So far we have considered <u>unconstrained</u> nonlinear optimization problems:

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

Now we will consider <u>equality constraints</u>:

$$(P): \begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & h_1(x) = 0 \\ & \vdots \\ & h_m(x) = 0. \end{cases}$$

We will learn the following result:

<u>Theorem</u>. If x_0 is a local minimizer for (P) and x_0 is "regular"

then $\exists\, u \in \mathbb{R}^m$ s.t. $\boxed{\nabla f(x_0) + u^T \nabla h(x_0) = 0}$

vector of Lagrange multipliers

Here: $\nabla h(x) := \begin{bmatrix} \nabla h_1(x) \\ \vdots \\ \nabla h_m(x) \end{bmatrix}.$

What do we mean by "regular point"?

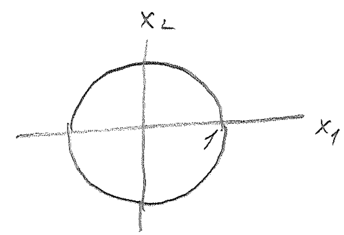
The feasible set $\mathcal{F}_e = \left\{ x \in \mathbb{R}^n : \begin{array}{c} h_1(x) = 0 \\ \vdots \\ h_m(x) = 0 \end{array} \right\}$.

A point $x \in \mathcal{F}_e$ is called a <u>regular point</u> if $\nabla h_1(x), \dots, \nabla h_m(x)$ are linearly independent vectors (in $\mathbb{R}^n$).

<u>Example</u> $\mathcal{F}_e = \{ x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1 \}$

$= \{ x \in \mathbb{R}^2 : h(x) = 0 \}$,

where $h(x) := x_1^2 + x_2^2 - 1$.



<u>Claim</u>: Every point in $\mathcal{F}_e$ is regular.

Note that $m=1$ and $\nabla h(x) = [2x_1 \quad 2x_2]$.

So $\nabla h(x)$ is independent if and only if $\nabla h(x) \neq 0$.

But if $\nabla h(x) = 0$, then $x_1 = x_2 = 0$ and then $h(x) = x_1^2 + x_2^2 - 1$
$= 0^2 + 0^2 - 1 = -1$
$\neq 0$!

So such an x can't be feasible.

Thus if x is feasible, $\nabla h(x) \neq 0$ and so x is regular.

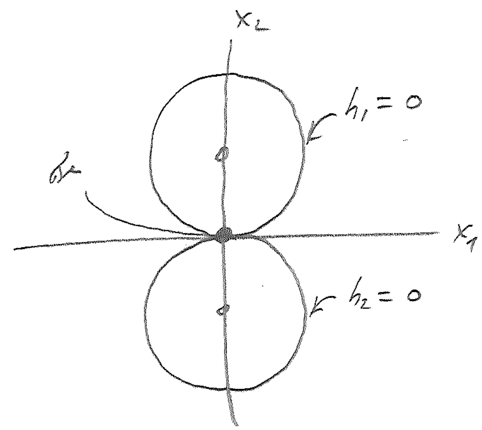
<u>Example</u>.

$$\mathcal{F} = \left\{ x \in \mathbb{R}^2 : \begin{array}{l} x_1^2 + (x_2-1)^2 = 1 \\ x_1^2 + (x_2+1)^2 = 1 \end{array} \right\}$$

$$= \left\{ x \in \mathbb{R}^2 : \begin{array}{l} h_1(x) = 0 \\ h_2(x) = 0 \end{array} \right\}, \text{ where}$$

$$h_1(x) := x_1^2 + (x_2-1)^2 - 1$$
$$h_2(x) := x_1^2 + (x_2+1)^2 - 1.$$

$$= \{(0,0)\} \quad \text{(see the figure).}$$



We have

$$\nabla h_1(x) = [2x_1 \quad 2(x_2-1)]$$
$$\nabla h_2(x) = [2x_1 \quad 2(x_2+1)].$$

Thus $\nabla h_1(0) = [0 \quad -2]$ and

$$\nabla h_2(0) = [0 \quad 2].$$

Since $1 \cdot \nabla h_1(0) + 1 \cdot \nabla h_2(0) = [0 \quad -2] + [0 \quad 2] = [0 \quad 0]$

it follows that $\nabla h_1(0), \nabla h_2(0)$ are <u>not</u> linearly independent.

