PLRQ:

$$q_A(x,y) = x^2 + 2xy + 3y^2$$
$$= x^2 + 2xy + y^2 + 2y^2$$
$$= (x+y)^2 + 2y^2$$

$$f(x,y) = \ln(x^2+y^2)$$
$$f_x = \frac{2x}{x^2+y^2}$$
$$f_{xx} = \frac{2(x^2+y^2) - (2x)(2x)}{(x^2+y^2)^2}$$

$$f_y = \frac{2y}{x^2+y^2}$$
$$f_{yy} = \frac{2(x^2+y^2) - (2y)(2y)}{(x^2+y^2)^2}$$

$$f_{xx} + f_{yy} = \frac{2(x^2+y^2) - (2x)(2x)}{(x^2+y^2)^2} + \frac{2(x^2+y^2) - (2y)(2y)}{(x^2+y^2)^2}$$
$$= \frac{4(x^2+y^2) - 4x^2 - 4y^2}{(x^2+y^2)^2} = \frac{4x^2 + 4y^2 - 4x^2 - 4y^2}{(x^2+y^2)^2}$$
$$= \frac{0}{(x^2+y^2)^2} = 0$$

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

$$\begin{bmatrix} 4 & 6 \\ 6 & -2 \end{bmatrix}$$

Lecture 25 - The Hessian Matrix, Quadratic Approximation, and Local Extrema

December 2, 2022

Goals: Analyze the Hessian matrix of a multivariable function f to draw conclusions about f

scalar-valued

Definition: 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 $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = f_{x_j x_i}(\mathbf{a})$:

$(\nabla f(\vec{a}))^T = Df(\vec{a})$

$$(Hf)(\mathbf{a}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{a}) \end{bmatrix}.$$

$\dfrac{\partial^2 f}{\partial r \, \partial c}$

Note that the Hessian is exactly $(D(\nabla f))(\mathbf{a})$. Since $f_{x_i x_j} = f_{x_j x_i}$, **this matrix is always symmetric.**

Previously, we have discussed the linear approximation $f(\mathbf{a}+\mathbf{h}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{h}$. We can now discuss the *quadratic approximation* (think second degree Taylor approximation).

Definition: The **quadratic approximation** to $f : \mathbb{R}^n \to \mathbb{R}$ near $\mathbf{a} \in \mathbb{R}^n$ is

$\vec{h} \in \mathbb{R}^n \therefore h^T$ is $1 \times n$
$Hf(a) \to n \times n$
$h \to n \times 1$

Taylor-series
$p(a+h) \approx f(a) + \dfrac{f'(a)h}{1!} + \dfrac{f''(a)h^2}{2!}$

$$f(\mathbf{a}+\mathbf{h}) \approx \underbrace{f(\mathbf{a}) + (\nabla f)(\mathbf{a}) \cdot \mathbf{h}}_{\text{linear}} + \underbrace{\frac{1}{2}\mathbf{h}^\top (Hf(\mathbf{a}))\mathbf{h}}_{\hookrightarrow \text{ quad}(h)}.$$

Example 1: Let $f(x,y) = \ln(xy-1)$. Use quadratic approximation to estimate $f(1.2, 1.8)$.

$\vec{a} = (1,2)$

$\nabla f = \begin{bmatrix} y/(xy-1) \\ x/(xy-1) \end{bmatrix}, \quad \nabla f(1,2) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$Hf(1,2) = \begin{bmatrix} -4 & -1 \\ -1 & -1 \end{bmatrix}$

$f(1,2) = \ln(1) = 0$

$D(\nabla f) = Hf(x,y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$

$f(1.2, 1.8) \approx 0 + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} .2 \\ -.2 \end{bmatrix} + \frac{1}{2} [.2 \;\; -.2] \begin{bmatrix} -4 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} .2 \\ -.2 \end{bmatrix}$

$= \boxed{.14}$

$f_{xx} = -y^2/(xy-1)^2$

$f_{yy} = -x^2/(xy-1)^2$

$f_{xy} = \dfrac{(xy-1) - (x)(y)}{(xy-1)^2}$

$Hf(x,y) = \begin{bmatrix} -y^2/(xy-1)^2 & -1/(xy-1)^2 \\ -1/(xy-1)^2 & -x^2/(xy-1)^2 \end{bmatrix}$

