Definition 12.3.

Given a feasible point x and the active constraint set $\mathcal{A}(x)$ of Definition 12.1, the set of linearized feasible directions $\mathcal{F}(x)$ *is*

$$\mathcal{F}(x) = \left\{ d \;\middle|\; \begin{array}{ll} d^T \nabla c_i(x) = 0, & \text{for all } i \in \mathcal{E}, \\ d^T \nabla c_i(x) \geq 0, & \text{for all } i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}.$$

Note : $\mathcal{F}(x)$ is a cone. [$C \subseteq \mathbb{R}^n$ is a cone if $x \in C, \lambda \geq 0 \Rightarrow \lambda x \in C$]

From the three examples, we should expect that in general :

ideas from local linear approximation

x^* solves $\min_x f(x)$ s.t. $c_i(x) = 0,\ i \in \mathcal{E}$, $c_i(x) \geq 0\ i \in \mathcal{I}$ $\Rightarrow$ $\underbrace{\nexists\, d \in \mathbb{R}^n \text{ s.t. } d \in \mathcal{F}(x^*) \text{ and } \nabla f(x^*)^T d < 0}$ $(\ast)$

and the remaining work is to convert this condition to a more convenient condition (one that involves Lagrange multipliers)

This step should only involves linear algebra!

We shall do exactly this to get our first major result of this course.

But notice an annoying technicality :

In our first example, if we change the constraint $\overbrace{x_1^2+x_2^2-2}^{c_1(x)}=0$ to the equivalent :

$$\underbrace{(x_1^2+x_2^2-2)^2}_{\text{new } c_1(x)} = 0, \quad \leftarrow \text{represent the same circle}$$

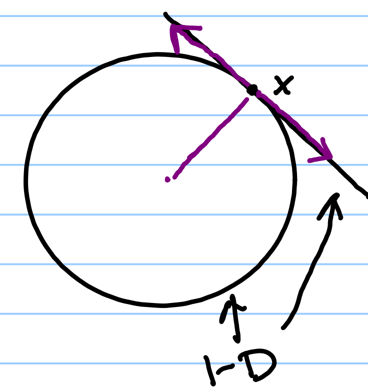
then

$$\nabla(\text{new } c_1)(x) = 2(x_1^2+x_2^2-2)\begin{bmatrix}2x_1\\2x_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \quad \forall \begin{bmatrix}x_1\\x_2\end{bmatrix} \text{ on the circle.}$$

with the original c_1 :

$$\mathcal{F}(x) = \{d : 2x^Td = 0\} = \text{the tangent line of the constraint set at } x.$$

$= \text{all vectors} \perp x$



with the new c_1 :

$$\neq$$

$$\mathcal{F}(x) = \{d : \begin{bmatrix}0\\0\end{bmatrix}^T d = 0\} = \mathbb{R}^2 \longleftarrow \text{2-D}$$

Summary: $\min \underbrace{x_1+x_2}^{=f(x)}$ has $x^* = \begin{bmatrix}-1\\-1\end{bmatrix}$ as its solution, but the
s.t. $\underbrace{(x_1^2+x_2^2-2)^2}_{c_1(x)}=0$ condition ⊛ does not hold!

We have actually seen this problem earlier:

In general, if $c(x)=0$, we expect $\{y: c(y)=c(x)\} = c^{-1}(c(x))$ to be a hypersurface near x, $\nabla c(x)$ is orthogonal to the hypersurface, and

$$\underbrace{\{d: \nabla c(x)^T d = 0\}}_{(n-1)\text{-dimensional when } \nabla c(x) \neq \vec{0}} = \text{the tangent plane of } c^{-1}(c(x)) \text{ at } x.$$

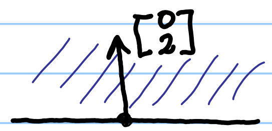
But this picture can totally fall apart if $\nabla c(x) = \vec{0}$! (recall my counter-examples.)

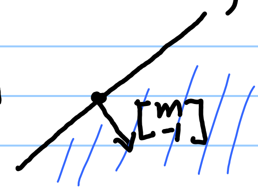
So maybe everything is fine (for the expected optimality theorem) if we impose $\nabla c_i(x^*) \neq \vec{0}$ $\forall i$?

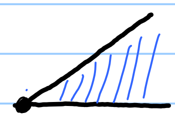
It turns out that we will still have a problem:

Consider the constraints
$$c_1(x) = 1 - x_1^2 - (x_2-1)^2 \geq 0$$
$$c_2(x) = -x_2 + m x_1 \geq 0, \quad (m \in \mathbb{R})$$