So $(0,0)$ is <u>not</u> a regular point.

We will need the following technical lemma to prove the theorem.

Technical lemma.

If x_0 is a local minimizer and
x_0 is a regular point
then there is no $d \in \mathbb{R}^n$ s.t. $\nabla f(x_0) \cdot d < 0$ and
$\nabla h(x_0) d = 0$.

Once we have the technical lemma, our theorem is easy to show.

Theorem. If x_0 is a local minimizer and
x_0 is a regular point
then $\exists u \in \mathbb{R}^m$ s.t. $\nabla f(x_0) + u^T \nabla h(x_0) = 0$.

Proof. From the technical lemma, we know that
there is no d s.t. $\nabla f(x_0) \cdot d < 0$ and $\nabla h(x_0) d = 0$.
But then there is no d s.t.
$\nabla f(x_0) \cdot d > 0$ and $\nabla h(x_0) d = 0$ either.

Since otherwise $\nabla f(x_0) . (-d) < 0$ and $\nabla h(x_0) \cdot (-d) = 0$!

So if d satisfies $\nabla h(x_0)\, d = 0$, then $\nabla f(x_0) \cdot d = 0$.

i.e., if $d \in \ker \nabla h(x_0)$, then $\nabla f(x_0)\, d = 0$

i.e., $\nabla f(x_0)^T \in \left(\ker\left(\nabla h(x_0)\right)\right)^{\perp}$

$$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \shortparallel$$

So $\nabla f(x_0)^T \in \operatorname{ran} \left(\nabla h(x_0)\right)^T$

i.e., $\exists v$ s.t. $\nabla f(x_0)^T = \nabla h(x_0)^T v$

i.e., $\nabla f(x_0)^T - \left(\nabla h(x_0)\right)^T v = 0$

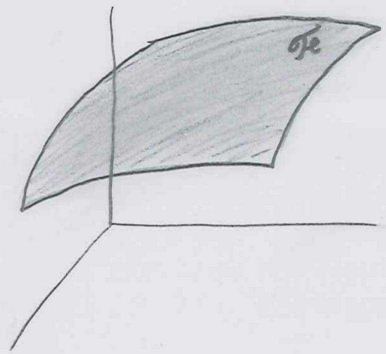
i.e., $\nabla f(x_0) - v^T \nabla h(x_0) = 0$

i.e., $\boxed{\nabla f(x_0) + u^T \nabla h(x_0) = 0}$ $(u := -v)$

$\square$

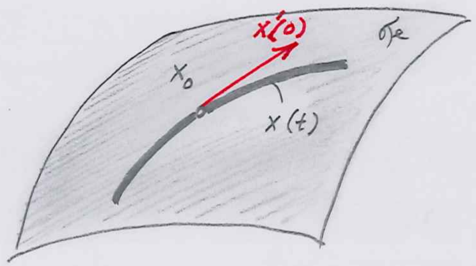
It remains to show the technical lemma.

Think of $\left\{\begin{array}{l} h_1(x) = 0 \\ \vdots \\ h_m(x) = 0 \end{array}\right\}$ as describing a surface.



Let $t \longmapsto x(t)$ be a curve in $\mathcal{S}$.

Let $x(0) = x_0$.



Then $h(x(t)) = 0 \quad \forall\, t$. So $\dfrac{h(x(t)) - h(x(0))}{t} = 0 \quad \forall\, t \neq 0$.

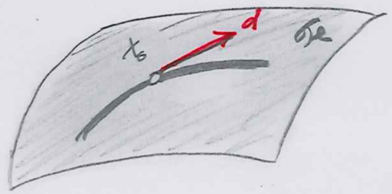
Thus $\left.\dfrac{d}{dt} h(x(t))\right|_{t=0} = 0$ i.e., $\nabla h(x_0) \cdot \underbrace{x'(0)}_{} = 0$.

