Quadratic Forms 13

Our study of extrema of multivariable functions has focused so far on:

- Conditions on the function *necessary* for a local or global extremum to occur ($\vec{\nabla} f(\vec{a}) = \vec{0}$, $\vec{a}$ = boundary / "bad" point)
- Conditions on the domain $\mathcal{S}$ that ensure any continuous function has global extrema ($\mathcal{S}$ is closed + bounded, i.e. compact)

Today we will discuss how the second partials of a C^2 function tell us when a stationary point ($\vec{\nabla} f(\vec{a}) = \vec{0}$) is a local maximum, local minimum, or saddle point. This is one of the first places matrix algebra really "barges in" to Calculus.

§1. The 2nd Derivative Test

In the first example of Lecture 12, we looked at $f\begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 4xy + y^2$ near $\vec{a} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. The surprise was that, although $f_{xx}(\vec{0}) \left(= \frac{\partial^2 f}{\partial x^2}(\vec{0}) \right) = 2 = f_{yy}(\vec{0})$, f actually had a saddle point at $\vec{0}$. We detected this by computing (with $\hat{u} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$)

$$D_{\hat{u}}^2 f(\vec{0}) = D_{\hat{u}}\{\sqrt{2}(y - x)\}(\vec{0}) = -2.$$

Note: even weirder things can happen. $f\begin{pmatrix} x \\ y \end{pmatrix} = 2x^4 - 3x^2 y + y^2$ has local minimum at $\vec{0}$ along all *lines* through the origin, but still has a saddle point there! (Consider its restriction to $y = \frac{3}{2}x^2$, which has a local MAX at $\vec{0}$.) This funny business is only possible, though, because one of the $D_{\hat{u}}^2 f(\vec{0})$'s is 0. If they're all *positive*, we'll have a local minimum.

To explore further, let $\vec{a}$ be a stationary point of a function $f(x,y)$, and set $\hat{u} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \begin{matrix} u_x \\ u_y \end{matrix}$

$$D^2_{\hat{u}} f = D_{\hat{u}}(u_x f_x + u_y f_y)$$

$$= \vec{\nabla}(u_x f_x + u_y f_y) \cdot \hat{u}$$

$$= (u_x f_{xx} + u_y f_{yx},\ u_x f_{xy} + u_y f_{yy}) \cdot (u_x, u_y)$$

$$= (u_x, u_y) \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}$$

$$= {}^t\hat{u}\, H\, \hat{u} \quad (= \hat{u} \cdot H\hat{u})$$

The 2×2-matrix-valued function $H(x,y)$ is called the Hessian of f.

Write H_0 for $H(\vec{a})$.

So $\hat{u} \cdot H_0 \hat{u}$ is the concavity at $\vec{a}$ in the direction $\hat{u}$, which depends on θ.

We want the directions of extreme concavity.

unit circle is COMPACT, so a max & min exist

To explore further, let $\vec{a}$ be a stationary point of a function $f(x,y)$, and set $\hat{u} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \begin{matrix} u_x \\ u_y \end{matrix}$

$$D^2_{\hat{u}} f = D_{\hat{u}}(u_x f_x + u_y f_y)$$

$$= \vec{\nabla}(u_x f_x + u_y f_y) \cdot \hat{u}$$

$$= (u_x f_{xx} + u_y f_{yx},\ u_x f_{xy} + u_y f_{yy}) \cdot (u_x, u_y)$$

$$= (u_x, u_y) \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}$$

$$= {}^t\hat{u}\, H\, \hat{u} \quad (= \hat{u} \cdot H\hat{u})$$

The 2×2-matrix-valued function $H(x,y)$ is called the Hessian of f.

Write H_0 for $H(\vec{a})$.

So $\hat{u} \cdot H_0 \hat{u}$ is the concavity at $\vec{a}$ in the direction $\hat{u}$, which depends on θ.