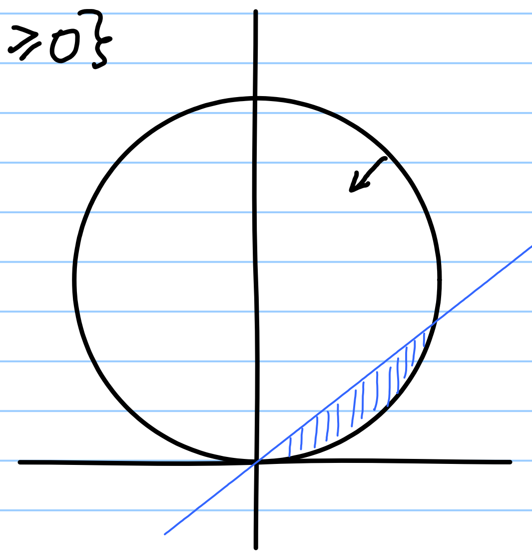
$$\mathcal{F}\left(\begin{bmatrix}0\\0\end{bmatrix}\right) = \left\{ d \in \mathbb{R}^2 : \overbrace{\nabla c_1(\vec{0})}^{\begin{bmatrix}0\\2\end{bmatrix}}{}^T d \geq 0 ,\ \overbrace{\nabla c_2(\vec{0})}^{\begin{bmatrix}m\\-1\end{bmatrix}}{}^T d \geq 0 \right\}$$

$=$ 

$\cap$ 

$=$ 

good first order approximation near $\begin{bmatrix}0\\0\end{bmatrix}$



what if $m=0$?

$$\{c_1(x) \geq 0, c_2(x) \geq 0\} = \{\underset{\substack{\uparrow\\ \text{0-dimensional}}}{\begin{bmatrix}0\\0\end{bmatrix}}\} \not\approx \mathcal{F}\left(\begin{bmatrix}0\\0\end{bmatrix}\right) = \{\underset{\substack{\uparrow\\ \text{1-dimensional}}}{\begin{bmatrix}d_1\\0\end{bmatrix}} : d_1 \in \mathbb{R}\}$$

If we consider $\min x_1+x_2$ s.t. $\begin{cases} 1-x_1^2-(x_2-1)^2 \geqslant 0 \\ -x_2 \geqslant 0 \end{cases}$

The minimizer is at $\begin{bmatrix}0\\0\end{bmatrix}$, as $\Omega = \left\{\begin{bmatrix}0\\0\end{bmatrix}\right\} \longleftarrow$ the feasible region

The hoped-for necessary condition will not hold:

$$\underbrace{\begin{bmatrix}1\\1\end{bmatrix}^T}_{\nabla f(\begin{bmatrix}0\\0\end{bmatrix})} \underbrace{\begin{bmatrix}d_1\\0\end{bmatrix}}_{\in \mathcal{F}(\begin{bmatrix}0\\0\end{bmatrix})} < 0 \quad \forall d_1 < 0$$

In this case, neither $\underbrace{\nabla c_1(\vec{0})}_{=\begin{bmatrix}0\\2\end{bmatrix}}$ nor $\underbrace{\nabla c_2(\vec{0})}_{=\begin{bmatrix}0\\-1\end{bmatrix}}$ is $\vec{0}$, but the two vectors are parallel.

Definition 12.4 (LICQ).

Given the point x and the active set $\mathcal{A}(x)$ defined in Definition 12.1, we say that the linear independence constraint qualification (LICQ) *holds if the set of active constraint gradients $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ is linearly independent.*

a.k.a. Karush-Kuhn-Tucker (KKT) conditions

Theorem 12.1 (First-Order Necessary Conditions).

Suppose that x^* is a local solution of (12.1), that the functions f and c_i in (12.1) are continuously differentiable, and that the LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^*, with components λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{12.34a}$$

$$c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E}, \tag{12.34b}$$

$$c_i(x^*) \geq 0, \quad \text{for all } i \in \mathcal{I}, \tag{12.34c}$$

$$\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I}, \tag{12.34d}$$

$$\underbrace{\lambda_i^* c_i(x^*)}_{\text{if constraint } i \text{ is inactive (i.e. } c_i(x)>0\text{), then the corresponding } \lambda_i = 0} = 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}. \tag{12.34e}$$

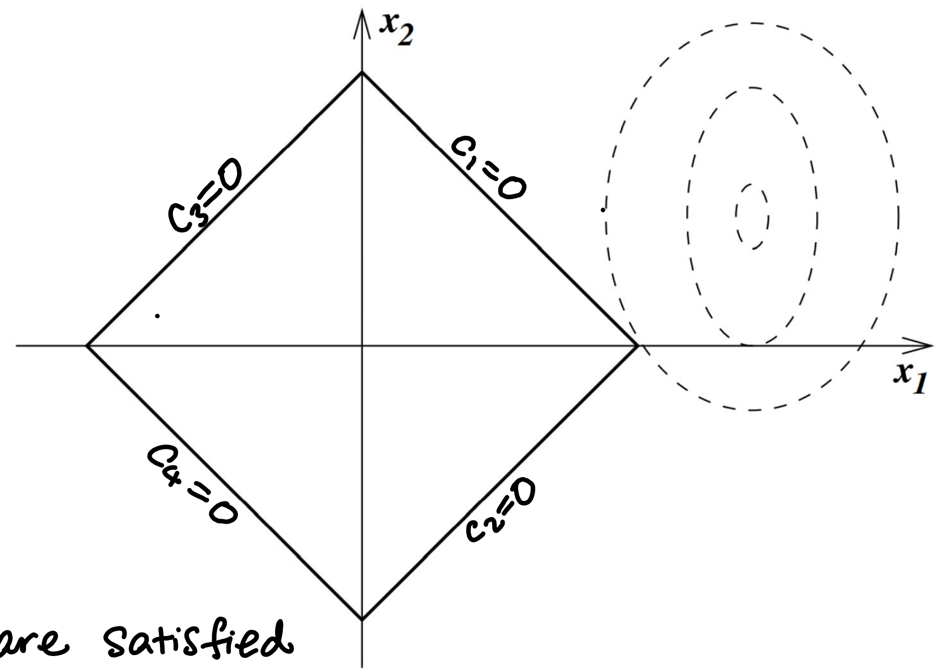
$\mathcal{L}(x, \lambda)$
$:= f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$

