MULTIVARIABLE DIFFERENTIAL CALCULUS

Let $c \in \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}^m$

$f(x, y) \quad (m = 2, m - 1)$

steepness at $c$ in the direction of $u$

$c \in S$ - interior

$\Rightarrow \exists \, B(c) \subset S$

DEF: THE DIRECTIONAL

derivative of $f$ at $c$ in the direction $u$, $f'(c; u)$, is

$$f'(c, u) = \lim_{h \to 0} \frac{f(c + hu) - f(c)}{h}$$

(sometimes assume $\|u\| = 1$)
\[ u = e_k = (0, 0, \ldots, 0, 1, 0, \ldots, 0) \]

Then
\[ f(c, e_k) = \Delta_k f(c) \]

is a partial derivative.

Usually, we look at \( f : \mathbb{R}^n \to \mathbb{R} \)
\[ f = f(x_1, \ldots, x_n) \]
and partial derivatives \( \partial f / \partial x_k \).

If \( f = (f_1, \ldots, f_m) \), then
\[ f(c, v) \text{ exists if } f_k(c, v) \text{ exist for } k = 1, \ldots, m \text{ and then} \]
\[ f(c, v) = (f_1(c, v), \ldots, f_m(c, v)) \]
for \( v \neq v_k \)
\[ \Delta_k f(c) = (\Delta_k f_1(c), \ldots, \Delta_k f_m(c)) \]
1. \( f(t) = f(c+tv) \), then
   \[ f'(t) = f'(c, v) \]
   In general
   \[ f'(t) = f'(c+vt, v) \]

2. \( f(x) = \|x\|^2 \), then
   \[ f(t) = f(c+tv) = (c+tv)\cdot(c+tv) \]
   \[ = \|c\|^2 + 2t \cdot c \cdot v + t^2 \|v\|^2 \]
   \[ \Rightarrow f'(t) = 2c \cdot v + 2t \|v\|^2 \]
   \[ \Rightarrow f'(0) = f'(c, v) = 2c \cdot v \]

3. \( f: \mathbb{R}^m \rightarrow \mathbb{R}^m \) is linear.

   \[ f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \]
   for all \( x, y \in \mathbb{R}^m \), \( \alpha, \beta \in \mathbb{R} \).

   \[ f'(c, v) = f'(v) \]
**Directional Derivatives & Continuity**

If \( f(c, v) \) exists for \( v \), then \( D_v f(c) \) if for \( k \in \mathbb{R} \), \(-j, ... \).

**Converse is not true:**

\[
 f(x, y) = \begin{cases}
 x + y & \text{if } x \leq 0 \text{ or } y \leq 0 \\
 1 & \text{otherwise}
\end{cases}
\]

\[
 \Rightarrow D_1 f(0, 0) = D_2 f(0, 0) = 1
\]

\[
 v = (\alpha_1, \alpha_2) \quad \alpha_1, \alpha_2 \neq 0
\]

\[
 \frac{f(0 + \alpha_2 v) - f(0)}{\alpha_1} = \frac{f(\alpha_2 v)}{\alpha_2} \rightarrow \frac{1}{\alpha_1}
\]

**Limit does not exist**

The function is not continuous at 0 either.
I can have \( f(1, u) \) for every \( u \), but not be continuous at \( u \).

\[
f(x, y) = \begin{cases} 
\frac{xy^2}{x^2 + y^4} & x \neq 0 \\
0 & x = 0
\end{cases}
\]

Let \( u = (a_1, a_2) \)

\[
\Rightarrow \frac{f(0 + h(u)) - f(0)}{h} = \frac{f(ha_1, ha_2)}{h}
\]

\[
= \frac{a_1 a_2^2}{a_1^2 + h^2 a_2^4}
\]

\[
\Rightarrow \frac{1}{f'(0, u)} = \frac{a_2^2}{a_1}, \text{ if } a_1 \neq 0
\]

\[f'(0, u) \neq 0 \text{ for } a_1 = 0
\]

\[
f'(0, u) \neq 0
\]
But take \( x = y^2 \)

\[ \Rightarrow f(y^2, y) = \frac{y^4}{2y^2} = \frac{1}{2} \]

\[ \Rightarrow f \text{ takes on the value } \frac{1}{2} \]

on \( x = y^2 \) except at \( x = y = 0 \) where \( f = 0 \)

\[ \Rightarrow \text{ DISCONTINUITY.} \]

---

**Total Derivative**

**One-dimensional case**

\[ f(x + h + \beta k) = f(x) + f'(x) \beta h + o(h) \]

\[ = f(x) + \beta f'(x) h + \beta f'(x) k + \ldots \]

**Linear Approximation**

\[ h \mapsto f'(x) h - \text{ linear function of } \beta \]
\[ f: \mathbb{S} \rightarrow \mathbb{R}^n, \quad \mathbf{s} \in \mathbb{R}^n \]

**Definition:**

\( f \) is differentiable at \( \mathbf{x} \in \mathbb{S} \) if for sufficiently small \( \mathbf{v} \) (i.e., \( \| \mathbf{v} \| < r \) for some \( r \))

\[ f(\mathbf{x} + \mathbf{v}) = f(\mathbf{x}) + f'(\mathbf{x}) \mathbf{v} + O(\| \mathbf{v} \|^2) \]

for some linear transformation \( f'(\mathbf{x}) \mathbf{v} \); \( f'(\mathbf{x}) \) - total derivative of \( f \) at \( \mathbf{x} \)

\[ f'(\mathbf{x})(\alpha \mathbf{v} + \beta \mathbf{w}) = \alpha f'(\mathbf{x}) \mathbf{v} + \beta f'(\mathbf{x}) \mathbf{w} \]

\( f'(\mathbf{x}) \) is a matrix

---

**Tangent Plane:**

\[ f(\mathbf{x}_1, \mathbf{x}_2) \]

\[ f(\mathbf{x}_1, \mathbf{x}_2) + f'(\mathbf{x}_1, \mathbf{x}_2) \mathbf{v} \]

\[ f(\mathbf{x}_1, \mathbf{x}_2) \]

\[ f(x_1, x_2) \]

\[ \mathbf{x}_2 \]

\[ (x_1, x_2) \]
Aside: symbols $O(h)$ and $o(h)$

(i) \( f(h) = O(h) \) as \( h \to 0 \) if \( |f(h)| < C|h| \), for some constant \( C \), or more precisely, if

\[ f(h) = h \cdot g(h), \]

where \( |g(h)| < c \) and \( \lim_{h \to 0} g(h) \neq 0 \).

(ii) \( f(h) = o(h) \) as \( h \to 0 \)

\[ \frac{f(h)}{h} \to 0. \]
Let $f(x)$ be differentiable at $x$. Then the directional derivative $f'(x, u)$ exists for every $u \in \mathbb{R}^n$, and
\[ f(x + hu) - f(x) \quad \frac{f(x + hu) - f(x)}{h} = f'(x) u + \frac{\sigma(h)}{h} \]
\[ f'(x, u) = \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h} = f'(x) u \]
**Theorem:** If \( f \) is differentiable at \( x \), then \( f \) is continuous at \( x \).

**Proof:** \( f(x + \epsilon) - f(x) = f'(x) \epsilon + o(\|\epsilon\|) \)

\[ = f'(x) (e_1 \epsilon_1 + \cdots + e_n \epsilon_n) + o(\|\epsilon\|) \]

\[ = \epsilon_1 f'(x) e_1 + \cdots + \epsilon_n f'(x) e_n + o(\|\epsilon\|) \]

\[ (e_j = (0, \ldots, 0, 1, 0, \ldots, 0)) \]

As \( \|\epsilon\| \to 0 \), so do \( \epsilon_j \).

\[ \Rightarrow \| f(x + \epsilon) - f(x) \| \to 0 \text{ as } \|\epsilon\| \to 0 \]
The matrix for $f'(x)$:

$\nu = \nu_1 e_1 + \cdots + \nu_m e_n$

$f'(x) \nu = \nu_1 f'(x) e_1 + \cdots + \nu_m f'(x) e_n$

$= \nu_1 f'(x, e_1) + \cdots + \nu_m f'(x, e_n)$

$= \nu_1 D_1 f(x) + \cdots + \nu_m D_m f(x)$

$f = (f_1, \ldots, f_m)$

$D_{e_k} f = (D_{e_k} f_1, \ldots, D_{e_k} f_m)$

$(f'(x) \nu)_k = \nu_1 D_1 f_k(x) + \cdots + \nu_m D_m f_k(x)$

$
\Rightarrow \int f'(x) \, df = D_1 f_1 (x) + \cdots + D_m f_m (x)
$

$\Rightarrow \text{Matrix for } f(x)$

$D f(x) = \begin{bmatrix}
D_1 f_1 (x) & \cdots & D_m f_1 (x) \\
D_1 f_m (x) & \cdots & D_m f_m (x)
\end{bmatrix}$
Alternative notation

\[
\begin{bmatrix}
\frac{df_1}{dx_1} & \frac{df_1}{dx_n} \\
\frac{df_2}{dx_1} & \frac{df_2}{dx_n} \\
\vdots & \vdots \\
\frac{df_m}{dx_1} & \frac{df_m}{dx_n}
\end{bmatrix}
\]