We want the directions of extreme concavity:

$$0 \overset{!}{=} \frac{d}{d\theta}\, \hat{u} \cdot H_0 \hat{u} = \hat{u}' \cdot H_0 \hat{u} + \hat{u} \cdot H_0 \hat{u}'$$

↑ to find stationary pts. as function of θ

To explore further, let $\vec{a}$ be a stationary point of a function $f(x,y)$, and set $\hat{u} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \begin{matrix} u_x \\ u_y \end{matrix}$

$$D^2_{\hat{u}} f = D_{\hat{u}}(u_x f_x + u_y f_y)$$

$$= \vec{\nabla}(u_x f_x + u_y f_y) \cdot \hat{u}$$

$$= (u_x f_{xx} + u_y f_{yx},\ u_x f_{xy} + u_y f_{yy}) \cdot (u_x, u_y)$$

$$= (u_x, u_y) \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}$$

$$= {}^t\hat{u}\, H\, \hat{u} \quad (= \hat{u} \cdot H\hat{u})$$

The 2×2-matrix-valued function $H(x,y)$ is called the Hessian of f.

Write H_0 for $H(\vec{a})$.

So $\hat{u} \cdot H_0 \hat{u}$ is the concavity at $\vec{a}$ in the direction $\hat{u}$, which depends on θ.

We want the directions of extreme concavity:

$$0 = \frac{d}{d\theta}\, \hat{u} \cdot H_0 \hat{u} = \hat{u}' \cdot H_0 \hat{u} + \hat{u} \cdot H_0 \hat{u}'$$

$$(=)\ 2\hat{u}' \cdot H_0 \hat{u} \Rightarrow \hat{u}' \perp H_0 \hat{u}$$

H_0 symmetric $(f_{xy} = f_{yx})$ assuming 2nd derivatives continuous

To explore further, let $\vec{a}$ be a stationary point of a function $f(x,y)$, and set $\hat{u} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \begin{matrix} u_x \\ u_y \end{matrix}$

$$D_{\hat{u}}^2 f = D_{\hat{u}}(u_x f_x + u_y f_y)$$

$$= \vec{\nabla}(u_x f_x + u_y f_y) \cdot \hat{u}$$

$$= (u_x f_{xx} + u_y f_{yx},\ u_x f_{xy} + u_y f_{yy}) \cdot (u_x, u_y)$$

$$= (u_x, u_y) \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}$$

$$= {}^t\hat{u}\, H\, \hat{u} \quad (= \hat{u} \cdot H\hat{u})$$

The 2×2-matrix-valued function $H(x,y)$ is called the Hessian of f.

Write H_0 for $H(\vec{a})$.

So $\hat{u} \cdot H_0 \hat{u}$ is the concavity at $\vec{a}$ in the direction $\hat{u}$, which depends on θ.

We want the directions of extreme concavity:

$$0 = \frac{d}{d\theta}\, \hat{u} \cdot H_0 \hat{u} = \hat{u}' \cdot H_0 \hat{u} + \hat{u} \cdot H_0 \hat{u}'$$

$$= 2\hat{u}' \cdot H_0 \hat{u} \Rightarrow \hat{u}' \perp H_0 \hat{u}$$

$\Rightarrow$ $\hat{u} \parallel H_0 \hat{u}$ $\Rightarrow$ $\hat{u}$ eigenvector of H_0!

$\hat{u}' = \frac{d}{d\theta}\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$

$= \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$

is $\perp$ to $\hat{u}$

To explore further, let $\vec{a}$ be a stationary point of a function $f(x,y)$, and set $\hat{u} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \begin{matrix} u_x \\ u_y \end{matrix}$

$$D^2_{\hat{u}} f = D_{\hat{u}}(u_x f_x + u_y f_y)$$

$$= \vec{\nabla}(u_x f_x + u_y f_y) \cdot \hat{u}$$

$$= (u_x f_{xx} + u_y f_{yx},\ u_x f_{xy} + u_y f_{yy}) \cdot (u_x, u_y)$$

$$= (u_x, u_y) \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}$$

$$= {}^t\hat{u}\, H\, \hat{u} \quad (= \hat{u} \cdot H\hat{u})$$

The 2×2-matrix-valued function $H(x,y)$ is called the Hessian of f. Write H_0 for $H(\vec{a})$.

So $\hat{u} \cdot H_0 \hat{u}$ is the concavity at $\vec{a}$ in the direction $\hat{u}$, which depends on θ.