The quadratic approximation is especially useful in understanding the behavior of a two variable function $f : \mathbb{R}^2 \to \mathbb{R}$ near a critical point via a contour plot. If $\mathbf{a}$ is a critical point of $f : \mathbb{R}^2 \to \mathbb{R}$, then by definition, $\nabla f(\mathbf{a}) = \mathbf{0}$, so the quadratic approximation simplifies to

$$f(\mathbf{a}+\mathbf{h}) \approx f(\mathbf{a}) + \frac{1}{2}\mathbf{h}^\top (Hf(\mathbf{a}))\mathbf{h} = \underbrace{f(\mathbf{a})}_{\text{we shift it to crit point}} + \underbrace{\frac{1}{2} q_{Hf(\mathbf{a})}(\mathbf{h})}_{\text{we find what this looks like at origin in tilted frame}}.$$

$$f(\mathbf{a}+\mathbf{h}) \approx \underbrace{f(\mathbf{a}) + (\nabla f)(\mathbf{a}) \cdot \mathbf{h}}_{\text{becomes zero}} + \frac{1}{2}\mathbf{h}^\top (Hf(\mathbf{a}))\mathbf{h}.$$

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

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

To summarize: 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 2: In lecture 24, we considered the symmetric matrix $A = \begin{bmatrix} 13 & -3 \\ -3 & 5 \end{bmatrix}$. Let us revisit this example.

Given: $\lambda_1 = 14$, $\lambda_2 = 4$, $\vec{w}_1 = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$, $\vec{w}_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$

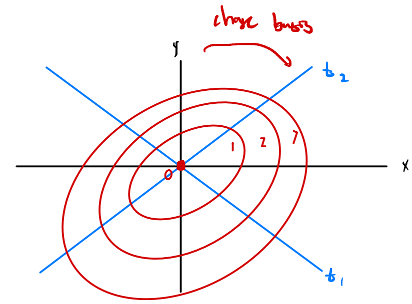
In diagonal form: $q_A(\vec{v}) = 14t_1^2 + 4t_2^2$

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

Level sets: $14t_1^2 + 4t_2^2 = c = 0$ point

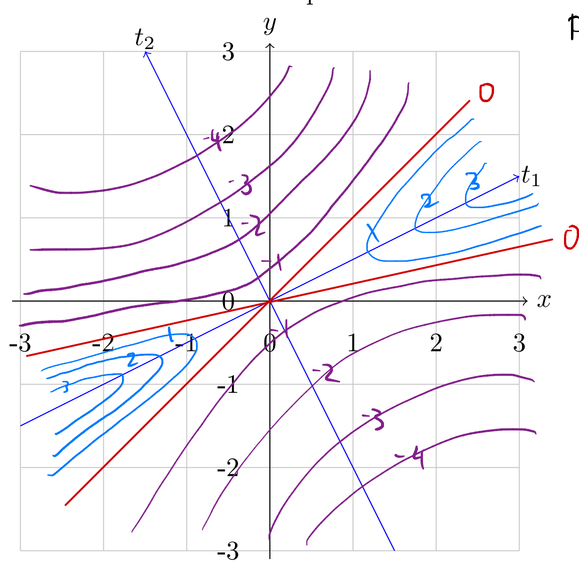
$14t_1^2 + 4t_2^2 = c > 0$ ellipse

$\Rightarrow \dfrac{t_1^2}{\left(\frac{1}{14}\right)} + \dfrac{t_2^2}{\left(\frac{1}{4}\right)} \Rightarrow$ ellipse is longer along the t_2-axis



So if a function $f : \mathbb{R}^2 \to \mathbb{R}$ has a critical point $\mathbf{a}$ such that $Hf(\mathbf{a}) = A$ in Example 2, then $\mathbf{a}$ will be a local minimum. Such an example is $f(x,y) = \frac{13}{2}(x-1)^2 + \frac{5}{2}y^2 - 3(x-1)y$ at the critical point $(1,0)$.

Example 3: Let $A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ 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? If $f : \mathbb{R}^2 \to \mathbb{R}$ has a critical point $\mathbf{a}$ such that $Hf(\mathbf{a}) = A$, then what kind of critical point is $\mathbf{a}$?



Find eigenvalues:

$p_A = \lambda^2 - \operatorname{tr}(A)\lambda + \det(A)$

$= \lambda^2 + \lambda - 6 = (\lambda+3)(\lambda-2) = 0$

$\underline{\lambda_1 = 2}$, $\underline{\lambda_2 = -3}$

Find eigenvectors:

$\underline{\lambda_1 = 2}$: Find $N(A - 2I_2) = N\left(\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix}\right) \to \begin{array}{l} -x_1 + 2x_2 = 0 \\ 2x_1 - 4x_2 = 0 \end{array}$

$\to x_1 = 2x_2 \to N = \operatorname{span}\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$

$\to$ eigenvector is $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$\underline{\lambda_2 = -3}$: Find $N(A + 3I_2) = N\left(\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}\right) \to \begin{array}{l} 4x_1 + 2x_2 = 0 \\ 2x_1 + x_2 = 0 \end{array}$