Jacobian matrix

\[ f: \mathbb{R}^n \rightarrow \mathbb{R}^m \]

\[
\det f'(x) = \frac{\partial (f_1, \ldots, f_m)}{\partial (x_1, \ldots, x_m)}
\]

Jacobian determinant
If \( f : S \rightarrow \mathbb{R}^n \), then
\[
\nabla f(x) \cdot \nu = \text{dot product}
\]

\[
\text{with} \quad \nabla f(x) = \left( D_1 f(x), D_2 f(x), \ldots, D_n f(x) \right)
\]

If \( f : S \rightarrow \mathbb{R}^m \), then
\[
f'(x) \nu = \left( \nabla f_1(x) \cdot \nu, \nabla f_2(x) \cdot \nu, \ldots, \nabla f_m(x) \cdot \nu \right)
\]

\[
\| f'(x) \nu \| = \| \sum_{k=1}^{m} (\nabla f_k(x) \cdot \nu) e_k \| \leq \|
\]

\[
\leq \sum_{k=1}^{m} | \nabla f_k(x) \cdot \nu | \| e_k \|
\]

\[
= \sum_{k=1}^{m} | \nabla f_k(x) \cdot \nu | \| e_k \|
\]

\[
\leq \sum_{k=1}^{m} \| \nabla f_k(x) \| \| \nu \|
\]

\[
= \| \nabla \| \frac{m}{2} \| \nabla f_k(x) \|
\]
THE CHAIN RULE

**Theorem**  
If \( g \) is differentiable at \( a \) and \( f \) is differentiable at \( g(a) = b \in \mathbb{R} \), then \( f \circ g \) is differentiable at \( a \) and

\[
(f \circ g)'(a) = f'(g(a)) \cdot g'(a)
\]

**Proof**  
Let \( h = f \circ g \).

\[
h(a + y) - h(a) = \]

\[
= f(g(a + y)) - f(g(a)) =
\]

\[
= f(g(a) + g'(a)y + o(y)) - f(g(a)) =
\]

\[
= f(g(a)) + f'(g(a)) (g'(a)y + o(y)) - f(g(a)) =
\]

\[
= f'(g(a)) (g'(a)y + o(y))
\]
But \( f'(g(a)) \cdot Df(\|y\|) = \sigma(\|y\|) \), so

\[
h'(a + y) - h'(a) =
\]
\[
= f'(g(a)) g'(a) y + o(\|y\|)
\]

\[
\Rightarrow D(h \circ g)(a) = Df(g(a)) \cdot Dg(a)
\]

\[\text{MATRICES}\]

\[\mathbb{R}^k \xrightarrow{g} \mathbb{R}^m \xrightarrow{f} \mathbb{R}^n \]

\[x \xrightarrow{g} y \xrightarrow{f} z\]

\[
\frac{\partial z_i}{\partial x_j} = \sum_{k=1}^{n} \frac{\partial z_i}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_j}
\]
EXAMPLE: \( f: \mathbb{R}^2 \rightarrow \mathbb{R} \), \( h: \mathbb{R}^2 \rightarrow \mathbb{R} \)

\[
\begin{align*}
\frac{\partial h}{\partial x} &= \frac{\partial h}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial h}{\partial v} \frac{\partial v}{\partial x} \\
\frac{\partial h}{\partial y} &= \frac{\partial h}{\partial u} \frac{\partial u}{\partial y} + \frac{\partial h}{\partial v} \frac{\partial v}{\partial y}
\end{align*}
\]

(1)

because

\[
Df = \left[ \begin{array}{cc}
\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\
\frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}
\end{array} \right] \quad Dh = \left( \frac{\partial h}{\partial u}, \frac{\partial h}{\partial v} \right)
\]

Note: Abuse of notation in (1), \( \frac{\partial h}{\partial x} \) is really \( \frac{\partial (h \circ f)}{\partial x} \), etc.
Sufficient condition for differentiability

Recall: If $f$ is differentiable at a point $a$, it is continuous at $a$.

$$ f(x, y) = \begin{cases} \frac{xy^2}{x^2 + y^4} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases} $$

is not continuous, yet it has directional derivatives in every direction.

$\Rightarrow$ existence of directional derivatives of $f(x, y)$ in every direction, still does not imply differentiability.
THM: \[ \text{Let } f : \mathbb{R}^n \to \mathbb{R}^m. \]

Then \( f \) is differentiable at \( a \in \mathbb{R}^n \) if and only if each component function \( D_j f_i \) is continuous at \( a \).

**Proof:** Clearly \( f : \mathbb{R}^n \to \mathbb{R}^m \) is differentiable at \( a \)

if \( f_i \) are differentiable for \( i = 1, \ldots, m \),

(\text{convergence in } \| \cdot \| \text{ is equivalent to convergence in components})
So, let $f: \mathbb{R}^n \to \mathbb{R}$. Then (2)

$$f(a + h) - f(a) =$$

$$= f(a_1 + h_1, a_2 + h_2, \ldots, a_n + h_n) - f(a_1, a_2, \ldots, a_n)$$

$$= f(a_1 + h_1, a_2 + h_2, \ldots, a_{n-1} + h_{n-1}, a_n + h_n) - f(a_1 + h_1, a_2 + h_2, \ldots, a_{n-1} + h_{n-1}, a_n)$$

$$+ f(a_1 + h_1, a_2 + h_2, \ldots, a_{n-1} + h_{n-1}, a_n) - f(a_1 + h_1, a_2 + h_2, \ldots, a_{n-1} + h_{n-1}, a_n)$$

$$+ f(a_1 + h_1, a_2 + h_2, \ldots, a_{n-1} + h_{n-1}, a_n) - f(a_1 + h_1, a_2 + h_2, \ldots, a_{n-1} + h_{n-1}, a_n)$$

$$\Rightarrow \text{the term is by the mean-value theorem}$$

$$h_1 \cdot D_1 f(a_1 + h_1, a_2 + h_2, \ldots, a_{n-1} + h_{n-1}, a_n) =$$

$$= h_1 \cdot D_1 f(c_1)$$

for some $c_1$. 
\[
\lim_{h \to 0} \frac{1}{n} \sum_{i=1}^{n} \left| D_i f(c_i) - D_i f(a) \right|
\leq \lim_{h \to 0} \frac{1}{n} \sum_{i=1}^{n} \left| D_i f(c_i) - D_i f(a) \right|
\leq 0
\]

since \(D_i f\) are continuous at \(a\).
**HIGHER PARTIAL DERIVATIVES**

**DEF:** \[ D_{ij} f(x) = D_i (D_j f(x)) \]

for \( f: \mathbb{R}^n \to \mathbb{R} \).

**Counterexample:**

\[ f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases} \]

\[ D_1 f(x, y) = \begin{cases} \frac{y(x^4+4xy^2-y^4)}{(x^2+y^2)^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases} \]

\[ D_2 f(x, y) = \begin{cases} \frac{x(x^4-4x^2y^2-y^2)}{(x^2+y^2)^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases} \]
\[ D_1 f(0, y) = -y \]
\[ D_{21} f(0, y) = -1, \quad D_{21} f(0, 0) = -1 \]
\[ D_2 f(x, 0) = x \]
\[ D_{12} f(x, 0) = 1, \quad D_{12} f(0, 0) = 1 \]
\[ \Rightarrow D_{12} f(0, 0) \neq D_{21} f(0, 0) \]

But:

This implies that \( D_{12} f(x, y) \) and \( D_{21} f(x, y) \) are not continuous in some open set \( G \subset \mathbb{R}^2 \), then

\[ D_{12} f(x, y) = D_{21} f(x, y) \]

on \( G \).

(Why is \( \mathbb{R}^2 \) enough?)
Proof: Consider the points $(x+yh, y+kh)$.

For $h$ small enough, all these points are in $S_j$ and so is the rectangle whose vertices they are.

Consider

$$A = f(x+yh, y+kh) - f(x+yh, y) - f(x, y+kh) + f(x, y).$$

Let

$$
\gamma(x) = f(x, y+kh) - f(x, y)
$$

Then

$$A = \gamma(x+yh) - \gamma(x) =
\gamma'(x+\theta h) h
= [D_1 f(x+\theta h, y+kh) - D_1 f(x+\theta h, y)] h
= D_2 f(x+\theta h, y+kh, y+\bar{\theta} h) h k.$$
On the other hand, let

\[ y'(y) = f(x + \theta \mathbf{e}, y) - f(x, y) \]

Then

\[ A = y'(y + \mathbf{e}) - y'(y) = \]

\[ = y'(y + \theta \mathbf{e}) \mathbf{e} = \]

\[ = \left[ \nabla_x f(x + \theta \mathbf{e}, y + \theta \mathbf{e}) - \nabla_x f(x, y + \theta \mathbf{e}) \right] \mathbf{e} \]

\[ = \nabla_{\mathbf{x}} f(x + \theta \mathbf{e}, y + \theta \mathbf{e}) \mathbf{e} \mathbf{e} \]

\[ \Rightarrow \nabla_{\mathbf{x}} f(x + \theta \mathbf{e}, y + \theta \mathbf{e}) = \]

\[ = \nabla_{\mathbf{x}} f(x + \theta \mathbf{e}, y + \theta \mathbf{e}) \]

As \( \mathbf{e}, \mathbf{e} \to 0 \)

\[ \nabla_{\mathbf{x}} f(x, y) = \nabla_{\mathbf{x}} f(x, y) \]

by continuity.
TAYLOR'S FORMULA FOR
FUNCTIONS $f: \mathbb{R}^n \to \mathbb{R}$

Recall: if $f'(x)$ exists, then for any vector $t$, the directional derivative $f'(x; t)$ exists and equals

$$f'(x; t) = f'(x) \cdot t =$$

$$= D_1 f(x) t_1 + \cdots + D_n f(x) t_n.$$

Define higher-order directional derivatives:

$$f''(x; t) = \sum_{i=1}^n \sum_{j=1}^n D_{ij} f(x) t_i t_j,$$

$$f'''(x; t) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n D_{ijk} f(x) t_i t_j t_k.$$
\( f^{(n)}(x,t) = \sum_{i=1}^{n-1} \sum_{i=m}^{n} \frac{n-i}{i} D_{i-m} f(x) t_i \cdot \cdot \cdot t_m \)

**Theorem:** Let \( f \) be \( m \) times continuously differentiable on \( B(a,r) \subseteq \mathbb{R}^n \). Then for every point \( \alpha + \theta \beta \in B(a,r) \),

\[
f(\alpha + \theta \beta) = \sum_{k=0}^{m-1} \frac{1}{k!} f^{(k)}(\alpha; \theta \beta) + \frac{1}{m!} f^{(m)}(\alpha + \theta \beta; \theta \beta)
\]

for some \( 0 < \theta < 1 \).