We want the directions of extreme concavity:

$$0 = \frac{d}{d\theta}\, \hat{u} \cdot H_0 \hat{u} = \hat{u}' \cdot H_0 \hat{u} + \hat{u} \cdot H_0 \hat{u}'$$

$$= 2\hat{u}' \cdot H_0 \hat{u} \Rightarrow \hat{u}' \perp H_0 \hat{u}$$

$$\Rightarrow \hat{u} \parallel H_0 \hat{u} \Rightarrow \hat{u} \text{ eigenvector of } H_0 !$$

If $\hat{u} = \vec{v}$:= eigenvector w/eigenvalue λ,

then $D^2_{\hat{u}} f = \vec{v} \cdot H_0 \vec{v} = \vec{v} \cdot \lambda \vec{v} = \lambda \|\vec{v}\|^2 \to 1$

$= \lambda$ is the concavity.

Writing $\lambda_1 \le \lambda_2$ for the 2 eigenvalues, these are the min/max. concavities of f at $\vec{a} \Rightarrow$

FACT 1 $\Delta := \det \begin{pmatrix} f_{xx}(\vec{a}) & f_{xy}(\vec{a}) \\ f_{yx}(\vec{a}) & f_{yy}(\vec{a}) \end{pmatrix}$ is the product of the max + min concavities of f at $\vec{a}$.

But if $\Delta > 0$ this tells me only that $\lambda_1, \lambda_2 > 0$ or $\lambda_1, \lambda_2 < 0$. Which is it?

Consider $f_{xx}(\vec{a})$, the concavity in the x-direction — as such, clearly in between the max/min concavities:

$$\lambda_1 \le f_{xx}(\vec{a}) \le \lambda_2 .$$

So if $f_{xx}(\vec{a}) > 0$, then $\lambda_2 > 0$; and since we assumed $\Delta = \lambda_1\lambda_2 > 0$, $\lambda_1 > 0$ too! $\Rightarrow$

FACT 2

- If $\Delta > 0$ and $f_{xx}(\vec{a}) > 0$ then $\lambda_1, \lambda_2 > 0$ and $\vec{a}$ is a local minimum.
- If $\Delta > 0$ and $f_{xx}(\vec{a}) < 0$ then $\lambda_1, \lambda_2 < 0$ and $\vec{a}$ is a local maximum.
- If $\Delta < 0$, then $\lambda_1 < 0$, $\lambda_2 > 0$, and $\vec{a}$ is a saddle point.
- If $\Delta = 0$, the test is inconclusive and you need to try something else!

So $\hat{u} \cdot H_0 \hat{u}$ is the concavity at $\vec{a}$ in the direction $\hat{u}$, which depends on θ.

We want the directions of extreme concavity:

$$0 = \frac{d}{d\theta}\, \hat{u} \cdot H_0 \hat{u} = \hat{u}' \cdot H_0 \hat{u} + \hat{u} \cdot H_0 \hat{u}'$$

$$= 2\hat{u}' \cdot H_0 \hat{u} \Rightarrow \hat{u}' \perp H_0 \hat{u}$$

$$\Rightarrow \hat{u} \parallel H_0 \hat{u} \Rightarrow \hat{u} \text{ eigenvector of } H_0 !$$

If $\hat{u} = \vec{v}$:= eigenvector w/eigenvalue λ,

then $D^2_{\hat{u}} f = \vec{v} \cdot H_0 \vec{v} = \vec{v} \cdot \lambda \vec{v} = \lambda \|\vec{v}\|^2 \to 1$

$= \lambda$ is the concavity.

Writing $\lambda_1 \le \lambda_2$ for the 2 eigenvalues, these are the min/max. concavities of f at $\vec{a}$ $\Rightarrow$

FACT 1 $\Delta := \det \begin{pmatrix} f_{xx}(\vec{a}) & f_{xy}(\vec{a}) \\ f_{yx}(\vec{a}) & f_{yy}(\vec{a}) \end{pmatrix}$ is the product of the max + min concavities of f at $\vec{a}$.

Given f twice continuously differentiable on a neighborhood of $\vec{a}$, with $\vec{\nabla} f(\vec{a}) = 0$.

FACT 1 $\Delta := \det \begin{pmatrix} f_{xx}(\vec{a}) & f_{xy}(\vec{a}) \\ f_{yx}(\vec{a}) & f_{yy}(\vec{a}) \end{pmatrix}$ is the product of the max + min concavities of f at $\vec{a}$.