$\to x_2 = -2x_1 \to N = \operatorname{span}\left\{\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right\}$

$\to$ eigenvector is $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$

$\to q_A(t_1, t_2) = \underset{\uparrow \lambda_1}{2}t_1^2 - \underset{\uparrow \lambda_2}{3}t_2^2$

$q_A(t_1, t_2) = c = 0$

$2t_1^2 = 3t_2^2$

$t_2 = \pm\sqrt{\frac{2}{3}}\, t_1$

$q_A(t_1, t_2) = c > 0$

$2t_1^2 - 3t_2^2 = c > 0$

hyperbola, opens up to t_1 dir bc $c > 0$

$q_A(t_1, t_2) = c < 0$

$2t_1^2 - 3t_2^2 = c < 0$

hyperbola, opens up to t_2 dir bc $c < 0$

Therefore, if f has crit pt @ $\vec{a}$, and $Hf(\vec{a}) = A$, then $\vec{a}$ is a saddle point

The previous two examples give insight to the following:

If $f(x,y)$ has a critical point $\mathbf{a}$ such that $Hf(\mathbf{a})$ has eigenvalues $\lambda_1, \lambda_2 \neq 0$, then in a "tilted reference frame" (i.e. the orthogonal eigenvectors of $Hf(\mathbf{a})$), the quadratic approximation to $f(\mathbf{x})$ near $\mathbf{a}$ has level sets

$$\lambda_1 t_1^2 + \lambda_2 t_2^2 = \underbrace{2(c - f(\mathbf{a}))}_{\text{just another constant.}}$$

$\lambda_1 t_1^2 + \lambda_2 t_2^2$ shows how f behaves at crit point, centered at origin. This moves it to crit point.

So near $\mathbf{a}$, we have the following possibilities:

(1) When λ_1 and λ_2 have the same sign (positive/negative definite Hessian), then f looks like ellipses. The longer direction of the ellipse is along the eigenvectors who eigenvalue has smaller absolute value.

$$\lambda_1 t_1^2 + \lambda_2 t_2^2 = c$$

$$\to \frac{t_1^2}{\left(\frac{1}{\lambda_1}\right)} + \frac{t_2^2}{\left(\frac{1}{\lambda_2}\right)} = c$$

(2) When λ_1 and λ_2 have opposite signs (indefinite Hessian), f looks like hyperbolas with asymptotes closer to the eigenvectors of smaller magnitude eigenvalue.

general hyperbola: $x^2 - y^2 = 1 \to$ opens in x-dir → incr in pos eigval

$x^2 - y^2 = -1 \to y^2 - x^2 = 1 \to$ opens in y-dir

↳ decr in neg eigval

We finally reach the second derivative test for critical points for two variable functions:

- if λ_1 and λ_2 have **opposite signs**, then f has a **saddle point** at the critical point $\mathbf{a}$.
- if λ_1 and λ_2 are **both positive**, then f has a **local minimum** at the critical point $\mathbf{a}$.
- if λ_1 and λ_2 are **both negative**, then f has a **local maximum** at the critical point $\mathbf{a}$.
- if either of the eigenvalues are 0, we need more information and the test is inconclusive (compare to the second derivative test for single variable functions).

We will elaborate on this test more in Lecture 26. Let us review everything from today:

Example 4: Let $f(x,y) = x^3 + 3x^2 + y^2 - 4y$.

(a) Find the critical points of f.

$f_x = 3x^2 + 6x = 0 \to 3x(x+2) \to x = 0, -2$

$f_y = 2y - 4 = 0 \to y = 2$

Two critical points: $(0,2)$ & $(-2,2)$

(b) Find the Hessian matrix of f and its eigenvalues/eigenvectors at each critical point.

$f_{xx} = 6x+6$ $\quad f_{yy} = 2$ $\quad Hf(x,y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 6x+6 & 0 \\ 0 & 2 \end{bmatrix}$

$f_{xy} = 0$ $\quad f_{yx} = 0$

1) $Hf(0,2) = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$

$\lambda_1 = 6$, $\lambda_2 = 2$

$\underbrace{\vec{w}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{\text{bc diag matrix}}$, $\underbrace{\vec{w}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\text{bc diag matrix}}$