**Proof:** Let \( s \in \mathbb{R} \) and

\[ g(s) = f(\alpha + s \beta) \]

Then

\[ g^{(k)}(s) = f^{(k)}(\alpha + s \beta; \beta) \]

by the chain rule.
To justify Taylor's theorem for functions of one variable with Lagrange's remainder,

\[ g(x) - g(0) = \sum_{k=1}^{m-1} \frac{1}{k!} g^{(k)}(0) + \frac{1}{m!} g^{(m)}(\xi) \]

\[ \Rightarrow f(x+h) - f(x) = \]

\[ = \sum_{k=1}^{m-1} \frac{1}{k!} f^{(k)}(a, h) + \frac{1}{m!} f^{(m)}(x+h, \xi) \]

**Corollary:**

**Another Mean-Value Theorem**

\[ f(x+h) - f(x) = \]

\[ = f'(a+\theta h, h) \]
COROLLARY

If \( f'(x) = 0 \) on an open ball \( c \in \mathbb{R}^n \), then \( f(x) = \text{constant} \).

Proof: \( f'(x, h) = f'(x) \cdot e \Rightarrow 0 \)

Note: A closer look at \( f^{(k)}(a; t) \):

\[
f^{(k)}(a; t) = \sum_{i_1=1}^{m} \cdots \sum_{i_k=1}^{m} D_{i_1 \cdots i_k} f(a; t_{i_1}, \ldots, t_{i_k})
\]

Some of these terms are the same, for instance:

\[
D_{123} f(a; t_1, t_2, t_3) = D_{213} f(a; t_2, t_1, t_3) = D_{321} f(a; t_3, t_2, t_1)
\]
How to find the right const:

\[ f (x; t) = \sum_{\alpha_1, \ldots, \alpha_m} \frac{\alpha_1! \cdots \alpha_m!}{\alpha_1 + \cdots + \alpha_m = k} D_{\alpha_1} \cdots D_{\alpha_m} f (a) t_1 \cdots t_m \]

This formula has the disadvantage of not having a product of \( k \) terms explicitly.

The estimate of the remainder:

\[ |f^{(m)} (x, h)| \leq \sum_{i=0}^{\infty} \sum_{i=0}^{m} \max_{x \in S (a_i)} |D_{i} \cdots D_{i_m} f (x)| \cdot |h_i| \cdot |h_{i_m}| \]

\[ \leq c \cdot M \cdot \| h \|_m^m \]

\[ \frac{|f^{(m)} (x, h)|}{\| h \|_m^{m-1}} = O (\| h \|) \to 0 \]
⇒ TRIVIAL: If \( f \in C^m(\mathbb{B}(a, r)) \)

then

\[
f(a + t h) = \sum_{k=1}^{m-1} \frac{1}{k!} f^{(k)}(a) t^k + O(||h||^m)
\]

E.g., \( V(r, \mathbf{h}) = \frac{1}{||r + \mathbf{h}||} - \frac{1}{||r - \mathbf{h}||} \quad r = (x, y, z) \)

Let \( f(x) = \frac{1}{x^2} \Rightarrow V(r, \mathbf{h}) = f(r + \mathbf{h}) - f(r - \mathbf{h}) \)

\[
= f(r) - f(r) + 2 \nabla f(r) \cdot \mathbf{h} + O(||h||^2)
\]

\[
= -2 \frac{r \cdot \mathbf{h}}{|r|^3} + O(||h||^2)
\]

Diagram: DIPole approximation
UNCONSTRAINED EXTREMA OF
FUNCTIONS ON $\mathbb{R}^n$

**Theorem** Let $f : A \to \mathbb{R}$, $A$ open in $\mathbb{R}^n$. If $a \in A$ is a local extremum of $f$ and $D_i f(a)$ exist, then $D_i f(a) = 0$

**Proof** Let

$$g_i(x) = f(a_1, \ldots, x_i, \ldots, a_n)$$

Then $g_i(x)$ has an extremum at $x = a_i$, so $g_i'(a_i) = D_i f(a) = 0$

**Examples**

1) $f(x,y)$: If there is a max or min at $(x_0, y_0)$, then $D_i f(x_0, y_0) = D_j f(x_0, y_0) = 0$

$\nabla f(x_0, y_0) = (f(x_0, y_0), f(x_0, y_0))$ - Tangent plane is horizontal at $(x_0, y_0, f(x_0, y_0))$
2.) \( f(x,y) = x^2 - y^2 \)

\[ x = 0 : \quad f(10, y) = - y^2 \]
\[ y = 0 : \quad f(x, 0) = x^2 \]
\[ D_1 f(10,0) = D_2 f(10,0) = 0 \]

**Sufficient condition for an extremum**

Let \( f \in C^2(A) \), \( A \subset \mathbb{R}^n \), \( a \in A \)

\[ f'(a) = 0. \]

By Taylor's formula

\[ f(a+h) = f(a) + \frac{1}{2!} f''(a, h) + o(h^2) \]
IDEA: a is a max, min, or saddle if \( f''(a, b) < 0 \), \( > 0 \), or changes sign, respectively, for all \( h \).

We need, and have, a bit more: take a closer look at \( f^n(a, b) \)

\[
f^n(a, b) = \sum_{j=1}^{m} \nabla_{h_j} f(a) \cdot h_i \cdot h_j = \\
= \langle h, f^n(a) h \rangle
\]

where

\[
f^n(a) = \\
= \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2}
\end{bmatrix}
\]
The Hessian, $f''(a)$, is a symmetric matrix

\[ f''(a) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} n_i \cdot n_j \]

$k$ can be diagonalized in an orthonormal basis $\{n_1, \ldots, n_k\}$.

The components of $f''(a)$ in this basis are

\[ f''_{ii} = \sum_{j=1}^{k} \frac{\partial^2 f}{\partial x_i \partial x_j} n_i \cdot n_j \quad \text{and} \quad f''_{ij} = 0 \quad \text{for} \quad i \neq j \]

Now, if $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of $f''(a)$, then

\[ f''(a) \cdot n_i = \sum_{j=1}^{k} \lambda_j n_i \cdot n_j n_j \]
\[ \langle h, f(\alpha) h \rangle = \sum_{j=1}^{n} \frac{1}{\alpha_j} - \langle h, v_j \rangle^2 \]

If all \( \alpha_j > 0 \), \( j = 1, \ldots, n \) let \( \alpha = \min \alpha_j \). Then
\[ \langle h, f(\alpha) h \rangle \geq \alpha \sum_{j=1}^{n} \langle h, v_j \rangle^2 = L \| h \|^2 \]

\( L \) in this case, we say \( f(\alpha) \) is pointwise definite.

\[ \Rightarrow \langle h, f(\alpha) h \rangle \geq \alpha \| h \|^2 > 0 (\| h \|^2) \]

and \( \alpha \) is really a minimum.

Likewise if all \( \alpha_j < 0 \), \( j = 1, \ldots, n \) then
\[ \langle h, f(\alpha) h \rangle \leq -L \| h \|^2 \]
\[ \rho_{x_0} = \min_{j=1, \ldots, n} H_{f_j} = -\max_{j=1, \ldots, n} \left( f_{x_0} \right) \] (positive definite) \[ a \text{ is a minimum} \]

If \( a_j \)'s have mixed signs
and \( \lambda_j \neq 0 \) \( j = 1, \ldots, n \),
a \( \text{is a saddle} \).

\[ \text{Hence let } f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}, \quad a \in A, \quad f'(a) = 0 \]

(i) If \( f''(a) \) is positive (negative) definite, \( a \) is a minimum (maximum).