FACT 2

- If $\Delta > 0$ and $f_{xx}(\vec{a}) > 0$ then $\lambda_1, \lambda_2 > 0$ and $\vec{a}$ is a local minimum.
- If $\Delta > 0$ and $f_{xx}(\vec{a}) < 0$ then $\lambda_1, \lambda_2 < 0$ and $\vec{a}$ is a local maximum.
- If $\Delta < 0$, then $\lambda_1 < 0$, $\lambda_2 > 0$, and $\vec{a}$ is a saddle point.
- If $\Delta = 0$, the test is inconclusive and you need to try something else!

Remark: these correspond to the quadratic form $Q(\vec{h}) = {}^t\vec{h}\, H(\vec{a})\, \vec{h}$ being positive definite / negative definite / indefinite.

Returning to the Example: $\vec{a} = \vec{0}$,

$$f(x,y) = x^2 + 4xy + y^2,$$

we have for the Hessian

$$H(x,y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 4 & 2 \end{pmatrix}$$

hence $\Delta = 2^2 - 4^2 = -12 < 0$

$\Rightarrow$ saddle.

Here are a couple of problems for you to try: find the local extrema and saddle points of

① $f(x,y) = x^4 + y^4 - 4xy + 1$

② $f(x,y) = x^3 - 3xy^2$

① $f(x,y) = x^4 + y^4 - 4xy + 1$

First solve $f_x = 0 = f_y$ $\rightsquigarrow$ $4x^3 - 4y = 0$, $4y^3 - 4x = 0$

$\rightsquigarrow$ $x^3 = y$, $y^3 = x$ $\rightsquigarrow (y^3)^3 = y$ $\rightsquigarrow 0 = y^9 - y = y(y^8 - 1)$

$\rightsquigarrow y = 0, 1, -1$ (since y real) $\rightsquigarrow$ the 3 stationary points are $\vec{a} = (0,0), (1,1), (-1,-1)$.

Now $H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}$

$\Rightarrow \det H = 144x^2y^2 - 16$. So

- at $\vec{a} = (0,0)$, $\Delta = -16 < 0$ $\rightsquigarrow$ saddle pt.
- at $\vec{a} = (1,1)$, $\begin{cases} \Delta = 128 > 0 \\ f_{xx} = 12 > 0 \end{cases}$ $\rightsquigarrow$ local minimum
- at $\vec{a} = (-1,-1)$, $\begin{cases} \Delta = 128 > 0 \\ f_{xx} = 12 > 0 \end{cases}$ $\rightsquigarrow$ local minimum

② $f(x,y) = x^3 - 3xy^2$

First solve $f_x = 0 = f_y$ $\rightsquigarrow$ $3x^2 - 3y^2 = 0$, $-6xy = 0$

$\rightsquigarrow (x,y) = (0,0)$ (why?).

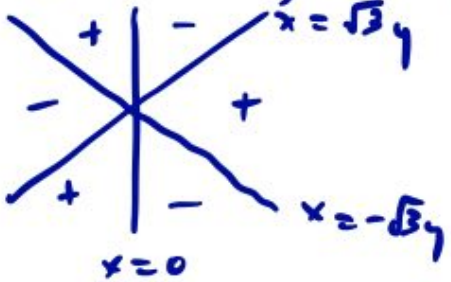
So $\vec{a} = (0,0)$ is the only stationary point.

But $\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 6x & -6y \\ -6y & -6x \end{pmatrix}$ is 0 at $\vec{a}$,

hence $\Delta = 0$ & test inconclusive.

Rather, it's enough to write $f(x,0) = x^3$ to see that it's a saddle point — even in the x-dir. we don't have a max or min.

You could also go into greater detail by writing $f(x,y) = x(x - \sqrt{3}y)(x + \sqrt{3}y)$



and observing that the sign [of $f(x,y) - f(0,0) = f(x,y)$] changes as you go from chamber to chamber.

§2. Eigenstuff (Cliff's Notes version)

Remembering a few "309" basics here really helps to understand what is going on in Shifrin §5.3.

Recall that for an $n \times n$ matrix A,

$0 \neq \vec{v} \in \mathbb{R}^n$ is an eigenvector of A with eigenvalue $\lambda \in \mathbb{R} \iff A\vec{v} = \lambda\vec{v}$

- $A\vec{v} = \lambda\vec{v} \iff 0 = (\lambda \mathbb{I}_n - A)\vec{v}$
  $\iff \vec{v} \in \mathrm{Nul}(\lambda \mathbb{I}_n - A)$
- λ eigenvalue $\iff \lambda \mathbb{I}_n - A$ not invertible
  $\iff \det(\lambda \mathbb{I}_n - A) = 0$
  $\iff f_A(x) := \det(\lambda \mathbb{I}_n - A)$ has λ as a root

If $\vec{v}_1, \dots, \vec{v}_n$ is a basis of $\mathbb{R}^n$ with $A\vec{v}_i = \lambda_i \vec{v}_i$ (an eigenbasis), set $S := \begin{pmatrix} \uparrow & & \uparrow \\ \vec{v}_1 & \cdots & \vec{v}_n \\ \downarrow & & \downarrow \end{pmatrix}$.