Example

$$\min_{x} \ (x_1 - 3/2)^2 + (x_2 - 1/2)^4 \quad \text{s.t.} \quad \begin{matrix} c_1 = \\ c_2 = \\ c_3 = \\ c_4 = \end{matrix} \begin{bmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{bmatrix} \geq 0 \qquad \mathcal{E} = \phi, \ \mathcal{I} = \{1, 2, 3, 4\}$$

Solution at $x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

$\mathcal{A}(x^*) = \{1, 2\}$



$$\nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$\nabla c_2(x^*) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -1/2 \end{bmatrix} = \frac{3}{4} \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\lambda^* = \begin{bmatrix} 3/4 \\ 1/4 \\ 0 \\ 0 \end{bmatrix}.$$

KKT conditions are satisfied at x^*.

Strategy for proving KKT :

(I) we have explained that the hoped-for result

$$\boxed{\begin{array}{ll} x^* \text{ solves} & \\ \displaystyle\min_x f(x) & \Rightarrow \underbrace{\nexists\, d \in \mathbb{R}^n \text{ s.t. } d \in \mathcal{F}(x^*) \text{ and } \nabla f(x^*)^T d < 0}\\ \text{s.t. } c_i(x) = 0,\ i \in \mathcal{E} & \\ \quad c_i(x) \geq 0\ i \in \mathcal{I} & \Leftrightarrow \nabla f(x^*)^T d \geq 0, \text{ for all } d \in \mathcal{F}(x^*) \end{array}}$$

does not hold without any assumption on $\nabla c_i(x)$, $i \in \mathcal{A}(x^*)$.

Notice that the solution should depend only on f and the feasible set Ω set itself, but not the algebraic specification of Ω.

Yet, as we showed, the cone $\mathcal{F}(x^*)$ does depend on the algebraic specification of Ω.

To correct this problem, we show that there is a more geometrically defined cone, called the tangent cone and denoted by $T_\Omega(x^*)$, so that

$$x^* \text{ is a solution} \implies \nabla f(x^*)^T d \geq 0, \text{ for all } d \in \cancel{\mathcal{F}(x^*)}\ T_\Omega(x^*)$$

(II) We show that under the LICQ assumption, $T_\Omega(x^*) = \mathcal{F}(x^*)$.

This essentially follows from the implicit function theorem.

(III) We show that: $\nabla f(x^*)^T d \geq 0$, for all $d \in \mathcal{F}(x^*)$ is equivalent to the KKT conditions.

This essentially follows from Farkas' lemma.

Spirit:
- a minimizer $\Rightarrow$ a local minimizer
- Steps (I) and (II) convert the local minimizer problem, using <u>local linear approximations</u>, into a linear algebra problem.
- Step (III) is pure linear algebra.

Definition 12.2.

The vector d is said to be a tangent *(or* tangent vector*) to Ω at a point x if there are a feasible sequence $\{z_k\}$ approaching x and a sequence of positive scalars $\{t_k\}$ with $t_k \to 0$ such that*

$$\lim_{k \to \infty} \frac{z_k - x}{t_k} = d. \tag{12.29}$$

The set of all tangents to Ω at x^* is called the tangent cone *and is denoted by $T_\Omega(x^*)$.*

$T_\Omega(x^*)$ depends only on the geometry of Ω, not the algebraic specification of Ω.

Recall:

Definition 12.3.

Given a feasible point x and the active constraint set $\mathcal{A}(x)$ of Definition 12.1, the set of linearized feasible directions *$\mathcal{F}(x)$ is*

$$\mathcal{F}(x) = \left\{ d \;\middle|\; \begin{array}{ll} d^T \nabla c_i(x) = 0, & \text{for all } i \in \mathcal{E}, \\ d^T \nabla c_i(x) \geq 0, & \text{for all } i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}$$

depends on the algebraic specification of Ω

The glory details.

(I) should be obvious intuitively, and is actually easy to prove.

Theorem 12.3.

If x^* is a local solution of (12.1), then we have

means f doesn't $\uparrow$ in direction d

means d points towards Ω

$$\nabla f(x^*)^T d \geq 0, \quad \text{for all } d \in T_\Omega(x^*).$$

Proof: Assume the contrary that $\exists\, d \in T_\Omega(x^*)$ s.t. $\nabla f(x^*)^T d < 0$.

Let $\{z_k\}$ and $\{t_k\}$ be the sequences satisfying Definition 12.2 for this d.