(ii) If \( f''(a) \) has non-zero eigenvalues if mixed
\( f_{x_0} \), \( a \) is a saddle.
Example

\[ f(x, y) = ax^2 + by^2 \]

\[ D_1 f(x, y) = 2f(x, y) = 0 \]

\[ f''(x, y) = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} = f''(0, 0) \]

1.) \( a, b > 0 \)

\[ z = f(x, y) \]

minimum

2.) \( a, b < 0 \)

\[ z = f(x, y) \]

maximum
3. $a > 0$ $b < 0$

4. $a = 0$ $b = 1$

$f(x, y) = \frac{y^2}{x}$

"Teough"
Special case: \( m = 2 \)

\[
\dot{f}''(a) = \begin{bmatrix}
\frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\
\frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2}
\end{bmatrix}
\]

\[
\det \left[ \dot{f}''(a) - \lambda I \right] =
\]

\[
= \lambda^2 - \nabla^2 f \lambda + \det \dot{f}''(a) \equiv P(\lambda)
\]

\[
\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}
\]

\[
\det \dot{f}''(a) = \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2 - \left( \frac{\partial^2 f}{\partial x^2} \right) \left( \frac{\partial^2 f}{\partial y^2} \right)
\]

Since \( P(\lambda) = \lambda^2 - (\lambda_1 + \lambda_2) \lambda + \lambda_1 \lambda_2 \)

\[
\Rightarrow \lambda_1 \text{ and } \lambda_2 \text{ have the same sign if } \det \dot{f}''(a) > 0
\]

\[
\Rightarrow \text{ maximum or minimum}
\]

\[
\min \Rightarrow \nabla^2 f(a) < 0
\]

\[
\max \Rightarrow \nabla^2 f(a) > 0
\]
Saddle if \( \det f''(a) < 0 \)

Alternatively, if \( \det f''(a) > 0 \)
we must have that
\[
\frac{\partial^2 f}{\partial x^2} \quad \text{and} \quad \frac{\partial^2 f}{\partial y^2}
\]
have the same sign.

\[\Rightarrow\] enough just to check
the sign of one, say
\[
\frac{\partial^2 f}{\partial x^2},
\]
to determine max or min.
**Implicit Functions**

**Examples**

1.) $x^2 + y^2 - \alpha = 0$

- $\alpha > 0$:
  - A circle of solutions

- $\alpha = 0$:
  - A single point solution

- $\alpha < 0$:
  - No solution

2.) $xy = 0$

- Two solutions passing through $(0, 0)$
**Closer Look**

**Example 1**.

If \( y \neq 0 \), we can express \( y \) in terms of \( x \) as

\[
y = \pm \sqrt{a - x^2}
\]

**Note:** \( \frac{d}{dy} (x^2 + y^2 - a) = 2y \neq 0 \) there and at \( y = 0 \), \( \frac{d}{dy} (x^2 + y^2 - a) = 0 \)

\( a < 0 \): Since the equation is non-linear, it may not have a solution.

\( a = 0 \) and **Example 2**:

**Note that** \( \nabla (x^2 + y^2) = (2x, 2y) = \nabla (xy) \) at \( x = y = 0 \) \( \Rightarrow \) the tangent plane is horizontal in both cases.

\( \Rightarrow \) only a single point solution

or multiple solutions?
General problem: m-equations

for m-unknowns \((y_1, \ldots, y_m) = g\)

depending on m-variables \((x_1, \ldots, x_n) = x\)

\[ f(x, y) = 0 \quad \forall \in \mathbb{R}^m \to \mathbb{R}^m \]

Because the problem is nonlinear, we must have some particular solution already, say \((x_0, y_0)\), i.e., \(F(x_0, y_0) = 0\).

We are looking for a function \(y = y(x)\), \(y(x) = \tilde{y}\) such that

\[ F(x, y(x)) = 0 \quad \text{for all } x \text{ near } x_0. \]

Linear approximation:

\[
D_x F(x, \tilde{y}) (x - x_0) + D_y F(x, \tilde{y}) (y - \tilde{y}) = 0 \\
\text{maximize}
\]

If \([D_y F(x, \tilde{y})]^{-1}\) exists

\[
y = \tilde{y} - [D_y F(x, \tilde{y})]^{-1} D_x F(x, \tilde{y}) (x - x_0)
\]
If \( \tilde{y} \neq 0 \Rightarrow D_y F(x, \tilde{y}) = 2\tilde{y} \neq 0 \)

Linear approximation to \( y(x) \) is
\[
y = \tilde{y} - \frac{\tilde{y}}{y}(x - \tilde{x})
\]

If \( \tilde{y} = 0 \), there is a vertical tangent and \( y(x) \) cannot be approximated linearly.

In fact, two branches
\[
y = \pm \sqrt{1-x^2}
\]
merge near \( \tilde{y} \to 0, \quad x = \pm 1 \).
The linear approximation indicates the general case:

**Theorem (Implicit Function)** Let \( F(x, y) \) be a \( C^1 \) function defined in a neighborhood of \( x \in \mathbb{R}^n \) and \( y \in \mathbb{R}^m \), taking values in \( \mathbb{R}^m \), on the \( F(x, y) = c \). Then if \( D_y F(x, y) \) is invertible there exists a neighborhood \( U \) of \( x \) and a \( C^1 \) function \( y : V \to \mathbb{R}^m \) such that \( F(x, y(x)) = c \) for every \( x \in U \). Furthermore, \( y \) is unique in that there exists a neighborhood \( V \) if \( y \) \((V = y(U)) \) such that there is only one function \( z \) in \( V \) if \( F(x, z) = c \), namely \( z = y(x) \). Finally, the derivative of \( y \) can be computed by implicit differentiation as

\[
y'(x) = -\left[ \frac{\partial F(x, y(x))}{\partial y} \right]^{-1} \frac{\partial F(x, y(x))}{\partial x}
\]
Preliminaries: Matrix norms: Def. If \( M \) is a matrix,

\[
\|M\| = \sup_{\|x\| = 1} \frac{\|Mx\|}{\|x\|} = \text{Samp} \|Mx\|_{\|x\| = 1}
\]

The last expression is the sup of a continuous function on a compact set \( \Rightarrow M \) exists.

(Warning: \( \|M\| \) is not \( M_{ij} \) on \( \mathbb{R}^{m \times n} \)).

**Prop 1:** \( \|Mx\| \leq \|M\| \|x\| \) (From definition)

**Prop 2:** \( M_{ij} \leq \|M\| \), \( \|M\|^2 = \sum_{ij} M_{ij}^2 \)

**Proof:** Let \( e_i = (0, \ldots, e_i^j, \ldots, 0) \), \( \|e_i\| = 1 \)

\[
\Rightarrow (Me_i)_j = M_{ij}
\]

\[
\Rightarrow |M_{ij}| = |(Me_i)_j| \leq \|Me_i\| \leq \|M\| \|e_i\| = \|M\|
\]

\[
\leq \|M\| \|e_i\| = \|M\|
\]
On the other hand,

\[ \|M\|_2^2 = \sum_{j,k} \left( \sum_{i=1}^{n} M_{ij} x_i \right)^2 \leq \sum_{j,k} \left( \sum_{i=1}^{n} \left( \sum_{i=1}^{n} M_{ij} x_i \right)^2 \right) \]

\[ \leq \left( \sum_{j,k} \|M_{jk}\| \right) \|x\|_2^2 \]

Since \( \|M\|_2 = \sup \frac{\|Mx\|}{\|x\|} \),

\[ \|M\|_2^2 \leq \sum_{j,k} \|M_{jk}\| \]

---

Proof:

\[ \|AB\| \leq \|A\| \|B\| \]

Proof:

\[ \|ABx\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| \|x\| \]

**Lemma 1**: If \( \|B\| \|A^{-1}\| < 1 \), then \( A + B \) is invertible.

**Proof**: Write a "perturbation" series for \((A + B)^{-1}:\)

\[ A + B = (I + BA^{-1}) A \]

\[ \Rightarrow (A + B)^{-1} = A^{-1} (I + BA^{-1})^{-1} \]
\[ = A^{-1} \sum_{k=0}^{\infty} (-1)^k (BA^{-1})^k \]

Now
\[ \| S_m - S_n \| = \| \sum_{k=m+1}^{\infty} (-1)^k (BA^{-1})^k \| \leq \]
\[ \leq \| A^{-1} \| \sum_{k=m}^{\infty} \left( \| B \| \| A^{-1} \| \right)^k \leq \]
\[ \leq \| A^{-1} \| \left( \| B \| \| A^{-1} \| \right)^m \frac{1}{1 - \| B \| \| A^{-1} \|} \]

\[ \Rightarrow \| B \| \| A^{-1} \| < 1. \]

By Prop 2, a matrix series converges component-wise if and only if it converges in the norm.

\[ \Rightarrow \text{even when the matrix norm is complete and the series for } (A + B)^{-1} \text{ converges if } \| B \| \| A^{-1} \| < 1. \]
\[(A + B)^{-1} - A^{-1} = \sum_{k=1}^{\infty} (-1)^k \left( B A^{-1} \right)^k \]

which again converges if \( \|B\| \|A^{-1}\| < 1 \).

\[\text{Lemma } \quad \| (A + B)^{-1} - A^{-1} \| \to 0 \quad \text{if } \|B\| \to 0 \]

\[\text{Proof: } \quad \| (A + B)^{-1} - A^{-1} \| \leq \frac{\|B\| \|A^{-1}\|}{1 - \|B\| \|A^{-1}\|} \to 0 \]

\[\Rightarrow \quad \text{Conclusion: } A^{-1} \text{ is a continuous function of the entries of } A \]

\[\text{Newton's method: } \quad \text{Find a zero of a function } f(x) \]

\[\text{Graph: } y = f(x) \quad \text{with iterations } x_0, x_1, x_2, \ldots, x_n \]
Approximate \( f(x) \) by its tangent at each approximate zero:

\[
f(x) = f'(x_n)(x - x_n) + f(x_n)
\]

\[
\Rightarrow \text{ since } f(x_{n+1}) < 0,
\]

\[
f'(x_n)(x_{n+1} - x_n) + f(x_n) = 0
\]

\[
x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}
\]

This may work if \( f'(x_n) \neq 0 \)

A simpler, but "slower" scheme:

Have all "tangents" leave slope \( f'(x_0) \)
\[ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \]

This is the scheme we will use in the proof.