- $S^{-1}AS\,\hat{e}_i = S^{-1}A\vec{v}_i = S^{-1}\lambda_i\vec{v}_i = \lambda_i \hat{e}_i$

$\Rightarrow S^{-1}AS = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: D$

$\Rightarrow A = SDS^{-1}$.

We'll prove later that $\begin{cases} \det(BC) = (\det B)(\det C) \\ \det(B^{-1}) = 1/(\det B), \end{cases}$

so

- $\det(A) = \det(D) = \lambda_1 \cdots \lambda_n$.

If the $\{\vec{v}_i\}$ are orthonormal $(\vec{v}_i \cdot \vec{v}_j = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases})$ then ${}^tSS = \begin{pmatrix} \leftarrow {}^t\vec{v}_1 \rightarrow \\ \vdots \\ \leftarrow {}^t\vec{v}_n \rightarrow \end{pmatrix} \begin{pmatrix} \uparrow & & \uparrow \\ \vec{v}_1 & \cdots & \vec{v}_n \\ \downarrow & & \downarrow \end{pmatrix} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$

$\Rightarrow {}^tS = S^{-1} \Rightarrow$ $\vec{y} := {}^tS\vec{x}$

- $Q(\vec{x}) := {}^t\vec{x}A\vec{x} = {}^t\vec{x}SD{}^tS\vec{x} \overset{\downarrow}{=} {}^t\vec{y}D\vec{y}$
  $= (y_1 \cdots y_n) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$
  $= \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$.

Conclude that (in this case) if all $\lambda_i \overset{<}{>} 0$, then $Q(\vec{x}) \overset{<}{>} 0$ for any $\vec{x} \neq \vec{0} \Rightarrow$ has (maximum / minimum) at $\vec{0}$.

Now suppose A is a symmetric matrix ($a_{ij} = a_{ji}$). Then writing $Q(\vec{x}) = A\vec{x} \cdot \vec{x}$,

$$(DQ)\vec{h} = (DA\vec{x})\vec{h} \cdot \vec{x} + A\vec{x} \cdot (D\vec{x})\vec{h}$$

($DA = A$, $D\vec{x} = I_n$)

$$= 2A\vec{x} \cdot \vec{h} \quad \Rightarrow \vec{\nabla} Q = 2A\vec{x}$$

Since the unit sphere $g(\vec{x}) := \|\vec{x}\|^2 = 1$ ($x_1^2 + \cdots + x_n^2$) is compact, $Q(\vec{x})$ has a maximum there at some $\vec{v}_1$. Finding $\vec{v}_1$ is a constrained extremum problem, which you learned to solve by using Lagrange multipliers, i.e. writing

$$\vec{\nabla} Q(\vec{v}_1) = \lambda_1 \vec{\nabla} g(\vec{v}_1)$$

and solving for λ_1 & $\vec{v}_1$:

($\vec{\nabla} g = \begin{pmatrix} 2x_1 \\ \vdots \\ 2x_n \end{pmatrix} = 2\vec{x}$)

$$2A\vec{v}_1 = \lambda_1 2\vec{v}_1$$

$$\Rightarrow A\vec{v}_1 = \lambda_1 \vec{v}_1.$$

So we have found an eigenvector (of unit length)!