Then:

$$f(z_k) = f(x^*) + \nabla f(x^*)^T \underbrace{(z_k - x^*)}_{= t_k d + o(t_k)} + \underbrace{o(\|z_k - x^*\|)}_{= o(t_k)}$$

$$= f(x^*) + \underbrace{\nabla f(x^*)^T d}_{<0} + o(t_k)$$

So, $\underset{\Omega}{\overset{\Downarrow}{f(z_k)}} < f(x^*) + 0.99\, \nabla f(x^*)^T d$ for large enough k.

This means x^* cannot be a local solution. Q.E.D.

Of course, the converse of this result is not true.

Counterexamples we have seen before:

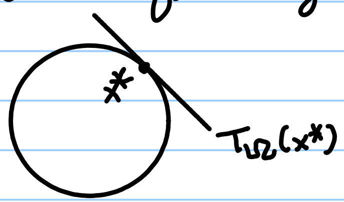
(No constraint:) $\min -\sum_i x_i^2$

$\Omega = \mathbb{R}^n$, $x^* = \vec{0}$, $\nabla f(x^*) = \vec{0}$

$T_\Omega(x^*) = \mathbb{R}^n$ $\quad \nabla f(x^*)^T d = 0 \;\; \forall d.$

But x^* is not a local minimizer. (It is a maximizer.)

(1 equality constraint:) $\min x_1 + x_2$ s.t. $x_1^2 + x_2^2 = 2$

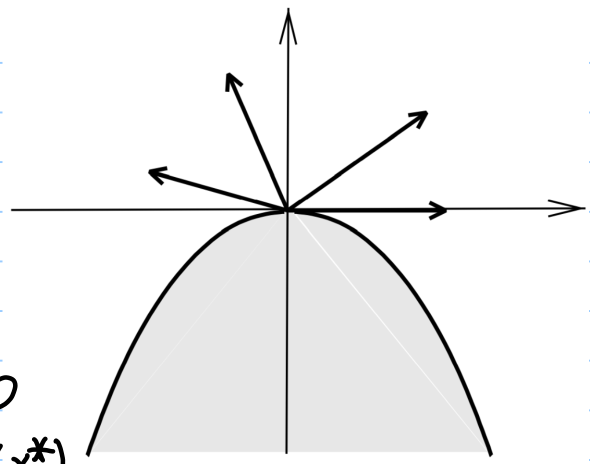


$x^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $T_\Omega(x^*) = \{ d : \begin{bmatrix} 1 \\ 1 \end{bmatrix}^T d = 0 \}$

maximizer! $\nabla f(x^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\nabla f(x^*)^T d = 0 \;\; \forall d \in T_\Omega(x^*)$

A counterexample with one inequality constraint (less obvious):

$$\min_{x \in \mathbb{R}^2} x_2 \quad \text{s.t.} \quad x_2 \geq -x_1^2$$



$x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $T_\Omega(x^*) = \left\{ \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} : d_2 \geq 0 \right\}$

(why?)

$\nabla f(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\nabla f(x^*)^T d = d_2 \geq 0$

$\forall d \in T_\Omega(x^*)$

But x^* is not a local minimizer.

Step II

Lemma 12.2.

Let x^* be a feasible point. The following two statements are true.

(i) $T_\Omega(x^*) \subset \mathcal{F}(x^*)$.

(ii) *If the LICQ condition is satisfied at x^*, then $\mathcal{F}(x^*) = T_\Omega(x^*)$.*

Recall:

Theorem A.2 (Implicit Function Theorem).

Let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be a function such that

(i) $h(z^*, 0) = 0$ *for some* $z^* \in \mathbb{R}^n$,

(ii) *the function* $h(\cdot, \cdot)$ *is continuously differentiable in some neighborhood of* $(z^*, 0)$, *and*

(iii) $\nabla_z h(z, t)$ *is nonsingular at the point* $(z, t) = (z^*, 0)$.

Then there exist open sets $\mathcal{N}_z \subset \mathbb{R}^n$ *and* $\mathcal{N}_t \subset \mathbb{R}^m$ *containing* z^* *and* 0, *respectively, and a continuous function* $z : \mathcal{N}_t \to \mathcal{N}_z$ *such that* $z^* = z(0)$ *and* $h(z(t), t) = 0$ *for all* $t \in \mathcal{N}_t$. *Further,* $z(t)$ *is uniquely defined. Finally, if* h *is* p *times continuously differentiable with respect to both its arguments for some* $p > 0$, *then* $z(t)$ *is also* p *times continuously differentiable with respect to* t, *and we have*

$$\nabla z(t) = -\nabla_t h(z(t), t)[\nabla_z h(z(t), t)]^{-1}$$

for all $t \in \mathcal{N}_t$.

