example 3. Consider $f(x, y) = \cos\left(\frac{x}{y}\right)$. Use quadratic approximation to estimate $\cos\left(\frac{h_1}{1+h_2}\right)$ for small values of h_1 and h_2.

$$\cos\left(\frac{h_1}{1+h_2}\right) = f(h_1, 1+h_2)$$

We compute $\nabla f = \begin{bmatrix} -\frac{1}{y}\sin\left(\frac{x}{y}\right) \\ \frac{x}{y^2}\sin\left(\frac{x}{y}\right) \end{bmatrix}$ and $Hf = \begin{bmatrix} -\frac{1}{y^2}\cos\left(\frac{x}{y}\right) & \frac{1}{y^2}\sin\left(\frac{x}{y}\right) + \frac{x}{y^3}\cos\left(\frac{x}{y}\right) \\ \frac{1}{y^2}\sin\left(\frac{x}{y}\right) + \frac{x}{y^3}\cos\left(\frac{x}{y}\right) & -\frac{2x}{y^3}\sin\left(\frac{x}{y}\right) - \frac{x^2}{y^4}\cos\left(\frac{x}{y}\right) \end{bmatrix}$. So,

$$f(h_1, 1+h_2) \approx f(0,1) + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} h_1 & h_2 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

$$= 1 - \frac{h_1^2}{2} \longleftarrow \text{deg. 2 Taylor poly. for } \cos(h_1)!$$

The quadratic approximation can also be used to understand the behavior of a function $f : \mathbb{R}^2 \to \mathbb{R}$ near a critical point in terms of the contour plot. The basic idea is: suppose $\mathbf{a} \in \mathbb{R}^2$ is a *critical point* of $f : \mathbb{R}^2 \to \mathbb{R}$; i.e. $\nabla f(\mathbf{a}) = \mathbf{0}$. Hence, the gradient term in the quadratic approximation vanishes, and we have, for $\mathbf{h}$ near $\mathbf{0}$,

$$f(\mathbf{a} + \mathbf{h}) \approx f(\mathbf{a}) + \frac{1}{2} q_{Hf(\mathbf{a})}(\mathbf{h}).$$

Hence, for a scalar c near $f(\mathbf{a})$, the level curve $f(\mathbf{a} + \mathbf{h}) = c$ for $\mathbf{h}$ near $\mathbf{0}$ is well-approximated by the level curve $f(\mathbf{a}) + \frac{1}{2} q_{Hf(\mathbf{a})}(\mathbf{h}) = c$, or equivalently,

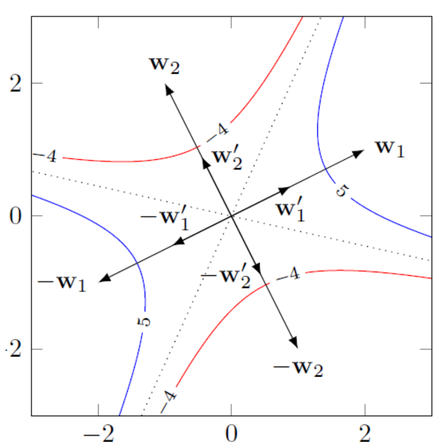
$$q_{Hf(\mathbf{a})}(\mathbf{h}) = 2(c - f(\mathbf{a})).$$

In other words, the level curves of f near a critical point $\mathbf{a} \in \mathbb{R}^2$ are well-approximated by the level curves of the quadratic form $q_{Hf(\mathbf{a})}$ near the origin.

Example 4. Consider $A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ and the associated quadratic form $q_A(x, y) = x^2 + 4xy - 2y^2$. What do the level curves of q_A look like around the origin?

A has eigenvectors $\vec{w}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{w}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ with eigenvalues 2 and -3. So, we

$\vec{w}_1' = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{w}_2' = \frac{1}{\sqrt{5}}\begin{bmatrix} -1 \\ 2 \end{bmatrix}$. If $\vec{v} = t_1\vec{w}_1' + t_2\vec{w}_2'$,



$$q_A(\vec{v}) = 2t_1^2 - 3t_2^2$$

by the diagonalization formula.

The hyperbola open towards t_1-axis if $c > 0$ and towards t_2-axis if $c < 0$.

Thus, the origin is a saddle point.

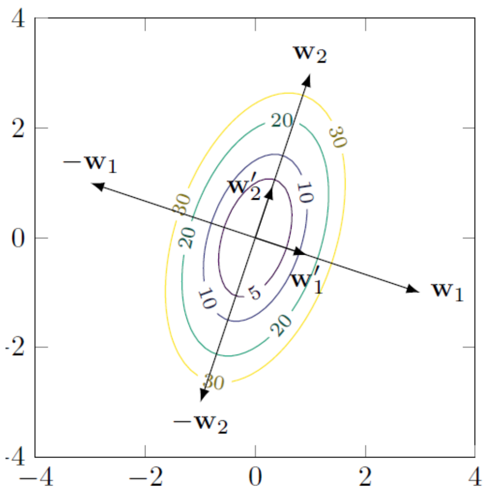
Example 5. Consider $A = \begin{bmatrix} 13 & -3 \\ -3 & 5 \end{bmatrix}$ and the associated quadratic form $q_A(x, y) = 13x^2 - 6xy + 5y^2$. What do the level curves of q_A look like around the origin?