Observe that for any $\vec{w} \perp \vec{v}_1$,

$$\vec{v}_1 \cdot A\vec{w} = A\vec{v}_1 \cdot \vec{w} = \lambda_1 \vec{v}_1 \cdot \vec{w} = 0 \Rightarrow A\vec{w} \perp \vec{v}_1.$$

Now the unit sphere intersected with the orthogonal complement

$$W := \{f(\vec{x}) = 0\} \quad (f(\vec{x}) := 2\vec{v}_1 \cdot \vec{x})$$

of $\vec{v}_1$ is compact, so again $Q(\vec{x})$ has a maximum at some $\vec{v}_2$. Since we have 2 constraints, we use 2 Lagrange multipliers:

$$\underbrace{\vec{\nabla} Q(\vec{v}_2)}_{2A\vec{v}_2} = \lambda_2 \underbrace{\vec{\nabla} g(\vec{v}_2)}_{2\vec{v}_2} + \mu_2 \underbrace{\vec{\nabla} f(\vec{v}_2)}_{2\vec{v}_1}$$

$$A\vec{v}_2 = \lambda_2 \vec{v}_2 + \mu_2 \vec{v}_1$$

and dotting both sides with $\vec{v}_1$ gives $0 = \mu_2$ hence

$$A\vec{v}_2 = \lambda_2 \vec{v}_2.$$

Continuing in this fashion, we get an orthonormal eigenbasis of $\mathbb{R}^n$ (basis b/c independent b/c orthogonal – cf. HW 5!).

Spectral Theorem: Given an $n \times n$ symmetric matrix A, there is an orthonormal basis $\vec{v}_1, \ldots, \vec{v}_n$ of $\mathbb{R}^n$ with $A\vec{v}_i = \lambda_i \vec{v}_i$.
Moreover, the associated quadratic form $Q(\vec{x}) = \vec{x} \cdot A\vec{x}$ can be rewritten as $\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$ in the coordinates $\vec{y} = {}^t S \vec{x}$.

Definition: (i) Q is

positive definite $\Leftrightarrow$ $Q(\vec{x}) > 0$ $\forall \vec{x} \neq \vec{0}$
positive semidefinite $\Leftrightarrow$ $Q(\vec{x}) \geq 0$ $\forall \vec{x}$
negative definite $\Leftrightarrow$ $Q(\vec{x}) < 0$ $\forall \vec{x} \neq \vec{0}$
negative semidefinite $\Leftrightarrow$ $Q(\vec{x}) \leq 0$ $\forall \vec{x}$
indefinite $\Leftrightarrow$ $Q(\vec{x})$ takes both positive & negative values

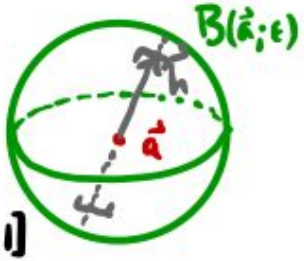
(ii) A is

positive definite $\Leftrightarrow$ all $\lambda_i > 0$
positive semidefinite $\Leftrightarrow$ all $\lambda_i \geq 0$
negative definite $\Leftrightarrow$ all $\lambda_i < 0$
negative semidefinite $\Leftrightarrow$ all $\lambda_i \leq 0$
indefinite $\Leftrightarrow$ A has both positive & negative eigenvalues.

Corollary: Q is [ ] $\Leftrightarrow$ A is [ ].

§3. The 2nd Derivative Test (v. 2)

Now we generalize to n variables and examine $f(\vec{x})$ on the grey axis in the picture:



$g(u) := f(\vec{a}+u\vec{h})$ for $u \in [-1,1]$

$$\Downarrow$$

$$g'(u) = \vec{\nabla} f(\vec{a}+u\vec{h}) \cdot \vec{h} = \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(\vec{a}+u\vec{h})$$

$$\Downarrow \text{ chain rule again}$$

$$g''(u) = \vec{\nabla}\left(\sum_j h_j \frac{\partial f}{\partial x_j}(\vec{a}+u\vec{h})\right)\cdot\vec{h} = \sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a}+u\vec{h})$$

$$= {}^t\vec{h}\, H(\vec{a}+u\vec{h})\, \vec{h} \qquad \left(H(\vec{x}) := \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x})\right]\right)$$

Hessian of f

$$\Downarrow$$

(Taylor remainder $c \in (0,1)$)

$$f(\vec{a}+\vec{h}) - f(\vec{a}) = g(1) - g(0) = g'(0) + \tfrac{1}{2} g''(c)$$

$$= \vec{\nabla} f(\vec{a})\cdot\vec{h} + \tfrac{1}{2}\, {}^t\vec{h}\, H(\vec{a}+c\vec{h})\,\vec{h}$$

$$= \vec{\nabla} f(\vec{a})\cdot\vec{h} + \tfrac{1}{2}\, {}^t\vec{h}\, H(\vec{a})\,\vec{h} + \|\vec{h}\|^2 E_2(\vec{a};\vec{h})$$

where $E_2(\vec{a};\vec{h})$ is $o(1)$ $(\to 0$ as $\vec{h} \to \vec{0})$ assuming f is twice continuously differentiable on $B(\vec{a};\epsilon)$.