Proof of (i): Let $d \in T_\Omega(x^*)$. By definition, $\exists\ z_k \in \Omega,\ t_k > 0$ s.t.

$$z_k \to x^*,\quad t_k \to 0 \quad \text{and} \quad \frac{z_k - x^*}{t_k} \to d.$$

$$\Updownarrow$$

$$\frac{z_k - x^* - t_k d}{t_k} \to 0 \Leftrightarrow z_k = x^* + t_k d + o(t_k).$$

If $i \in \mathcal{A}(x^*) \cap \mathcal{E}$,
then

$$0 = \frac{1}{t_k} c_i(z_k) = \frac{1}{t_k}\Big[\underbrace{c_i(x^*)}_{=0} + \nabla c_i(x^*)^T(\underbrace{z_k - x^*}_{= t_k d + o(t_k)}) + \underbrace{o(\|z_k - x^*\|)}_{= o(t_k)}\Big] \qquad (1)$$

$$= \nabla c_i(x^*)^T d + \underbrace{\frac{o(t_k)}{t_k}}_{\to 0}. \quad \text{so } \nabla c_i(x^*)^T d = 0.$$

If $i \in \mathcal{A}(x^*) \cap \mathcal{I}$,
then

$$0 \le \frac{1}{t_k} c_i(z_k) \underset{\text{similar to (1)}}{=} \nabla c_i(x^*)^T d + \underbrace{\frac{o(t_k)}{t_k}}_{\to 0}. \quad \text{so } \nabla c_i(x^*)^T d \ge 0.$$

$$\boxed{\text{Note: } t_k \gg o(t_k), \quad \frac{o(t_k)}{t_k} \gg o(t_k)}$$

Idea for the proof of (ii)

For $d \in \mathcal{T}(x^*)$ we need to find $\underset{\in\Omega}{z_k}$, $t_k \nearrow 0$, $t_k > 0$ st

$$\frac{z_k - x^*}{t_k} \to d \quad \text{— (I)}$$

choosing $z_k = x^* + t_k d$ works if d points towards the interior of Ω.

But doesn't work if d is tangent to one of the level surface $c_i(x) = 0$

If $c_i(z_k) = t_k \nabla c_i^T(x^*) d \quad \forall i \in \mathcal{A}(x^*)$ } a system of $m = |\mathcal{A}(x^*)|$ nonlinear equations in n variables. (*)

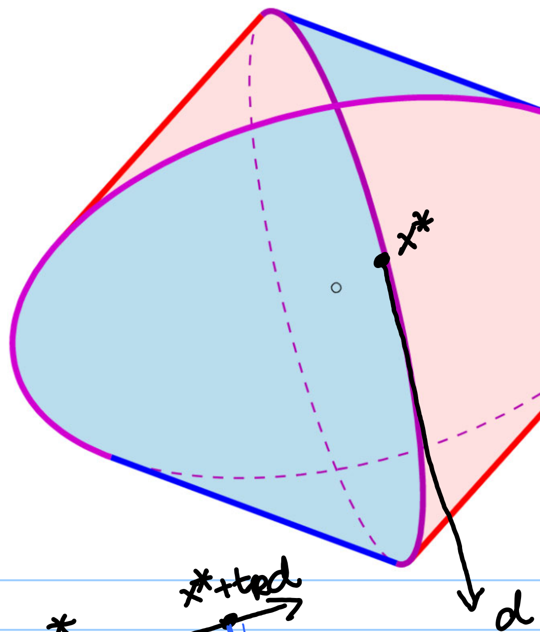
then $z_k \in \Omega$. (why?)

On the other hand, if z_k, t_k satisfy (I), then

$$c_i(z_k) \approx \underbrace{c_i(x^*)}_{=0} + \nabla c_i(x^*)^T \underbrace{(z_k - x^*)}_{\approx t_k d} \approx t_k \nabla c_i(x^*)^T d$$

So the problem is essentially about solving the under-determined nonlinear system (*).

Generically, there should be $(n-m)$ d.o.f. in choosing the solution z_k for a fixed t_k.



$$c_i(z_k) \approx \nabla c_i(x^*)^T (z_k - x^*)$$

not okay to choose $z_k = x^* + t_k d$,
(as this may make z_k infeasible.)
but it should be okay to choose $z_k \in \Omega$ so that

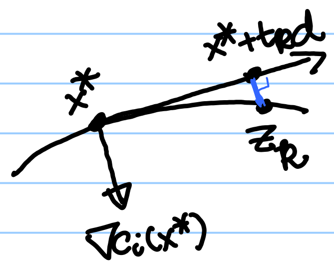
$$z_k - (x^* + t_k d) \;/\!/\; \nabla c_i(x^*)$$

$$\Longleftrightarrow$$

$$P\big(z_k - (x^* - t_k d)\big) = 0 \quad \text{— } (**)$$

$\uparrow$ ortho-projection onto $\bigcap_{i=1}^{m} \nabla c_i(x^*)^{\perp}$

$$\overset{\shortparallel}{\text{null}}\left(\underbrace{\begin{bmatrix} \nabla c_1(x^*)^T \\ \vdots \\ \nabla c_m(x^*)^T \end{bmatrix}}_{\substack{\mathbb{R}^n \to \mathbb{R}^m \\ \text{full rank}}}\right)$$



So, pick any basis $b_1, -, b_{n-m}$ of the null space, and set

$$Z^T = \underbrace{\begin{bmatrix} b_1^T \\ \vdots \\ b_{n-m}^T \end{bmatrix}}_{(n-m)\times n}.$$

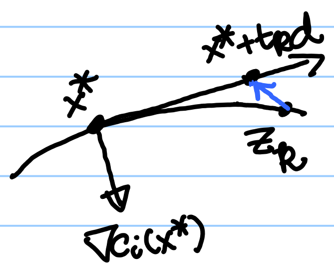
Then $(**) \Longleftrightarrow Z^T(z_k - x^* - t_k d) = 0.$

So we look for z_k by solving the (square) system of nonlinear equations

$$(***) \quad \begin{cases} c_i(z) - t_k \nabla c_i(x_*)^T d = 0 & \leftarrow n \text{ vars, } m \text{ eqts} \\ Z^T(z - x^* - t_k d) = 0 & \leftarrow (n-m) \text{ extra eqts to remove redundancy} \end{cases}$$

for every small $t_k > 0$.