So $x'(0)$ belongs to the kernel of $\nabla h(x_0)$.

Vice versa suppose that d is s.t. $\nabla h(x_0) . d = 0$.

If $\underline{x_0 \text{ is a regular point}}$, then it can be shown that there is a curve $t \longmapsto x(t)$ $(t \in \mathbb{R})$ in $\mathcal{S}$ s.t. $x(0) = x_0$ and $x'(0) = d$.

(This uses the Implicit Function Theorem.)



Technical lemma

If x_0 is a local minimizer and x_0 is a regular point

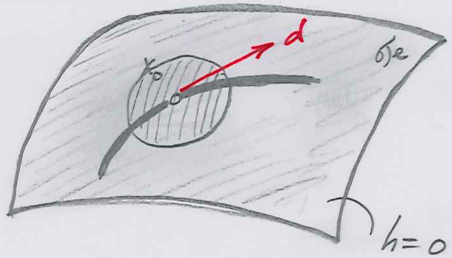
then there is no d s.t. $\nabla f(x_0)\, d < 0$ and $\nabla h(x_0)\, d = 0$.

Proof Suppose d is s.t. $\nabla h(x_0)\, d = 0$. (We want: $\nabla f(x_0)\, d \geqslant 0$)

By above discussion, $\exists$ a curve $t \mapsto x(t)$ $(t \in \mathbb{R})$ in $\mathcal{H}$ s.t.

$x(0) = x_0$ and

$x'(0) = d$.



But x_0 is a local minimizer.

So for small ts, $f(x(t)) \geqslant f(x_0) = f(x(0))$.

Hence $\dfrac{f(x(t)) - f(x(0))}{t} \geqslant 0$ for small positive ts.

Thus $\nabla f(x_0) \cdot x'(0) \geqslant 0$

i.e., $\nabla f(x_0) \cdot d \geqslant 0$. Done!

$\square$

Example $\begin{cases} \text{minimize} & x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 = 1 \end{cases}$ $\quad \left\{ \begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & h(x) = 0 \end{array} \right\}$, where $f(x) := x_1 + x_2$, $h(x) := x_1^2 + x_2^2 - 1$.

We had seen that every point in the feasible set is regular.

So our theorem says that if x_0 is a local minimizer,

then $\exists\, u$ s.t. $\nabla f(x_0) + u^T \nabla h(x_0) = 0$

$$\nabla f(x_0) = \begin{bmatrix} 1 & 1 \end{bmatrix}$$
$$\nabla h(x_0) = \begin{bmatrix} 2x_1 & 2x_2 \end{bmatrix}.$$

$\nabla f(x_0) + u^T \nabla h(x_0) = 0$ becomes $\begin{array}{l} 1 + 2ux_1 = 0 \\ 1 + 2ux_2 = 0. \end{array}$

Also since x_0 is feasible, we have $x_1^2 + x_2^2 = 1$.

So we should solve the system

$$\begin{cases} 1 + 2ux_1 = 0 \\ 1 + 2ux_2 = 0 \\ x_1^2 + x_2^2 = 1. \end{cases}$$

$x_1 = \frac{-1}{2u}$, $x_2 = \frac{-1}{2u}$ $\quad x_1^2 + x_2^2 = 1$ gives $\frac{1}{4u^2} + \frac{1}{4u^2} = 1$ i.e., $u^2 = \frac{1}{2}$

and so $u = \frac{1}{\sqrt{2}}$ or $u = -\frac{1}{\sqrt{2}}$.

So <u>possible</u> local minimizers are $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ or $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$.

But does there exist a global minimizer?

<u>Answer:</u> Yes.

To see why, we will learn the following result from calculus which is very important in optimization:

<u>Weierstrass' theorem</u>

If $K \subset \mathbb{R}^n$ is "compact" and $f: K \longrightarrow \mathbb{R}$ is continuous, then f has a global minimizer on K.

(So the problem $\left\{ \begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & x \in K \end{array} \right\}$ always has an optimal solution!)