A has eigenvectors $\vec{w}_1 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ and $\vec{w}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ with eigenvalues 14 and 4, respectively. The unit eigenvectors are $\vec{w}_1' = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ and $\vec{w}_2' = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. So, for $\vec{v} = t_1 \vec{w}_1' + t_2 \vec{w}_2'$, we get

$$q_A(\vec{v}) = 14t_1^2 + 4t_2^2$$

by the diagonalization formula.

The level curves exist only for $c > 0$, and they are ellipses longer in the t_2-axis. Hence, the origin is a local minimum.



If $f(x, y)$ has a critical point $\mathbf{a}$ at which the Hessian $(Hf)(\mathbf{a})$ has eigenvalues $\lambda_1, \lambda_2 \neq 0$, then in a **typically tilted reference frame** given by its orthogonal eigenlines, the *quadratic approximation* to $f(\mathbf{x})$ near $\mathbf{a}$ has level sets

$$\lambda_1 t_1'^2 + \lambda_2 t_2'^2 = 2(c - f(\mathbf{a})).$$

These are good approximation to the level sets $f(\mathbf{x}) = c$ for $\mathbf{x}$ near $\mathbf{a}$. Thus,

1. these level sets for f near $\mathbf{a}$ look like hyperbolas when λ_1 and λ_2 have opposite signs (indefinite Hessian). Asymptotes are closer to the eigenline corresponding to the smaller magnitude eigenvalue.

2. these level sets for f near $\mathbf{a}$ look like ellipses when λ_1 and λ_2 have the same sign (definite Hessian). The longer direction of the ellipse is along the eigenline whose eigenvalue has smaller absolute value.

Eventually, in Chapter 26, we will see that, for $\lambda_1, \lambda_2 \neq 0$,

- if λ_1 and λ_2 are of opposite signs, then f has a saddle point at $\mathbf{a}$.
- if λ_1 and λ_2 are both positive, then f has a local minimum at $\mathbf{a}$.
- if λ_1 and λ_2 are both negative, then f has a local maximum at $\mathbf{a}$.

Example 6. This exercise provides practice with a technique that will relate contour plots to the multivariable second derivative test in Chapter 26.

For each quadratic form $q(x, y)$ below, compute the eigenvalues λ_1, λ_2 of the associated symmetric 2×2 matrix and find an orthogonal basis $\{\mathbf{w}_1, \mathbf{w}_2\}$ of corresponding eigenvectors. (The eigenvalues are integers in both cases.)
Also sketch qualitatively correct level sets, including justification in terms of the eigenvalues: in a definite case, draw ellipses aligned with with eigenlines and longer along the correct eigenline, and in an indefinite case draw hyperbolas $q(x, y) = \pm c$ aligned with the eigenlines and with asymptotes drawn "closer" to the correct eigenline.

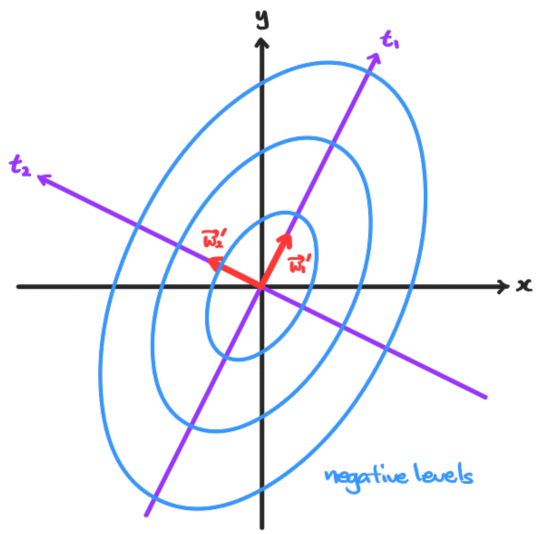
In an indefinite case, indicate as well which hyperbolas are $q(x, y) = c$ with $c > 0$ and which are $q(x, y) = c$ with $c < 0$.

(a) $q(x, y) = -13x^2 + 8xy - 7y^2$

The associated symm. matrix is $A = \begin{bmatrix} -13 & 4 \\ 4 & -7 \end{bmatrix}$, and $p_A(\lambda) = \lambda^2 + 20\lambda + 75$. The eigenvectors of A are $\vec{w}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{w}_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ with eigenvalues -5 and -15, respectively. Taking $\vec{w}_1' = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{w}_2' = \frac{1}{\sqrt{5}}\begin{bmatrix} -2 \\ 1 \end{bmatrix}$, if $\vec{v} = t_1\vec{w}_1' + t_2\vec{w}_2'$,

$$q_A(\vec{v}) = -5t_1^2 - 15t_2^2,$$

which is negative-definite.