where $E_2(\vec{a};\vec{h}) := \begin{cases} {}^t\vec{h}\,\dfrac{H(\vec{a}+c\vec{h}) - H(\vec{a})}{2\|\vec{h}\|^2}\,\vec{h}, & \vec{h} \neq \vec{0} \\ 0, & \vec{h} = \vec{0} \end{cases}$

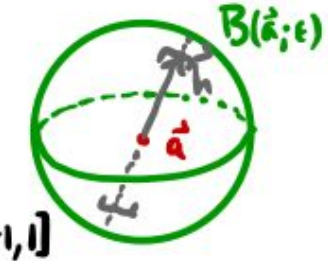
has $|E_2(\vec{a};\vec{h})| = \dfrac{1}{2\|\vec{h}\|^2}\left|\sum_i\sum_j h_i h_j \left(\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a}+c\vec{h}) - \dfrac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a})\right)\right|$

$$\le \frac{1}{2\|\vec{h}\|^2}\sum_i\sum_j |h_i||h_j| \left|\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a}+c\vec{h}) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a})\right|$$

$$\le \frac{\|\vec{h}\|^2}{2\|\vec{h}\|^2}\sum_i\sum_j \left|\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a}+c\vec{h}) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a})\right|$$

$\to 0$ as $\vec{h} \to \vec{0}$ by continuity of 2nd partials

Now we generalize to n variables and examine $f(\vec{x})$ on the grey axis in the picture:



$g(u) := f(\vec{a}+u\vec{h})$ for $u \in [-1,1]$

$\Downarrow$

$$g'(u) = \vec{\nabla} f(\vec{a}+u\vec{h}) \cdot \vec{h} = \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(\vec{a}+u\vec{h})$$

$\Downarrow$ chain rule again

$$g''(u) = \vec{\nabla}\Big(\sum_j h_j \frac{\partial f}{\partial x_j}(\vec{a}+u\vec{h})\Big)\cdot\vec{h} = \sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a}+u\vec{h})$$

$$= {}^t\vec{h}\, H(\vec{a}+u\vec{h})\, \vec{h} \qquad \left(H(\vec{x}) := \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x})\right]\right)$$

$\Downarrow$

(Taylor remainder $c \in (0,1)$) Hessian of f

$$f(\vec{a}+\vec{h}) - f(\vec{a}) = g(1) - g(0) = g'(0) + \tfrac{1}{2} g''(c)$$

$$= \vec{\nabla} f(\vec{a})\cdot\vec{h} + \tfrac{1}{2}\, {}^t\vec{h}\, H(\vec{a}+c\vec{h})\,\vec{h}$$

$$= \vec{\nabla} f(\vec{a})\cdot\vec{h} + \tfrac{1}{2}\, {}^t\vec{h}\, H(\vec{a})\,\vec{h} + \|\vec{h}\|^2 E_2(\vec{a};\vec{h})$$

where $E_2(\vec{a};\vec{h})$ is $o(1)$ ($\to 0$ as $\vec{h} \to \vec{0}$) assuming f is twice continuously differentiable on $B(\vec{a};\epsilon)$.

Lemma on Quadratic Forms: Let $Q(\vec{h}) := {}^t\vec{h}\, A\, \vec{h}$, A symmetric $n\times n$ with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Then

(i) $Q(\vec{h}) > 0 \ \forall \vec{h} \neq \vec{0} \iff$ all $\lambda_j > 0 \iff A$ pos.-definite

(ii) $Q(\vec{h}) < 0 \ \forall \vec{h} \neq \vec{0} \iff$ all $\lambda_j < 0 \iff A$ neg.-definite

(iii) $Q(\vec{h})$ takes +/- values $\iff \lambda_1 < 0 < \lambda_n \iff A$ indefinite.

(This is just the Corollary to the Spectral Theorem, better stated.)