**Proof of the implicit function theorem:**

For simplicity, let \( D_x F = F_x \)
\[ D_y F = F_y \]

We are looking for a zero \((m, y)\) of the function \( F(x, y) - c \).

Use the above scheme: iterate the mapping.
\[ Ty = y + \left[ F_y(x, y) \right]^{-1} (c - F(x, y)) \]

Show there exists a neighborhood \( V \) of \( y \), say \( V = \{ y \mid \|y - \hat{y}\| < \delta \} \),

such that:

(i) \( T : V \to V \)

(ii) \( T \) is a contraction on \( V \)

\( \Rightarrow T \) has a fixed point in \( V \).

This will hold, we expect, if \( x + U \)

\[ U = \left\{ \frac{3}{2} \parallel x - \hat{x} \parallel < \varepsilon \right\} . \]
Show $T$ is a contraction:

$$T_y - T_z = y - z + \frac{1}{f'(x,y)} \left[ F(x,z) - F(x,y) \right]$$

$$= f_y(x,y) \left[ F(x,z) - F(x,y) - f_y(x,y)(z - y) \right]$$

The line segment between $y$ and $z$:

$$y + t(z - y), \quad 0 \leq t \leq 1$$

Consider

$$h(t) = F(x, y + t(z - y)), \quad (x = \text{const})$$

$$\Rightarrow F(x,z) - F(x,y) = h(1) - h(0)$$

$$= \int_0^1 h'(t) \, dt =$$

$$= \int_0^1 f_y(x, y + t(z - y)) \cdot (z - y) \, dt$$
\[ F(x_2) - F(x,y) = F_y(x, \bar{y})(2 - y) \]

\[ \frac{1}{b} \int_0^b \left[ F_y(x, y + t(2 - y)) - F_y(x, \bar{y}) \right] (2 - y) \, dt \]

\[ (b/0 \quad \frac{1}{b} \int_0^b \, dt = 1) \]

Since \( F_y \) is continuous,
\[
\| x - x_1 \| < \varepsilon, \quad \| y - \bar{y} \| < \delta
\]
\[
\Rightarrow \| y + t(2 - y) - \bar{y} \| \leq \| y - \bar{y} \| + t \| 2 - \bar{y} \| < 3\delta
\]

Then
\[
\| F_y(x, y + t(2 - y)) - F_y(x, \bar{y}) \| \leq 1
\]
given any \( \varepsilon \). (If \( \varepsilon, \delta \) are small enough.)
\[ \Rightarrow \| F_y(x, y + t(z - y)) - F_y(x, y) \| \leq \lambda \| z - y \| \]

From (*)

\[ \| F(x, z) - F(x, y) - F_y(x, y)(z - y) \| \]

\[ \leq \frac{1}{\lambda} \| z - y \| \leq \lambda \| z - y \| \]

\[ \Rightarrow \| T_y - Tz \| = \]

\[ = \| [F_y(x, y)]^{-1} [F(x, z) - F(x, y) - F_y(x, y)(z - y)] \| \]

\[ \leq M \lambda \| z - y \| \]

\[ M = \| [F_y(x, y)]^{-1} \| \]

Choose \( \lambda \) so that \( M \lambda = \rho < 1 \)

(i.e., choose \( \delta \) small)

\[ \Rightarrow \| T_y - Tz \| \leq \rho \| y - z \|, \quad \rho < 1 \]

\[ \Rightarrow T \text{ is a contraction on } \{ y : \| y - y' \| \leq \delta \} \]
Show $T: V \to V$, i.e. $\nabla T$

is small enough $\|y - \tilde{y}\| \leq \delta$

implies $\|Ty - \tilde{y}\| < \delta$.

$$Ty - \tilde{y} = y - \tilde{y} + [F_x(x, \tilde{y})]^{-1} (c - F(xy))$$

$$= [F_x(x, \tilde{y})]^{-1} \left[ F(xy) - F(x, \tilde{y}) + F_x(x, \tilde{y})(\tilde{y} - y) \right] \tag{1}$$

$\therefore c = F(xy)$

$$[1] = [F(x, \tilde{y}) - F(xy) + F_x(x, \tilde{y})(x - x)]$$

$$+ F_y(x, \tilde{y})(y - \tilde{y}) - F_x(x, \tilde{y})(x - x)]$$

Since $T$ is differentiable at $(x, \tilde{y})$

$$\|F(x, \tilde{y}) - F(xy) + F_x(x, \tilde{y})(x - x) + F_y(x, \tilde{y})(y - \tilde{y})\| \leq \lambda (\varepsilon + \tilde{\varepsilon})$$

for any $\lambda$ of $\varepsilon$ and $\tilde{\varepsilon}$ are small enough.
Also \( \| F(x, y) \| (x - x_0) \| \leq K \varepsilon \), \( K = \| F(x, y) \| \)

\[
\Rightarrow \| T_y - y_0 \| \leq M \| F(x, y) - F(x, y) - F_y(x, y)(y - y_0) \|
\leq M (\lambda (\delta + \delta) + K \varepsilon)
\]

\[
M = \| [F_y(x, y)]^{-1} \|
\]

\[
\Rightarrow \| T_y - y_0 \| < \delta \iff \lambda < \frac{1}{2M}
\]

fix the required \( \delta \), and then let \( \varepsilon < \frac{\delta}{2M(x + K)} \).

\[\Rightarrow \text{The unique fixed point } y(x) \neq T\]

lies in \( V = \{ \| y - y_0 \| < \delta \} \) and \( F(x, y_0) = 0 \)

Also \( y(x) \) is continuous in \( x \). We can relocate \( y_0 = T \) and then \( y(x) = \lim y_n. \) The convergence

\( \rightarrow \) uniform in \( x \), \( \| x - x_0 \| < \delta \)

because the contraction constant, \( \rho \), is independent of \( x \).
Show: \( y(x) \) is differentiable

\[
F(x, y) = c, \quad F(x, y(x)) = c
\]

\[
F(x, y(x)) - F(x, y) = 0
\]

Since \( F \) is differentiable at \( (x, y) \)

\[

\begin{align*}
0 &= F(x, y(x)) - F(x, y) \\
&= F_x(x, y) (x - x) + F_y(x, y) (y(x) - y) \\
&\quad + R(x)
\end{align*}
\]

\[
R(x) = o \left( \| x - x \|^2 + \| y(x) - y \|^2 \right)
\]

\[
\Rightarrow \quad y(x) - y = - \left[ F_y(x, y(x)) \right]^{-1} F_x(x, y) (x - x) \\
- \left[ F_y(x, y(x)) \right]^{-1} R(x)
\]

(II)
\[ y'(x) \text{ is differentiable at } x \]
\[ y'(x) = -\left( F_y(x, y(x))\right)^{-1} F_x(x, y(x)) \]

provided
\[ \left( F_y(x, y(x))\right)^{-1} R(x) = \mathcal{O}(\|x - x^*\|) \]

First show: \[ \|y(x) - y^*\| \leq a\|x - x^*\| \]
for some constant \( a \) and \( x \) near \( x^* \).

From (10)
\[ \|y(x) - y^*\| \leq \|F_y(x, y(x))\| \|F_x(x, y(x))(x - x^*)\| + \|F_y(x, y(x))^{-1} R(x)\| \]
\[ \leq M \|x - x^*\| + M \|R(x)\| \]
\[ M = \|F_y(x, y(x))\|, \quad K = \|F_x(x, y(x))\| \]

\[ 10 \]
Now, since \( y(x) \) is continuous in \( x \), if \( x \) is close to \( x^* \), \( y(x) \) will be close to \( y^* \) and we can make

\[
\| R(x) \| \leq \frac{1}{2M} \left( \| x - x^* \| + \| y(x) - y^* \| \right)
\]

\[
\Rightarrow \| y(x) - y^* \| \leq M \| x - x^* \| + \frac{1}{2} \left( \| x - x^* \| + \| y(x) - y^* \| \right)
\]

\[
\Rightarrow \| y(x) - y^* \| \leq \left( 2Mk + 1 \right) \| x - x^* \|
\]

Now, since \( \| R(x) \| = \eta \left( \| x - x^* \| + \| y(x) - y^* \| \right) \)

if \( \| x - x^* \| \) and \( \| y(x) - y^* \| \) are small enough

\[
\| R(x) \| \leq \eta \left( \| x - x^* \| + \| y(x) - y^* \| \right)
\]

for any small \( \eta \).
\[
\begin{align*}
\left\| \frac{1}{(F_y(x,y))^{-1}} R(x) \right\| &\leq \\
&\leq M \left\| R(x) \right\| \leq \eta \left( \| x - x_0 \| + \| y(x) - y_0 \| \right) \\
&\leq M \eta \left( \| x - x_0 \| + (Mk+1) \| x - x_0 \| \right) \\
&\leq 2M(Mk+1) \eta \| x - x_0 \| = \eta \| x - x_0 \| \\
\Rightarrow \quad y'(x) &= -\left[ F_y(x,y) \right]^{-1} \dot{x}(x,y)
\end{align*}
\]

Now show that \( y(x) \in C' \) by Lemma 2, \( \left[ F_y(x,y(x)) \right]^{-1} \) exists for \( x \) close to \( x_0 \), since \( F_y(x,y) \) and \( y(x) \) are continuous.