(b) $q(x, y) = 7x^2 + 18xy - 17y^2$

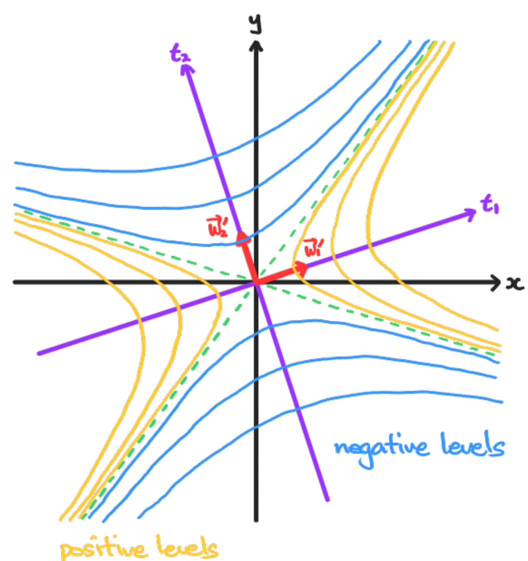
$A = \begin{bmatrix} 7 & 9 \\ 9 & -17 \end{bmatrix}$ has characteristic polynomial $p_A(\lambda) = \lambda^2 + 10\lambda - 200$. The eigenvectors of A are $\vec{w}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\vec{w}_2 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$ with eigenvalues 10 and -20, respectively. For $\vec{w}_1' = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\vec{w}_2' = \frac{1}{\sqrt{10}}\begin{bmatrix} -1 \\ 3 \end{bmatrix}$, if $\vec{v} = t_1\vec{w}_1' + t_2\vec{w}_2'$,

$$q_A(\vec{v}) = 10t_1^2 - 20t_2^2,$$

which is indefinite.

The hyperbola have asymptotes

$$t_2 = \pm\frac{1}{\sqrt{2}}t_1.$$



Example 7. Consider the function

$$f(x,y) = xy + 2x - \ln x^2 y$$

in the first quadrant $(x > 0, y > 0)$.

(a) Find the unique critical point.

Computing $f_x = y + 2 - \frac{2}{x}$ and $f_y = x - \frac{1}{y}$, we set them equal to 0 to get

$$x = \frac{1}{y} \quad \text{and} \quad 0 = y + 2 - 2y.$$

Hence, $y = 2$, $x = \frac{1}{2}$, and the unique critical point is $\left(\frac{1}{2}, 2\right)$.

(b) Find the Hessian of f at the critical point.

Using f_x and f_y from (a), we get

$$Hf(x,y) = \begin{bmatrix} \frac{2}{x^2} & 1 \\ 1 & \frac{1}{y^2} \end{bmatrix}.$$

Thus, $Hf\left(\frac{1}{2}, 2\right) = \begin{bmatrix} 8 & 1 \\ 1 & \frac{1}{4} \end{bmatrix}$.

(c) Give a quadratic approximation for f at the critical point.

The quadratic approximation of f at $\left(\frac{1}{2}, 2\right)$ is

$$f\left(\tfrac{1}{2}+h, 2+k\right) \approx f\left(\tfrac{1}{2}, 2\right) + \nabla f\left(\tfrac{1}{2}, 2\right) \cdot \begin{bmatrix} h \\ k \end{bmatrix} + \frac{1}{2} q_{Hf(\frac{1}{2},2)}(h,k)$$

$$= 2 + \ln 2 + 4h^2 + hk + \frac{1}{8}k^2.$$

(d) Sketch the contour plot of the quadratic approximation for f at the critical point.

The Hessian $Hf\left(\frac{1}{2}, 2\right)$ has characteristic polynomial $P_{Hf(\frac{1}{2},2)}(\lambda) = \lambda^2 - \frac{33}{4}\lambda + 1$.

$Hf\left(\frac{1}{2}, 2\right)$ has eigenvectors $\vec{v}_1 = \begin{bmatrix} 31+5\sqrt{41} \\ 8 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 31-5\sqrt{41} \\ 8 \end{bmatrix}$ with corresponding eigenvalues $\lambda_1 = \frac{33+5\sqrt{41}}{8}$ and $\lambda_2 = \frac{33-5\sqrt{41}}{8}$.

For $\vec{w}_1' = \frac{\vec{v}_1}{\|\vec{v}_1\|}$ and $\vec{w}_2' = \frac{\vec{v}_2}{\|\vec{v}_2\|}$,

$$q_{Hf(\frac{1}{2},2)}(\vec{v}) = \boxed{\lambda_1} t_1^2 + \boxed{\lambda_2} t_2^2$$

Both positive,
$|\lambda_1| > |\lambda_2|$.

if $\vec{v} = t_1\vec{w}_1' + t_2\vec{w}_2'$.

Recall that $\lambda_1 t_1'^2 + \lambda_2 t_2'^2 = 2\left(c - f\left(\frac{1}{2}, 2\right)\right)$ is a good approximation for $f(\vec{x}) = c$ for $\vec{x}$ near $\left(\frac{1}{2}, 2\right)$.

$2 + \ln 2$