Consider the general constrained problem (smooth functions)

$$\boxed{\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0 \quad i \in \mathcal{E} \\ & c_i(x) \geq 0 \quad i \in \mathcal{I} \end{aligned}} \qquad (*)$$

$\mathcal{E}, \mathcal{I}$ are index sets for identifying the equality and inequality constraint functions, respectively.

If $\mathcal{E} = \phi$ and $\mathcal{I} = \phi$ then we have an unconstrained problem for which we know

(a) x^* is a local minimizer $\Rightarrow$ $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ p.s.d.

(b) $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ p.d. $\Rightarrow$ x^* is a strict local minimizer.

We would like to have similar tests for constrained problems.

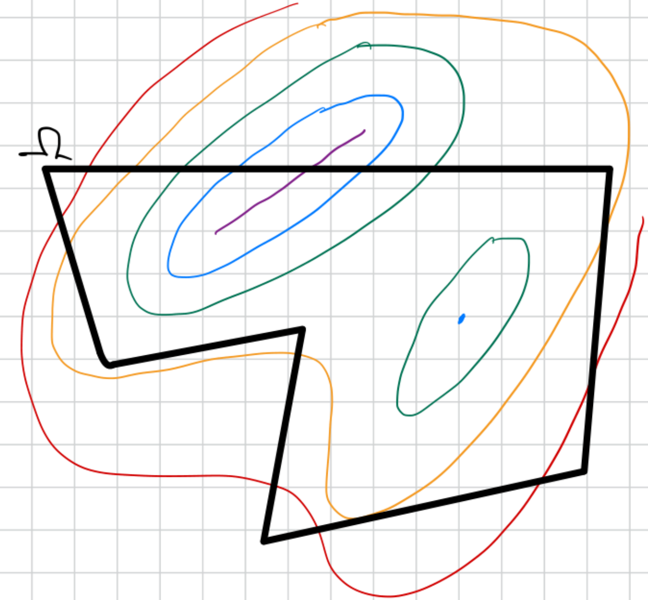
Some Definitions

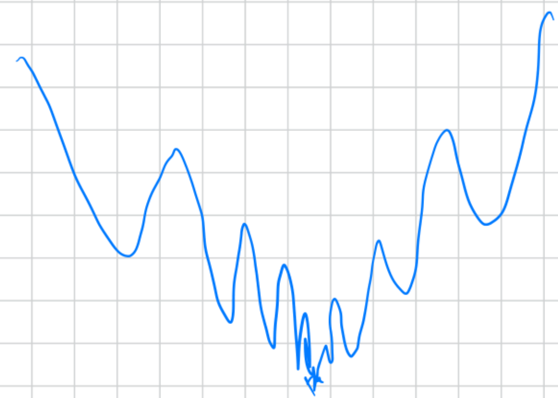
- x^* is a local minimizer of ⊛ if $f(x^*) \le f(x)$ $\forall x \in \Omega \cap B(x^*, r)$ for some $r > 0$.

- x^* is a strict local minimizer of ⊛ if $f(x^*) < f(x)$ $\forall x \in \Omega \cap B(x^*, r)$ $x \neq x^*$, for some $r > 0$.

- x^* is an isolated local minimizer of ⊛ if $\exists$ neighborhood $\Omega \cap B(x^*, r)$, $r > 0$ for which x^* is the unique local minimizer.

- A constraint $c_i(x) = 0$ or $c_i(x) \le 0$ is said to be active at x if $c_i(x) = 0$.





An Example

$$\boxed{\begin{aligned} \min \quad & f(x) = x_1 - x_2 \\ \text{s.t.} \quad & x_2 = x_1^3 \\ & x \in \mathbb{R}^2 \end{aligned}}$$

– or –

$$\boxed{\begin{aligned} \min \quad & f(x) = x_1 - x_2 \\ \text{s.t.} \quad & c(x) = x_1^3 - x_2 = 0 \\ & x \in \mathbb{R}^2 \end{aligned}}$$