Theorem 2: If f is twice continuously diff'able, with a stationary point at $\vec{x} = \vec{a}$, then:

(i) $H(\vec{a})$ pos.-definite $\Rightarrow f$ has rel. minimum at $\vec{a}$

(ii) $H(\vec{a})$ neg.-definite $\Rightarrow f$ has rel. maximum at $\vec{a}$

(iii) $H(\vec{a})$ indefinite $\Rightarrow f$ has saddle point at $\vec{a}$.

Proof: We record (before erasing) the fact that

$$f(\vec{a}+\vec{h}) - f(\vec{a}) = \vec{\nabla} f(\vec{a})\cdot\vec{h} + {}^t\vec{h}\, H(\vec{a})\,\vec{h} + \|\vec{h}\|^2 E(\vec{a};\vec{h}).$$

Since $\vec{\nabla} f(\vec{a}) = \vec{0}$ at a stationary point,

$$f(\vec{a}+\vec{h}) - f(\vec{a}) = {}^t\vec{h}\, H(\vec{a})\vec{h} + \|\vec{h}\|^2 E_2(\vec{a};\vec{h}).$$

The main idea is then that this term doesn't matter, and so we are done by the Lemma. Here is how this works for (i):

- let $0 < \lambda_1 \le \cdots \le \lambda_n$ be eigenvalues of $H(\vec{a})$.
  pick $\epsilon > 0$ s.t. $|E_2(\vec{a};\vec{h})| < \frac{1}{4}\lambda_1$ for $0 < \|\vec{h}\| < \epsilon$.
- let $u \in (0, \lambda_1)$, so that the eigenvalues $\lambda_j - u$ of $H(\vec{a}) - uI_n$ are all positive.
  Lemma $\Rightarrow {}^t\vec{h}\,[H(\vec{a}) - uI_n]\vec{h} > 0 \quad \forall\, \vec{h} \neq \vec{0}$.
- so ${}^t\vec{h}\, H(\vec{a})\vec{h} > u\|\vec{h}\|^2 \quad \forall\, \vec{h} \neq \vec{0}$, and taking $u = \frac{1}{2}\lambda_1$ gives

$$\tfrac{1}{2}\,{}^t\vec{h}\, H(\vec{a})\vec{h} > \tfrac{1}{4}\lambda_1\|\vec{h}\|^2 > \|\vec{h}\|^2 |E_2(\vec{a};\vec{h})| \ge 0$$

$$\Rightarrow f(\vec{a}+\vec{h}) - f(\vec{a}) = \tfrac{1}{2}\,{}^t\vec{h}\, H(\vec{a})\vec{h} + \|\vec{h}\|^2 E_2(\vec{a};\vec{h})$$
$$\ge \tfrac{1}{2}\,{}^t\vec{h}\, H(\vec{a})\vec{h} - \|\vec{h}\|^2 |E_2(\vec{a};\vec{h})| > 0 \quad \text{for } \vec{h} \neq \vec{0},$$

and f has a relative minimum at $\vec{a}$. □

Lemma on Quadratic Forms: Let $Q(\vec{h}) := {}^t\vec{h}\, A\, \vec{h}$, A symmetric $n \times n$ with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Then

(i) $Q(\vec{h}) > 0 \ \forall \vec{h} \neq \vec{0} \iff$ all $\lambda_j > 0 \iff A$ pos.-definite

(ii) $Q(\vec{h}) < 0 \ \forall \vec{h} \neq \vec{0} \iff$ all $\lambda_j < 0 \iff A$ neg.-definite

(iii) $Q(\vec{h})$ takes +/− values $\iff \lambda_1 < 0 < \lambda_n \iff A$ indefinite.

(This is just the Corollary to the Spectral Theorem, better stated.)

Theorem 2: If f is twice continuously diff'able, with a stationary point at $\vec{x} = \vec{a}$, then:

(i) $H(\vec{a})$ pos.-definite $\Rightarrow f$ has rel. minimum at $\vec{a}$

(ii) $H(\vec{a})$ neg.-definite $\Rightarrow f$ has rel. maximum at $\vec{a}$

(iii) $H(\vec{a})$ indefinite $\Rightarrow f$ has saddle point at $\vec{a}$.

Proof: We record (before erasing) the fact that

$$f(\vec{a}+\vec{h}) - f(\vec{a}) = \vec{\nabla} f(\vec{a}).\vec{h} + {}^t\vec{h}\, H(\vec{a})\vec{h} + \|\vec{h}\|^2 E_2(\vec{a};\vec{h}).$$