Of course, (***) isn't truely necessary, an obique projection should also work.



It is fine to replace Z^T above by any $B \in \mathbb{R}^{(m-n)\times n}$ so that

$$\begin{bmatrix} - \nabla c_1(x^*)^T - \\ \vdots \\ - \nabla c_m(x^*)^T - \\ B \end{bmatrix}$$

is invertible

[End of the intuitive explanation of the proof, now the rigorous proof:]

Proof of (ii). For notational convenience, assume $c_1, \cdots, c_m$ are the active constraints, i.e. $\mathcal{A}(x^*) = \{1, \cdots, m\}$.

Write $C(z) = \begin{bmatrix} c_1(z) \\ \vdots \\ c_m(z) \end{bmatrix}$, $A(x^*) = \begin{bmatrix} \nabla c_1(x^*)^T \\ \vdots \\ \nabla c_m(x^*)^T \end{bmatrix} \in \mathbb{R}^{m \times n}$. ($m \le n$ as the gradients are linearly indep.)

Now, assume $d \in \mathcal{F}(x^*)$, i.e. $\nabla c_i^T(x^*) d = 0$, $i \in \mathcal{E}$

$\nabla c_i^T(x^*) d \geq 0$, $i \in \mathcal{I}$

Let $t_k > 0$ be s.t. $t_k \to 0$. Our goal is to find $z_k \in \Omega$ s.t. $\dfrac{z_k - x^*}{t_k} \to d$. (C)

Let $B \in \mathbb{R}^{(n-m) \times m}$ so that $\begin{bmatrix} A(x^*) \\ B \end{bmatrix} \in \mathbb{R}^{n \times n}$ is non-singular, and consider the parametrized system of equations

$$(\ast) \quad R(z, t) := \begin{bmatrix} C(z) - tA(x^*)d \\ B(z - x^* - td) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad R: \mathbb{R}^{n+1} \to \mathbb{R}^n.$$

Claim: the solutions $z = z_k$ of this system for small $t = t_k > 0$ give a feasible sequence that approaches x^* and satisfies the condition (C).

Note: $\nabla_z R(x^*, 0) = \nabla_z \begin{bmatrix} c(z) \\ B(z - x^*) \end{bmatrix} = \begin{bmatrix} A(x^*) \\ B \end{bmatrix} \leftarrow$ invertible

So, by the implicit function theorem, $\exists\, \varepsilon > 0$ st $\forall\, t \in (-\varepsilon, \varepsilon)$, there is a unique solution $z(t)$ of the system (*), ie.

$$R(z(t), t) = 0.$$

The solutions $z_k = z(t_k)$ are what we need! Check:

- $i \in \mathcal{E} \Rightarrow c_i(z_k) = t_k \underbrace{\nabla c_i(x^*)^T d}_{=0} = 0$

  $i \in \mathcal{I} \cap \mathcal{A}(x^*) \Rightarrow c_i(z_k) = \underbrace{t_k}_{>0} \underbrace{\nabla c_i(x^*)^T d}_{\geq 0} \geq 0.$

  $\left[i \notin \mathcal{A}(x^*),\ c_i(x^*) > 0 \quad z_k \to x^*, \text{ so } c_i(z_k) > 0 \text{ for large } k. \right]$

  So z_k is indeed feasible.

- $0 = R(z_k, t_k) = \begin{bmatrix} c(z_k) - t_k A(x^*) d \\ B(z_k - x^* - t_k d) \end{bmatrix}$

$$= \begin{bmatrix} \overbrace{c(x^*)}^{=0} + \overbrace{\nabla c(x^*)^T}^{=A(x^*)} (z_k - x^*) + o(\|z_k - x^*\|) - t_k A(x^*) d \\ B(z_k - x^* - t_k d) \end{bmatrix}$$

invertible

$$= \begin{bmatrix} A(x^*) \\ B \end{bmatrix} (z - x^* - t_k d) + o(\|z_k - x^*\|)$$

so $$\frac{z-x^*}{t_k} = d + o\left(\frac{\|z_k - x^*\|}{t_k}\right)$$

$\uparrow$ call it q_k $\qquad$ $\frac{\|z_k - x^*\|}{t_k}$ " $\|q_k\|$

$\leftarrow$ This relation says that $\|q_k\|$ has to be bounded.
But then $o(\|q_k\|) \to 0$.

we have $\frac{z-x^*}{t_k} \to d$ as $k \to \infty$.

Q.E.D.

Note: the LICQ condition can be dispensed with for linear constraints, and can be replaced by a weaker condition in general.

more about these later. And I'll show you a fun example in the HW.