For a feasible step s (small enough) at feasible point x, it must be true that $c(x+s) = 0$. So

$$c(x+s) = c(x) + s^T \nabla c(x) + \cdots$$

$$\Rightarrow \boxed{\nabla c(x)^T s = 0}$$

If x is a local minimizer then

$$f(x \pm s) \geq f(x). \quad \text{So}$$

$$f(x \pm s) = f(x) \pm s^T \nabla f(x) + \cdots$$

$$\Rightarrow \pm \nabla f(x)^T s \geq 0$$

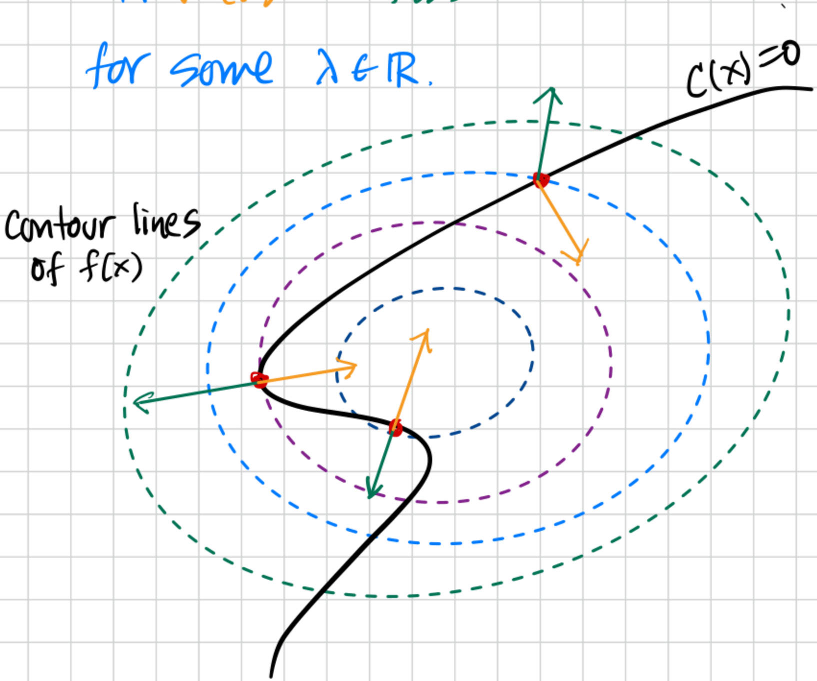
$$\Rightarrow \boxed{\nabla f(x)^T s = 0}$$

So, for a single constraint, the space of vectors perpendicular to $c(x) = 0$ is dimension 1. Thus, $\nabla f(x)$ and $\nabla c(x)$ are colinear.

If x is a local minimizer then

$$\lambda \nabla c(x) = \nabla f(x)$$

for some $\lambda \in \mathbb{R}$.



We define the Lagrangian as

$$L(x,\lambda) = f(x) - \lambda c(x).$$

Then we recover the first order necessary conditions as

$$\nabla L(x,\lambda) = 0.$$

In particular:

$$\nabla_x L = \nabla f(x) - \lambda \nabla c(x) = 0$$

$$\Rightarrow \boxed{\nabla f(x) = \lambda \nabla c(x)}$$

$$\nabla_\lambda L = -c(x) = 0$$

$$\Rightarrow \boxed{c(x) = 0}$$

For the current example

$$L(x,\lambda) = x_1 - x_2 - \lambda(x_1^3 - x_2)$$

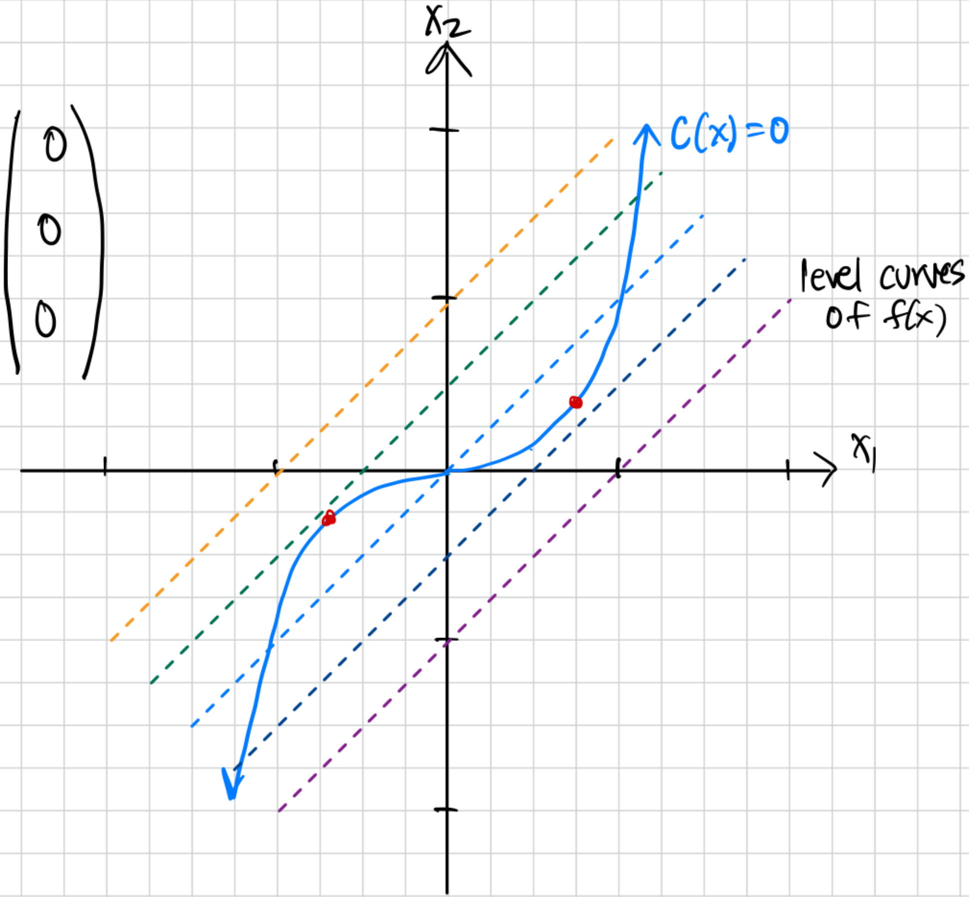
$$\nabla L(x,\lambda) = \begin{pmatrix} 1 - 3\lambda x_1^2 \\ -1 + \lambda \\ -x_1^3 + x_2 \end{pmatrix} \overset{\text{set}}{=} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

We find two solutions:

$$x_1 = \pm\sqrt{1/3}$$

$$x_2 = \pm\left(\sqrt{1/3}\right)^3$$

$$\lambda = 1$$

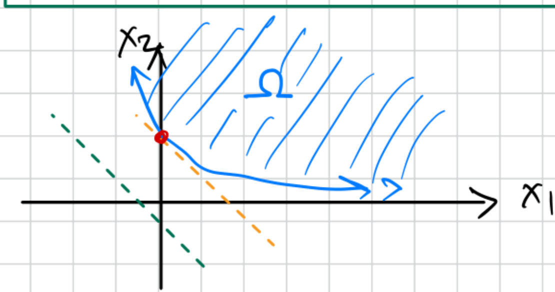


A Second Example

$$\boxed{\begin{array}{ll} \min & f(x) = x_1 + x_2 \\ \text{s.t.} & x_2 \geq e^{-x_1} \\ & x \in \mathbb{R}^2 \end{array}}$$

- or -

$$\boxed{\begin{array}{ll} \min & f(x) = x_1 + x_2 \\ \text{s.t.} & c(x) = x_2 - e^{-x_1} \geq 0 \\ & x \in \mathbb{R}^2 \end{array}}$$



Let's look at this case slighty differently

A feasible descent direction s to first order must satisfy

$$\nabla f(x)^T s < 0 \quad \text{and} \quad \nabla c(x)^T s \geq 0$$

So for there to be no feasible descent directions, it must be true that

$$\boxed{\nabla f(x) = \lambda \nabla c(x), \quad \lambda \geq 0}$$

(∇f and ∇c are not just colinear, but point in the same direction)

OK, let's make sure that these conditions work as desired.

- If $c(x)$ is not active at x then it must follow that $\lambda = 0$ at a stationary point. $(\nabla f(x) = 0)$
- If $c(x)$ is active at x then $\lambda \geq 0$.