Our function $f(x) = x_1 + x_2$ is continuous.

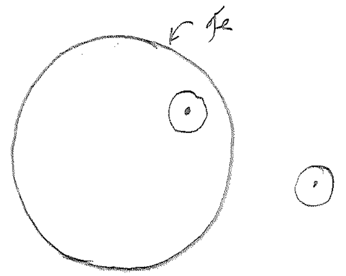
But is $\mathcal{F} = \{x : x_1^2 + x_2^2 = 1\}$ compact?

<u>What is a compact set?</u> A $\underbrace{\text{closed}}$ and $\underbrace{\text{bounded}}$ set is compact.

it is contained in some big ball.

for every point y in the complement, there is a small ball with center y which is contained in the complement.

<u>Example</u>



$\mathcal{F}$ is closed in $\mathbb{R}^2$
$\|$
$\{x : x_1^2 + x_2^2 = 1\}$

$\mathcal{F}$ is also bounded. $\left(\because \mathcal{F} \subset \{x : \|x\| \leq 1\}.\right)$

So $\mathcal{F}$ is compact.

So the problem $\left\{ \begin{array}{ll} \text{minimize} & x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 = 1 \end{array} \right\}$ has a global minimizer.

As $f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \sqrt{2} > -\sqrt{2} = f\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$, it follows that $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ is the <u>global minimizer!</u>

For checking compactness, it is useful to know that intersection of closed sets is closed.

<u>Example</u>

$$\mathcal{K} = \left\{ x \in \mathbb{R}^3 : \begin{array}{l} x_1^2 + x_2^2 + x_3^2 = 1 \\ x_1 + x_2 + x_3 = 0 \end{array} \right\} \begin{array}{l} \text{sphere} \\ \text{plane.} \end{array}$$

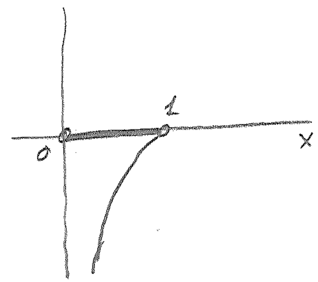
is compact.

<u>Examples</u>

(1) $(0,1)$ is bounded, but it is not closed in $\mathbb{R}$.

So $(0,1)$ is not compact.

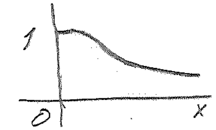
$f : (0,1) \longrightarrow \mathbb{R}$ given by $f(x) = \log x$ does not have a global minimizer on $(0,1)$.



(2) $[0, \infty)$ is closed in $\mathbb{R}$, but it is not bounded.

So $[0,\infty)$ is not compact.

$f : [0,\infty) \longrightarrow \mathbb{R}$ given by $f(x) = \dfrac{1}{1+x^2}$ does not have a global minimizer on $[0,\infty)$.



(3). $[0,1]$ is closed and bounded in $\mathbb{R}$

So $[0,1]$ is compact.

$f : [0,1] \longrightarrow \mathbb{R}$ given by

$$f(x) = \frac{x^{2010} \sin x - 3x^2 + 9}{e^x + e^{-x} + 1}$$ is continuous.

So $\left\{ \begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & x \in [0,1] \end{array} \right\}$ has a global minimizer, by Weierstrass' theorem.

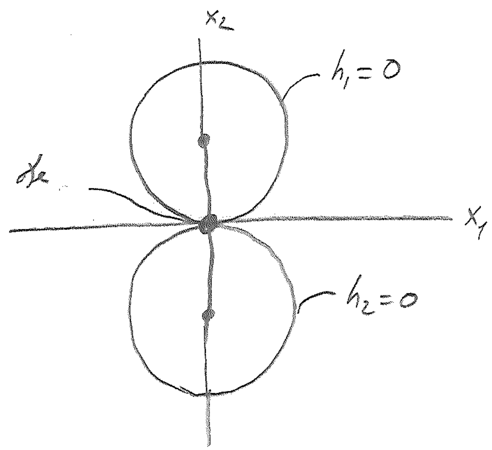
<u>Example</u> Consider the problem

$$\left\{ \begin{array}{ll} \text{minimize} & x_1 \\ \text{s.t.} & x_1^2 + (x_2 - 1)^2 = 1 \\ & x_1^2 + (x_2 + 1)^2 = 1. \end{array} \right.$$

i.e., $$\left\{ \begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & h_1(x) = 0 \\ & h_2(x) = 0 \end{array} \right.$$

where
$$f(x) := x_1$$
$$h_1(x) := x_1^2 + (x_2 - 1)^2 - 1$$
$$h_2(x) := x_1^2 + (x_2 + 1)^2 - 1.$$