Repeat the argument from \( x = x_0 \).
at any nearby \( x \):

\[
y'(x) = -\left[ F_y(x, y(x)) \right]^{-1} F_x(x, y(x)) \quad (\Delta)
\]

since \( y(x) \) is continuous, and \( \Delta y \) and \( \Delta x \), then so must be \( y'(x) \). \( \Rightarrow y' \in C' \).

**Corollary** If \( F \in C^k \), \( y(x) \in C^k \).

**Proof:** (\( \Delta \)) and bootstrapping.

**Special case:** let \( F(x, y) = f(y) - x \) and \( c = 0 \), with \( x, y \in \mathbb{R}^n \)

\( \Rightarrow y(x) \) solves \( f(y(x)) = x \)

and we get the
Inverse function THEOREM: let $f$ be a $C^1$ function defined in a neighborhood of $y$ in $\mathbb{R}^n$ taking values in $\mathbb{R}^n$. If $Df(y)$ is invertible, then there exists a neighborhood $U$ of $x = f(y)$ and a $C^1$ function $g : U \to \mathbb{R}^n$ such that $f(g(x)) = x$ for every $x$ in $U$. Furthermore, $g$ maps $U$ one-to-one onto a neighborhood $V$ of $\tilde{y}$ and $g(f(x)) = y$. The function $g$ is unique in that for any $x \in U$ there is only one $z \in V$ with $f(z) = x$, namely $z = f(x)$. Finally $Dg(x) = [Df(g(x))]^{-1}$. 
Counterexample: \( f(x, y) = (x^2 - y^2, 2xy) \)

\[
Df(x, y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}
\]

\[
\det Df(x, y) = 4(x^2 - y^2) \]

\( \Rightarrow Df(x, y) \) is invertible \( \forall (x, y) \neq (0, 0) \)

\( \Rightarrow \) Locally, there is a unique inverse, but not globally:

\[
f(-x, -y) = f(x, y)
\]

Geometrically:

\[
(x + iy) = r e^{i\theta}
\]

\[
\rightarrow r e^{2i\theta} = (x^2 - y^2 + 2ixy)
\]
REPRESENTATION OF CURVED SURFACES

1) Parametric: \( g: U \rightarrow \mathbb{R}^n \)
   - Given \( u \in U \subset \mathbb{R}^m \), \( n < m \)
   - \( A = g(U) \)
   - Surface

\((t, \ldots, t_n) = t \in U \rightarrow \) curvilinear coordinates on \( A \)

Often \( A = \{ y \in \mathbb{R}^n : f(y) = 0 \} \)

2) Implicit: \( F: \mathbb{R}^n \rightarrow \mathbb{R}^k \)

\( A = \{ x \in \mathbb{R}^n : F(x) = 0 \} \)

\( k = \text{codimension of } A \)

\( m = n - k = \text{dimension of } A \)
2) Explicit (on a graph)

\((x_1, \ldots, x_n) = (t_1, \ldots, t_m, 0, \ldots, 0)\)

with \(t = x\). For some \(f: \mathbb{R}^m \rightarrow \mathbb{R}^n\)

\[ A = \frac{1}{2} (t_1 e_1 + \mathbb{R}^m) \]

\[ \mathbf{s} = f(1 + \frac{1}{2}) \]

\[ t = (x, y) \]

\[ z = 2 \]

**Tangent space**

Two descriptions

1) All vectors \(v\) in \(\mathbb{R}^n\) (starting at \(0 + x\))
   in the directions tangent to \(A\).

2) The plane, but translated to \(x + \frac{1}{2}\)
We will use mostly $1)$, but they are equivalent.

Examples

$1)$ $S^1 \subset \mathbb{R}^2$: $x^2 + y^2 = 1$

Parametric description: $g(t) = (\cos t, \sin t)$

Boundary: $g: [0, 2\pi) \rightarrow S^1$ leaves $(0, 0)$ out.

Orientation: two switches:

$g: [0, 2\pi) \rightarrow \mathbb{R}^2$ and $g: [-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}^2$

$g$ - one-to-one on each.

Graphed: $x = \pm \sqrt{1 - y^2}$ or $y = \pm \sqrt{1 - x^2}$

$2)$ $y^2 - x^2 = 0$ - implicitly

$g(t) = (t, t^2)$ - parametrically

$y = x^{3/2}$ - explicitly (as a graph)
In class today, we touched upon implicit and parametric representations are smooth.

**Explicit** \( f(x) = x^{\frac{2}{3}} + 1(0) = \text{non-existent} \)

3) \( \mathbb{R}^2 \subset \mathbb{R}^3 \) 

**Implicit** \( x^2 + y^2 + z^2 = 1 \) 

**Explicit** 6 of spheres: \( x = \pm \sqrt{1-y^2-z^2} \), \( y = \pm \sqrt{1-x^2-z^2} \), \( z = \pm \sqrt{1-x^2-y^2} \)

Parametric \((x, y, z) = (\sin \phi \cos \theta, \sin \phi \sin \theta, \cos \phi)\) 

(just like \( x^2 + y^2 = 1 \) \( \phi = \pm \pi, \theta = 0 \))

4) \( \mathbb{S}^1 \subset \mathbb{R}^3 \)

\[ \{x^2 + y^2 + z^2 = 1 \} \cap \{ z = 0 \} \]

\( f(x, y, z) = 0 \)

\( f(x, y, z) = (x^2 + y^2 + z^2 - 1, z) \)
Given:
\[ g(t) = (t^3, t^2) \]

Parameter:
\[ g(t) = (\cos \theta, \sin \theta, 0) \]

Explicit:
\[ (y, z) = (\pm \sqrt{1-x^2}, 0) \]
\[ (x, z) = (\pm \sqrt{1-y^2}, 0) \]

Not possible to express \( z \) as a graph over \( x \).

5. \( xy < 0 \Rightarrow x < 0 \) or \( y < 0 \)

Near 0: no 1-1

Parametric representation (or explicit) of \( xy = 0 \)

But: perfectly good representations of \( x = 0 \) and \( y = 0 \).

Back to:
\[ g(t) = (t^3, t^2) \Leftrightarrow y^2 - x^3 = 0 \]

At the cusp, the circle tangent (velocity) vanishes.
(Necessary, but not sufficient; \[ g'(1) = (3t^2, 2t) \]
Extend \( g(t,s) = (t^3 + t^2, s) \)

\[
\begin{pmatrix}
3t^2 & 0 \\
2t & 0 \\
0 & 1
\end{pmatrix}
\]

At \( t = 0 \), \( g'(0,s) = (0,0,0) \)

\( g'(0,s) \) does not have full rank.

Aside: Let \( A \in \mathbb{R}^{m \times n} \)

\[ \text{rank}(A) = \text{dim} \left( \text{image}(A) \right) = \]

\# of linearly independent rows or columns of \( A \)

= size of the largest \( k \times k \) submatrix in \( A \) with \( \det \neq 0 \) =

\[ = m - \dim \left( \text{kernel}(A) \right) \]

\[ \text{kernel}(A) = \mathbb{R}^m, \quad Ax = \mathbb{0} \]
Consider $g : U \to \mathbb{R}^n$, $U$ open in $\mathbb{R}^m$, $g \in C^1(U)$, $n \geq m$.

**Def:** $g$ is an **immersion** if

\[
\text{rank } (g'(x)) = m \quad \text{for} \quad x \in U
\]

**Def:** $g$ is an **embedding** if

- $g$ is an immersion and
- $g : U \to \mathbb{R}^m$ is 1-1
- $g : g(U) \to \mathbb{R}^m$ is continuous

*on $g(U)$ as a metric subspace of $\mathbb{R}^n$*

**Immersions that are not embeddings**

\[
\begin{array}{c}
g \underset{\text{not} \ 1-1} \text{at the crossing}
\end{array}
\]

\[
\begin{array}{c}
U
\end{array}
\]
If it is not continuous at the point of touching (since \( f \) is defined on an open interval, the point is only covered once.)

**Neighborhood of the touching point in the metric of \( g(u) \):**

However:

- It is an embedding because the touching point is not in \( g(u) \), so small enough neighborhoods are all one to one.
Let \( g : U \rightarrow \mathbb{R}^m \) be an embedding.

Carry the usual coordinates on \( U \subset \mathbb{R}^m \) by \( g \) onto a set of curvilinear coordinates on \( g(U) \).