Step III

Lemma (Farkas)

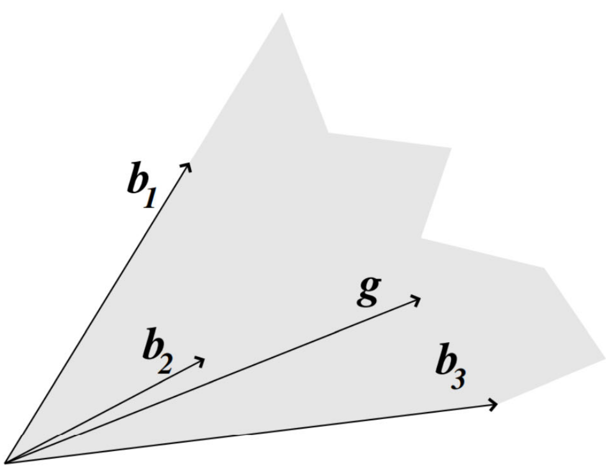
Given any vectors $g, b_1, \cdots, b_m, c_1, \cdots, c_p \in \mathbb{R}^n$.
We have either:

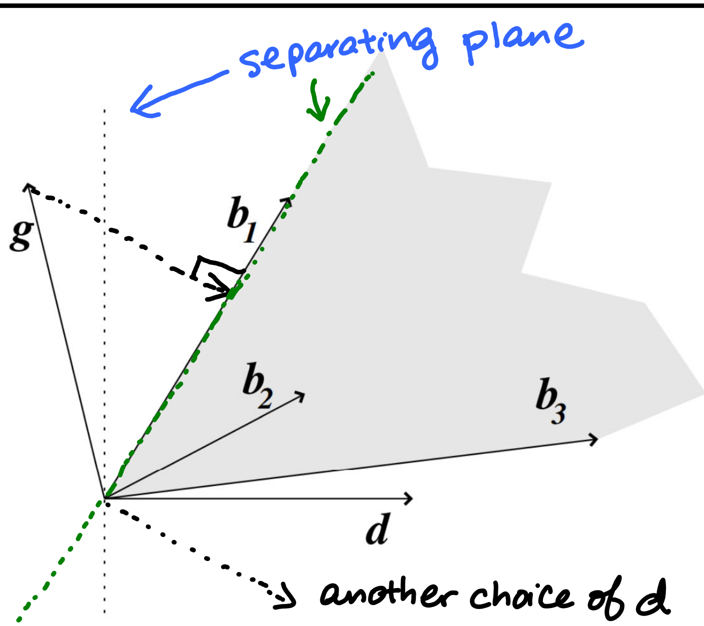
- $g \in K = \{ \overset{n\times m}{B} y + \overset{n \times p}{C} w : y \geq 0 \}$, $B = [b_1, \cdots, b_m]$, $C = [c_1, \cdots, c_p]$

OR else

- $\exists\, d \in \mathbb{R}^n$ st $g^T d < 0$, $B^T d \geq 0$, $C^T d = 0$.

But not both.





For <u>any</u> such separating plane $d^\perp$,

$g^T d < 0$

$s^T d \geq 0 \quad \forall s \in K$

$\Rightarrow \underbrace{(By + Cw)^T d}_{= y^T B^T d + w^T C^T d} \geq 0 \quad \forall y \geq 0, w$

$\Rightarrow B^T d \geq 0, \; C^T d = 0$

So all we need is the existence of a separating plane, which seems obvious!

Put differently, $\boxed{\nexists\ d \in \mathbb{R}^n \ \text{ st } \ g^T d < 0,\ B^T d \geq 0,\ C^T d = 0 \iff g \in K.}$

<u>Proof</u> (modulo a technical step):

(Easy step) We first show that the two alternatives cannot hold simultaneously. If $g \in K$, i.e.

$$g = By + Cw \quad , \ y \geq 0,$$

and also

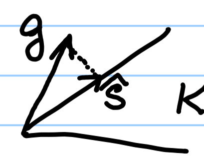
$$g^T d < 0,\ B^T d \geq 0,\ C^T d = 0$$

then

$$0 > g^T d = (By + Cw)^T d = \underbrace{y^T}_{\geq 0} \underbrace{B^T d}_{\geq 0} + w^T \underbrace{C^T d}_{=0} \geq 0, \text{ a contradiction.}$$

(Harder step):

Assume $g \notin K$, we construct d satisfying the properties. We choose d in the following way:



$$\boxed{\text{Let } \hat{s} \in \operatorname*{argmin}_{s \in K} \|s - g\|_2^2 \ , \quad \text{and } \ d = \hat{s} - g.}$$

A minimizer exists because <u>K is closed</u>, but it requires some care to prove it. (We omit the argument here. see Beck Ch6 or N-W Lemma 12.15)

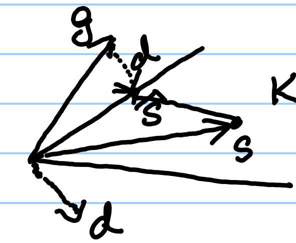
[It should be intuitively clear that $d^\perp$ separates g from K. Here is a proof:]

It should be clear that $\hat{s}^T \perp \hat{s}-g$. (If not, it is because you forgot about the "conspiracy" between length and angle : $\|x\|_2^2 = \langle x,x\rangle = x^T x$.)
Precisely,

K is a cone, so $\alpha\hat{s} \in K \ \forall \alpha \geq 0$. So $\|\alpha\hat{s}-g\|_2^2$ is minimized by $\alpha=1$.