2) $Hf(-2,2) = \begin{bmatrix} -6 & 0 \\ 0 & 2 \end{bmatrix}$

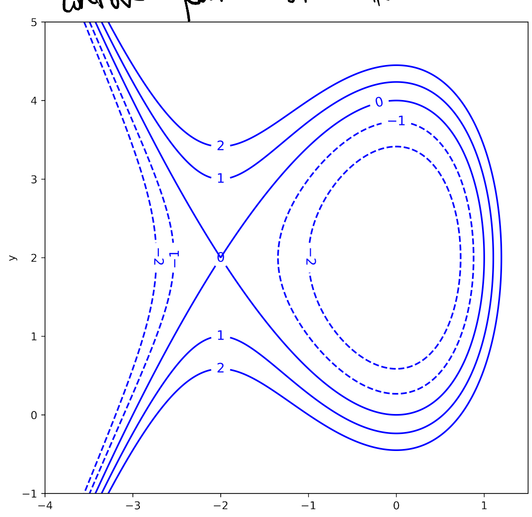
$\lambda_1 = -6$ $\quad \lambda_2 = 2$

$\vec{w}_1 = \vec{e}_1$ $\quad \vec{w}_2 = \vec{e}_2$

for diag matrix, eigvecs are standard basis vectors

(c) Sketch the contour plot of the quadratic approximation of f near each critical point and classify the critical points as a local max/min or saddle point.

contour plot of f



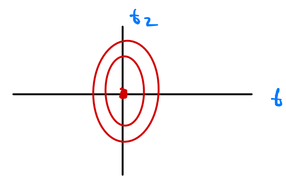
@ (2, −2), eigen vals are opp sign so hyperbola

@ (0, 2), eigvals are same sign so ellipse

(0, 2)

$q_{Hf(0,2)}(t_1, t_2) = 6t_1^2 + 2t_2^2 = c$

→ ellipse, longer along t_2 axis



eigvecs are standard basis

(2, −2)

look at quadratic of Hf (2, −2) in $\underbrace{\text{tilted reference}}$ frame

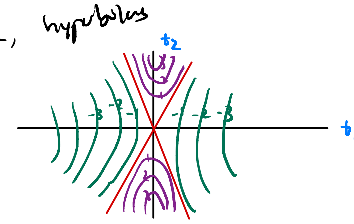
$q_{Hf(2,-2)}(t_1, t_2) = -6t_1^2 + 2t_2^2 = c$

@ c = 0, it's a pair of lines closer to t_2

otherwise, hyperbolas

c > 0, open to t_2

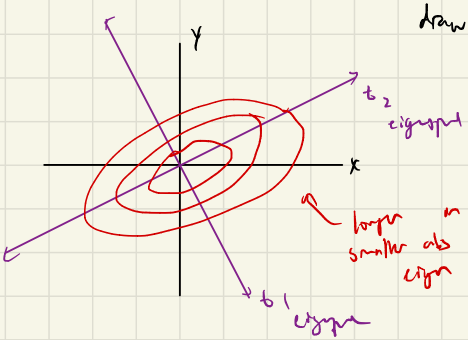
c < 0, open to t_1



Ex 1

$\lambda_1 = 2, \quad \lambda_2 = 1$

$\begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad \begin{pmatrix} 2 \\ 1 \end{pmatrix}$



draw w/ respect to $\begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ as basis vectors

$q_{H}(u) = 2t_1^2 + t_2^2 = 0$ point

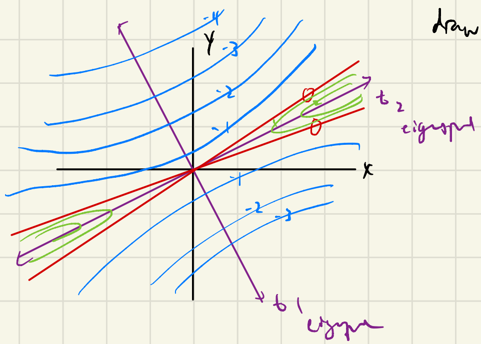
$2t_1^2 + t_2^2 > 0$ ellipse

$2t_1^2 + t_2^2 < 0$ empty set, not possible

Ex 2

$\lambda_1 = -2, \quad \lambda_2 = 1$

$\begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad \begin{pmatrix} 2 \\ 1 \end{pmatrix}$



draw w/ respect to $\begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ as basis vectors

$q_{H}(u) = -2t_1^2 + t_2^2 = 0$ point

$-2t_1^2 + t_2^2 > 0$ hyperbola

$-2t_1^2 + t_2^2 < 0$ hyperbola

$-2t_1^2 + t_2^2 = 1$ pos same sign, opens to t_2 axis

$-2t_1^2 + t_2^2 = -1$ neg same sign, opens to t_1 axis

$q_A = x^2 + y^2 - 2xy$

$\Rightarrow A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

$\Rightarrow \lambda^2 - (2)\lambda + 0 = 0 \to \lambda(\lambda - 2) = 0 \longrightarrow \lambda_1 = 0, \ \lambda_2 = 2$

$\Rightarrow q_A = 0t_1^2 + 2t_2^2$