Fix \( x \in U \) \( \Rightarrow \) \( g(x) = y \in g(U) \)

Fix \( x_2, \ldots, x_m \)-coordinates of \( \mathbb{R}^m \), vary \( x_1 \) \( \Rightarrow \) tangent line on \( U \)

\[ f(t) = g(t, x_2, \ldots, x_m) \]

\[ \frac{df}{dt}(x) = \frac{df}{dx_1}(x) \neq 0 \]

Tangent vector at \( y \):

\[ \frac{df}{dt} \] of \( t \) is a column of \( f'(x) \).
Vary every $x_k$, $k = 1, \ldots, m$.

\[ \begin{array}{c}
\begin{array}{c}
\text{\textbullet} \\
\text{\textbullet}
\end{array}
\end{array} \]

\[ \begin{array}{c}
\begin{array}{c}
\text{\textbullet} \\
\text{\textbullet}
\end{array}
\end{array} \]

For curves $\mathbf{y}(\mathbf{v})$ with tangent vectors at $\mathbf{y}$ that are column vectors $\mathbf{y}'(\mathbf{v})$, they are linearly independent (rank $p'(x) = m$).

They span an $m$-dimensional fullspace of $\mathbb{R}^m$ (image $\mathbf{y}'(\mathbf{v})$).

They are the tangent space of $\mathbf{y}(\mathbf{v})$ at $\mathbf{y}$.

They are $T_{\mathbf{y}} \mathbf{y}(\mathbf{v})$.

Parametric representation of $T_{\mathbf{y}} \mathbf{y}(\mathbf{v})$:

\[ T_{\mathbf{y}} \mathbf{y}(\mathbf{v}) = \{ \sum_{k=1}^{m} \mathbf{x}_k (\mathbf{y}'(\mathbf{v}))_k c_k = \mathbf{g}(\mathbf{v}) | c = (c_1, \ldots, c_m) \in \mathbb{R}^m \} \]
Let \( h : (a,b) \to U \) be a \( C^1 \) curve

\[ h(c) = x \]

\[ g \circ h : (a,b) \to f(U) \] is a \( C^1 \) curve in \( f(U) \) through \( y \).

\[ g \circ h(c) = y \]

Its tangent vector

\[ \frac{d}{dt} (g \circ h)(c) = \sum_{k=1}^{m} \frac{\partial g^k}{\partial x^k} (y) \frac{dh^k}{dt}(c) = \]

\[ = g'(y) \cdot h'(c) \]

\[ \in T_{g(y)} f(U) \]

\[ \Rightarrow \text{Prop } T_{g(y)} f(U) = \{ \text{tangent vectors to curves in } f(U) \text{ through } y \} \]

Also: \( g : f(U) \to \mathbb{R}^m \) is continuous \( \Rightarrow \) any curve on \( f(U) \) through \( y \) is in the image of a curve in \( U \) through \( x \).
PARAMETRIC DESCRIPTIONS OF SURFACES

Def. A $C^1$ m-dim surface (parametrix) in $\mathbb{R}^n$ (m $\leq$ n) is $M_{c} \subset \mathbb{R}^n$ such that for every point $y$ in $M_{c}$ there exists a neighborhood $V$ of $y$ in $\mathbb{R}^n$ and an embedding $g : V \rightarrow \mathbb{R}^n$ such that $g(V) = V \cap M_{c}$. Each embedding is $C^1$.

THM. Let $g : V \rightarrow \mathbb{R}^m$, $V$ open in $\mathbb{R}^m$ be a $C^1$ immersion. Then for every point $x \in V$ there exists a neighborhood $U$ of $x$ such that $g$ is 1-1 on $U$ and $g(U)$ is the graph of a $C^1$ function.
Remark: If \( g \) is an embedding
\[ g(\mathcal{U}) \text{ is a neighborhood of } g = g(x) \text{ in } g(\mathcal{U}) \text{ and the result is local on } \mathcal{U}. \]
If \( g \) is just an immersion, the result is local in \( \mathcal{U} \).

\[ g'(x) = \sqrt{\left( \frac{\partial g_i}{\partial x_j} \right)^2}, \quad m \leq m \rightarrow m \]

By assumption, rank \( (g'(x)) = m \)
\[ \Rightarrow g'(x) \text{ has } m \text{ linearly independent rows (and of } m). \]

(After possibly renaming true y's for \( y \in \mathbb{R}^m \)), we can choose the first \( m \) rows to be independent.

\[ \begin{pmatrix} v^T & c_{1:m} & \Rightarrow & g(\mathcal{U}) c_{1:m} \\ x = (x_1, \ldots, x_m) & g & y = (y_1, \ldots, y_m) \end{pmatrix} \]
Write
\[(y_1, \ldots, y_m) \equiv (t_1, \ldots, t_m, s_1, \ldots, s_{m-n})\]
Let \( h = (g_1, \ldots, g_m)^T \).
\[\Rightarrow h^T x_j \text{ is the first m rows of } f^T x_j\]
\[\Rightarrow [h^T x_j]^T \text{ exists}\]
\[\Rightarrow \text{By inverse function theorem}\]
There is an open neighborhood \( \hat{U} \) of \( x_j \) s.t. \( h \) has a \( C^1 \)
inverse \( h^{-1} : \hat{V} \to \hat{U} \), \((\hat{V} = \text{open net in } t\)-space)\)
\[x = h^{-1}(t) \quad \text{iff} \quad t = h(x)\]
\[\Rightarrow g \text{ is 1-1 on } \hat{V} .\]
\[\Rightarrow \varphi = (g_m, \ldots, g_n)^T \Rightarrow s = \varphi(x)\]
\[\Rightarrow s = \varphi(h^{-1}(t)) = f(t), \quad t = \varphi \circ h^{-1}\]
\[y = f(x) \text{ for } x \in \hat{V}, \quad \text{iff} \quad y = (t, s), \; t \in \hat{V}, \; s = f(t)\]
\[\Rightarrow \varphi(\hat{V}) \text{ is a graph of } f .\]
**EXAMPLES**

1. \( g(t) = (\cos t, \sin t)^T = (x, y)^T \)

    Function \( f: \mathbb{R} \to \mathbb{R}^2 \), \( g(\theta) = (\cos \theta, \sin \theta)^T \)

    \( (\theta = \text{unit circle}) \)

    \( g'(\theta) = (-\sin \theta, \cos \theta)^T \)

    \( \Rightarrow \) can write \( y = y(x) \) near

    where \( g_2'(\theta) = \cos \theta \neq 0 \)

    and for \( x = x(y) \) near

    where \( g_1'(\theta) = -\sin \theta \neq 0 \)

Choose \( (x_0, y_0) = (\cos \theta_0, \sin \theta_0) \), \( \sin \theta_0 \neq 0 \)

\( x_0 \neq \pm 1 \)

\( \Rightarrow \theta = \arccos x \) \( \{ \text{make the branch} \) \( \theta_0 = \arccos x_0 \)

\( \Rightarrow y = \sin (\arccos x) = \pm \sqrt{1-x^2} \)
$2) \quad (x, y, z)^T = g(\theta, \phi) = (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \theta)$

$g'(\theta, \phi) = \begin{pmatrix}
-\sin \theta \sin \phi & \cos \theta \sin \phi & 0 \\
-\cos \theta \sin \phi & -\sin \theta \sin \phi & 0 \\
0 & 0 & -1
\end{pmatrix}$

If $\dot{\phi} = 0$, 1st column $= D \Rightarrow \text{rank } g' < 1 < 2 
\Rightarrow g'$ is not an immersion

Everywhere else, $g'$ is an immersion

$\Rightarrow g'(0 < \phi < \pi) = S^2 - \text{two hemispheres}$

If we want $\theta = 2\pi x(y)$,
we want the first two rows
if $g'$ linearly independent, i.e.

$\left( \begin{array}{c}
-\sin \theta \sin \phi & \cos \theta \sin \phi \\
-\cos \theta \sin \phi & -\sin \theta \sin \phi \\
0 & 0
\end{array} \right)$ must be
\[
\det (2 \times 2) = - \sin \phi \cos \phi \Rightarrow \cos \phi \neq 0 \\
\Rightarrow \text{away from the equation}
\]

For \(x = \cos \Theta \sin \phi, y = \sin \Theta \sin \phi\):

\[\begin{align*}
x^2 + y^2 &= \sin^2 \phi \Rightarrow \phi = \arctan \sqrt{x^2 + y^2} \\
x &= \tan \Theta \Rightarrow \Theta = \arctan \frac{y}{x}
\end{align*}\]

\[z = \cos \phi = \cos (\arcsin \sqrt{x^2 + y^2}) = \pm \sqrt{1 - x^2 - y^2}
\]

In the equation:

\[
y' = \begin{pmatrix} \frac{-\sin \Theta}{\cos \phi} & 0 \\ \frac{\cos \Theta}{\cos \phi} & \cos \Theta \end{pmatrix}
\]

\[x = x(y, z) \Rightarrow \cos \Theta \\
y = y(x, z) \Rightarrow \cos \Theta \\
z = \cos \Theta \Rightarrow y = \tan \Theta + \phi, z = \cos \phi
\]

\[\Rightarrow \phi = \arccos z, \Theta = \arctan \left( \frac{y}{\cos \phi} \right) = \arctan \left( \frac{y}{\cos \phi} \right)
\]

\[x = \cos \Theta \sin \phi = \cos \left[ \arcsin \left( \frac{y}{\cos \phi} \right) \right] \cos \left( \arccos z \right) = \pm \sqrt{1 - y^2 - z^2}
\]
Implicit description of surface

\[ F : \mathbb{R}^n \rightarrow \mathbb{R}^{n-m}, \quad F \in C^1, \]
\[ F(x) = 0 \quad \exists x \mid F(x) = c_j^3 \text{ - level set } \quad F \]