Feasible set $\mathcal{F} = \{(0,0)\}$:



Obviously the unique global minimizer is $\hat{x} = (0,0)$.

We have: $\nabla f(x) = [1 \quad 0]$ and so $\nabla f(0) = [1 \quad 0]$

$\nabla h_1(x) = [2x_1 \quad 2(x_2-1)]$ and so $\nabla h_1(0) = [0 \quad -2]$.

$\nabla h_2(x) = [2x_1 \quad 2(x_2+1)]$ and so $\nabla h_2(0) = [0 \quad 2]$.

So the Lagrange equation becomes:

$$\nabla f(0) + u^T \nabla h(0) = 0$$

i.e., $[1 \quad 0] + [u_1 \quad u_2] \begin{bmatrix} 0 & -2 \\ 0 & 2 \end{bmatrix} = 0$

i.e., $[1 \quad -2u_1 + 2u_2] = 0$

which is never satisfied for any $u_1, u_2 \in \mathbb{R}$!

What went wrong?

What went wrong?

Answer: $(0,0)$ is <u>not</u> a regular point. $\left(1 \cdot \nabla h_1(0) + 1 \cdot \nabla h_2(0) = 1 \cdot [0 \;\; -2] + 1 \cdot [0 \;\; 2] = 0 \right.$.

So $\nabla h_1(0), \nabla h_2(0)$ are linearly independent.)

So there is no contradiction to our theorem!

Example. (Quadratic optimization under equality constraints revisited.)

$$\begin{cases} \text{minimize} & \frac{1}{2} x^T H x + c^T x + c_0 \\ \text{s.t.} & A x = b. \end{cases}$$

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & h(x) = 0 \end{cases}$$

where $f(x) := \frac{1}{2} x^T H x + c^T x + c_0$,

$h(x) := b - A x$.

Lagrange multiplier method

$$\left.\begin{array}{l} x_0 \text{ local minimizer and} \\ x_0 \text{ regular point} \end{array}\right\} \implies \exists u \text{ s.t } \nabla f(x_0) + u^T \nabla h(x_0) = 0 .$$

We have $\nabla f(x_0) = (Hx+c)^T$

$\nabla h(x_0) = -A$

x_0 regular point $\iff$ rows of $\nabla h(x_0)$ linearly independent

$\iff$ rows of A linearly independent.

Also: $\nabla f(x_0) + u^T \nabla h(x_0) = 0$

becomes $(Hx_0+c)^T + u^T(-A) = 0$

i.e., $Hx_0 + c - A^T u = 0$.

Moreover, $Ax_0 = b$. ($\because$ x_0 is feasible).

Together: $\exists u$ s.t. $\begin{bmatrix} H & -A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x_0 \\ u \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$.

Same as the Lagrange method equations from before.

So we have:

<u>Theorem</u>. If A has linearly independent rows and x_0 is a local minimizer, then $\exists u \in \mathbb{R}^m$ s.t. $\begin{bmatrix} H & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ u \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$.

Compare this to our earlier result:

<u>Theorem</u>. Let $\{z_1, \dots, z_k\}$ be a basis for the kernel of A, and $Z := [z_1 \dots z_k]$.

If $Z^T H Z$ is not p.s.d., then there is no global minimizer.

If $Z^T H Z$ is p.s.d., then

$$\boxed{\hat{x} \text{ is a global minimizer}} \iff \boxed{\exists u \text{ s.t. } \begin{bmatrix} H & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ u \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}}.$$