So
$$\frac{d}{d\alpha} \underbrace{\|\alpha\hat{s}-g\|_2^2}_{\alpha^2\hat{s}^T\hat{s}-2\alpha\hat{s}^Tg+g^Tg}\Big|_{\alpha=1} = 0 \iff 2\alpha\hat{s}^T\hat{s} - 2\hat{s}^Tg\Big|_{\alpha=1} = 0$$
$$\iff \hat{s}^T(\hat{s}-g) = 0.$$

Now, note that K is not just a cone, it is also convex. So:



$$\|\underbrace{(1-\theta)\hat{s}+\theta s}_{=\hat{s}+\theta(s-\hat{s})} - g\|_2^2 \geq \|\hat{s}-g\|_2^2 \quad \forall s \in K,\ \theta \in [0,1].$$

$$\Rightarrow \langle \hat{s}-g+\theta(s-\hat{s}),\ \hat{s}-g+\theta(s-\hat{s})\rangle \geq \langle \hat{s}-g, \hat{s}-g\rangle$$

$$\Rightarrow 2\theta(s-\hat{s})^T(\hat{s}-g) + \theta^2\|s-\hat{s}\|_2^2 \geq 0 \quad \xrightarrow{\theta\downarrow 0} \quad (s-\hat{s})^T(\underbrace{\hat{s}-g}_{d}) \geq 0$$

so $s^Td - \underbrace{\hat{s}^Td}_{=0} \geq 0$, or $s^Td \geq 0 \ \forall s \in K$

Also $d \neq 0$ since $g \notin K$, so $d^Tg = d^T(\hat{s}-d) = 0 - d^Td = -\|d\|_2^2 < 0$.

We have shown that $d^{\perp}$ separates K from g.

So $\quad d^T(By + Cw) \geq 0 \quad \forall\, y \geq 0, w$

set $w = 0$, $\underbrace{d^TBy}_{(B^Td)^Ty} \geq 0 \;\; \forall y \geq 0$. This is only possible if $B^Td \geq 0$

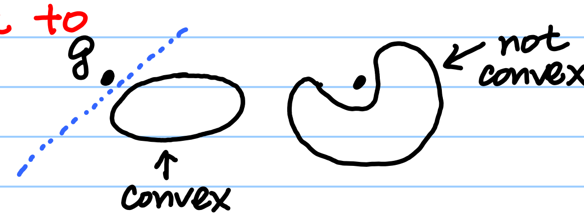
Similarly, set $y=0$, $\underbrace{d^TCw}_{(C^Td)^T} = 0 \;\; \forall w$. This is only possible if $C^Td = 0$.

Q.E.D.

Comments on the proof:

(i) That $K = \{Cy + Bw : y \geq 0\}$ is closed is essential for the proof.

(ii) That K is convex is very essential for the existence of a separating hyperplane. (The existence has little to do with the fact that K is a closed cone.)
This is also why I like the treatment of Farkas' lemma in [Becks] more, except that it is longer.



(iii) We can state Farkas' lemma in a slightly simpler form without losing generality :

Lemma (Farkas)

Given any vectors $g, b_1, \cdots, b_m, \cancel{c_1, \cdots, c_p} \in \mathbb{R}^n$.

We have either :

- $g \in K = \{ \overset{n\times m}{B} y \cancel{+ \overset{n\times p}{C} w} : y \geq 0 \}$, $B = [b_1, \cdots, b_m]$, $\cancel{C = [c_1, \cdots, c_p]}$

OR else

- $\exists\, d \in \mathbb{R}^n$ st $g^T d < 0$, $B^T d \geq 0$, $\cancel{C^T d = 0}$.

But not both.

Why doesn't it lose generality ?

It is because we can always write a real number as the difference of two non-negative numbers, so

$$K = \{ By + Cw : y \geq 0 \} = \{ \underbrace{[B, C, -C]}_{\text{call this the new } B} \begin{bmatrix} y \\ w^+ \\ w^- \end{bmatrix} : \begin{bmatrix} y \\ w^+ \\ w^- \end{bmatrix} \geq 0 \}.$$

the new y

Proof of the KKT theorem:

If x^* is a local solution, the LICQ is satisfied at x^*, then

$$\nabla f(x^*)^T d \geqslant 0, \quad \forall d \in T_\Omega(x^*) = \mathcal{F}(x^*)$$

So by Farkas' lemma,

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*)$$, for some Lagrange multipliers λ_i

$\lambda_i \geqslant 0$ for $i \in \mathcal{I} \cap \mathcal{A}(x^*)$

To complete the proof, all we need is to define the vector λ^* by

$$\lambda^*_i = \begin{cases} \lambda_i & i \in \mathcal{A}(x^*) \\ 0 & i \in \mathcal{I} \setminus \mathcal{A}(x^*). \end{cases}$$

The rest is trivial to check. Q.E.D.