E.g. \[ F(x,y) = x^2 + y^2 - 1 \]

Suppose \[ M_m = \{ x \mid F(x) = 0 \} \] is a \n\[ c^1 \text{ m-dimensional surface (manifold)} \]
\[ \text{If } x(t) \text{ in a curve in } M_m \]
\[ \Rightarrow F_j(x(t)) = 0 \quad j = 1, \ldots, n-m \]
\[ \Rightarrow 0 = \frac{d}{dt} F_j(x(t)) = \nabla F_j(x(t)) \frac{dx}{dt} \]
\[ \nabla f_j \perp \frac{dx}{dt}, \quad \text{tangent vector} \]

\[ j = 1, \ldots, n-m \]

Since \( x(t) \) is any curve in \( M \) \[ \Rightarrow \nabla f_j \perp T_x M \quad \Rightarrow \nabla f_j \in N_x M \]

\[ \nabla f = (N_x M_m) \quad \text{Normal space} \]

\[ M_m = \mathbb{R}^2 \quad \text{plane} \]

From linear algebra \[ N_x M_m \perp T_x M_m \]

\[ \text{dim}(n-m) \quad \text{dim}(m) \]

If a \( C^1 \), \( m \)-dimensional surface \( S \) is given by \( F(x) = 0 \Rightarrow \nabla f(x) \) are \( n-m \) vectors in \( N_x M_m \). If \( \nabla f(x) \) are linearly independent,

\[ \Rightarrow \nabla f_j(x) \quad \text{from } N_x M_m \]
\[ \nabla f_i(x) \text{ are rank-1} \Rightarrow \text{linearly independent if} \]

\[ \text{rank} \{f'(x)\} = n-m \text{ (maximal)} \]

\[ \text{e.g. } x^2 + y^2 = 1, \quad z = 0 \quad \in \]

\[ = \} \theta \in \mathbb{R}^3 \equiv M, \]

Let \( \mathbf{r} = (x, y, z) \)

\[ T_{\mathbf{r}} M_1 = \{ \mathbf{r} + \lambda (-\cos \theta, \sin \theta, 0) \} \]

\[ N_{\mathbf{r}} M_1 = \{ \mu (2x, 2y, 0) + \mu (0, 0, 1) \} \]
Let $f : \mathbb{R}^m \to \mathbb{R}^{m-m}$ be a $C^1$ function and let $F(x)$ have rank $m-m$ at every point on the level set $M = \{ x \mid F(x) = c \}$. Then this level set is a $C^1$ $m$-dimensional surface (manifold).

Proof: Let $x \in M$. Since $\text{rank}(F(x)) = m-m$, we can find $m-m$ variables $x_i$ such that columns of $DF(x)$ are linearly independent. Let these $m \times (x_1, \ldots, x_{m-m}) = s$. Call

$$(x_{m-m+1}, \ldots, x_m) = t \iff \text{the level set equation is } F(t, s) = c;$$

since $F$ is invertible, by implicit function theorem, we can write $s = f(t)$ for $t$ in some neighborhood of $s$.

For $s = f(t)$ for $t$ in some neighborhood of $s$.\[ \begin{array}{c}
F(t_1, t_2, s) = 0 \\
2s + 0 \to \text{normal is not horizontal} \Rightarrow \text{tangent plane is not vertical} \Rightarrow \text{can write for } s = f(t_1, t_2) \end{array} \]
Example 1. $x^2 + y^2 + z^2 = 1$

$F(x, y, z) = x^2 + y^2 + z^2$, $F'(x, y, z) = (2x, 2y, 2z)$

$\Rightarrow F'(x, y, z)$ has rank 1 \iff $(x, y, z) = (0, 0, 0)$

$\Rightarrow$ by TMN, all spheres $x^2 + y^2 + z^2 = c$, $c > 0$

are $C^1$, 2-dimensional surfaces

2.) $x^2 + y^2 + z^2 = 1$, $z = 0$

$F(x, y, z) = (x^2 + y^2 + z^2)$

$= (1, 0, 0)$

$F'(x, y, z) = (2x, 2y, 2z)$

$\Rightarrow \text{rank}(F') = 2 \Rightarrow x \neq 0, y \neq 0$

since $x = y = z = 0$ does not satisfy these equations.

$\Rightarrow \text{rank}(F') = 2$

and $H_1$ is a 1-dimensional surface

$\Rightarrow$ circle

Remark: In general if $N_x M_m$

is spanned by $\nabla \Phi_i(x)$ \hspace{1cm} (rank $\Phi_i(x) = m - n$)

$\Rightarrow T_x M_m$ is given implicitly as

all the solutions to equations $\nabla \Phi_i(x) = 0$, $i = 1, \ldots, m - n$. 

$\bullet$
MAXIMA AND MINIMA ON SURFACES

"Extreme with side conditions"

Maximize $f(x)$, $f:\mathbb{R}^n \to \mathbb{R}$
subject to the constraints
$g(x) = (g_1(x), \ldots, g_k(x))^T = 0$

($g_i: \mathbb{R}^n \to \mathbb{R}$, $f, g \in C^1$)

Lagrange Multipliers: Form $H: \mathbb{R}^{n+k} \to \mathbb{R}$:

$H(x, \lambda) = f(x) + \lambda_1 g_1(x) + \cdots + \lambda_k g_k(x)$

and find all critical points
(i.e., points where $\nabla H(x, \lambda) = 0$)
of $H(x, \lambda)$. The constrained extreme occur at $x$-values of
these critical points (see below)
THM \[ f: \mathbb{R}^n \to \mathbb{R} \quad \text{and} \quad g: \mathbb{R}^n \to \mathbb{R}^k \]
be \( C^1 \) functions and let \( x \) be a point where \( g(x) = 0 \) and \( \text{rank}(g'(x)) = k \).
If \( f(x) \) is a local minimum for \( f \) in \( \mathbb{R}^n \) and \( g(x) = 0 \), then there exists a neighborhood \( U \) of \( x \) in \( \mathbb{R}^n \) such that \( f(y) \leq f(x) \)
for all \( y \in U \) in \( \mathbb{R}^n \) where \( g(y) = 0 \).
Then there exists \( x' = x+y \) such that \( f(x') = f(x) + \frac{1}{2} \cdot \| y \| \).
\( g(x) \) has a critical point at \((x,1)\).

**Remark:** Of course this method also finds "saddles."
Proof: \( \text{rank } \langle G(x) \rangle = k \Rightarrow \) the level set \( G(x) = 0 \) is an \((n-k)\)-dimensional, \( C^1 \) surface close to \( x \).

\[ H_x(x) = 0 \] just says \( G(x) = 0 \), so we must show \( H_x(x; t) = 0 \) for some \( t \).

Consider any \( C^1 \) curve \( \lambda(t) \) on \( x \) \( G(x) = 0 \), passing through \( x \).

Say \( \lambda(0) = x \).

\( \Rightarrow \) \( \text{fol}(t) \) is a \( C^1 \) function and has a minimum at \( t_0 \).

\( \Rightarrow \frac{d}{dt} (\text{fol}(t)) (0) = 0 \)

(by the 1-d theorem)
But \( \frac{d}{dt} \langle f_0(x), h(t) \rangle = f'(x) \cdot h'(t) \) \( \equiv \) \\
\( \equiv \nabla f(x) \cdot h'(t) \) \\
\( \Rightarrow \nabla f(x) \perp h'(t) = \text{tangent vector to the curve} \) \\
by varying the curves \( \Rightarrow \) \\
\( \nabla f(x) \perp \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_j} g_i(x) = 0 \) \\
\( \Rightarrow \nabla f(x) \in \mathbb{N}_x \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_j} g_i(x) = 0 \) \\
But \( \mathbb{N}_x \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_j} g_i(x) = \text{span} \left\{ \frac{\partial f_i}{\partial x_j}, i=1, \ldots, n \right\} \) \\
\( \Rightarrow \nabla f(x) = - \sum_{i=1}^{n} \nabla g_i (x) \) \\
in some \( \lambda \in \mathbb{R} \) \\
\( \Rightarrow H_x (\tilde{x}, \lambda) = 0 \)
E.g. Points on the ellipse 
\[ \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \]
closest to and farthest from the origin.

⇒ Find the extrema of \( x^2 + y^2 \) under the constraint
\[ \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0 \]

\[ H(x, y, \lambda) = x^2 + y^2 + \lambda \left( \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 \right) \]

\[ H_x = 2x + 2 \frac{Ax}{a^2} = 0 \]
\[ H_y = 2y + 2 \frac{By}{b^2} = 0 \]
\[ H_\lambda = \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0 \]
1) \( x = 0, \quad y = \pm b \quad (\lambda = -b^2) \)

2) \( y = 0, \quad x = \pm a \quad (\lambda = -a^2) \)

Usually, we can figure out if we have a min, max, or saddle from the context.

The second derivative test for a min (max, saddle) of \( f(x) \) on \( f''(x) < 0 \) if the quadratic form

\[ \langle u, \frac{d^n f(x)}{d x^n} u \rangle > 0 \]

is positive (negative, zero) definite for all \( u \in T_x, \quad \frac{d^n f(x)}{d x^n} = 0 \).

Proof: Stichcharts