The way to guarantee that λ follows these requirements is the complementarity condition

$$\boxed{\lambda c(x) = 0}$$

We have the first order necessary conditions:

$$\boxed{\begin{aligned} \nabla f(x) &= \lambda \nabla c(x) \\ c(x) &\geq 0 \\ \lambda &\geq 0 \\ \lambda c(x) &= 0 \end{aligned}}$$

$(\nabla_x L(x, \lambda) = 0)$

For our example problem...

$$\nabla_x L(x,\lambda) = \begin{pmatrix} 1-\lambda e^{-x_1} \\ 1-\lambda \end{pmatrix} \overset{\text{set}}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\nabla_\lambda L(x,\lambda) = x_2 - e^{-x_1} \geq 0$$

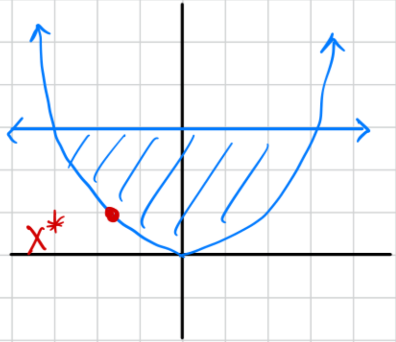
$$\lambda \geq 0$$

$$\lambda\left(x_2 - e^{-x_1}\right) = 0$$

$$\rightarrow \boxed{\begin{aligned} \lambda &= 1 \\ x_1 &= 0 \\ x_2 &= 1 \end{aligned}}$$

EXAMPLE

$$\min \quad x_1 + x_2$$
$$\text{s.t.} \quad x_2 \geq x_1^2$$
$$x_2 \leq 1$$
$$x \in \mathbb{R}^2$$



$$L(x,\lambda) = x_1 + x_2 - \lambda_1(-x_1^2 + x_2) - \lambda_2(1 - x_2)$$

$$\min f(x) = x_1 + x_2$$
$$\text{s.t.} \quad c_1(x) = -x_1^2 + x_2 \geq 0$$
$$c_2(x) = 1 - x_2 \geq 0$$
$$x \in \mathbb{R}^2$$

$$\nabla_x L = \begin{pmatrix} 1 + 2\lambda_1 x_1 \\ 1 - \lambda_1 + \lambda_2 \end{pmatrix} \overset{set}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$-x_1 + x_2 \geq 0$$

$$1 - x_2 \geq 0$$

$$\lambda_1 \geq 0 \qquad \lambda_2 \geq 0$$

$$\lambda_1(-x_1^2 + x_2) = 0 \qquad \lambda_2(1 - x_2) = 0$$

Solution: $\lambda^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad x^* = \begin{pmatrix} -1/2 \\ 1/4 \end{pmatrix}$

Tangent Cones

Before we make a general claim about first order necessary conditions we need to consider carefully the nature of the active constraints.

Thus far our examples have worked out well. But finding potential descent directions is a very geometric concept that may or may not be accurately represented by linear constraint approximations.

Do the active constraint gradients $\nabla c_i(x)$ accurately describe the feasible region geometry near x?

example: $f(x) = x_2$

$$c_1(x) = -x_2 \geq 0$$

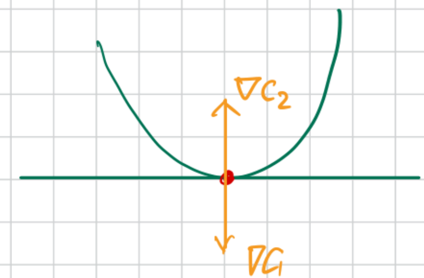
$$c_2(x) = x_2 - x_1^2 \geq 0$$

The sole feasible point is $x = (0,0)$.

The linearized feasible set at x is:

$$\{d = (d_1, d_2) \mid \nabla c_1^T d \geq 0, \nabla c_2^T d \geq 0\}$$

$$= \{d = (d_1, 0) \mid d_1 \in \mathbb{R}\}$$



lin. feas. set not geometrically correct

Let $\Omega \subseteq \mathbb{R}^n$, $x \in \Omega$.

Def: $\{z_k\}$ is said to be a feasible sequence approaching x if $z_k \in \Omega$ for all k sufficiently large and $z_k \to x$.

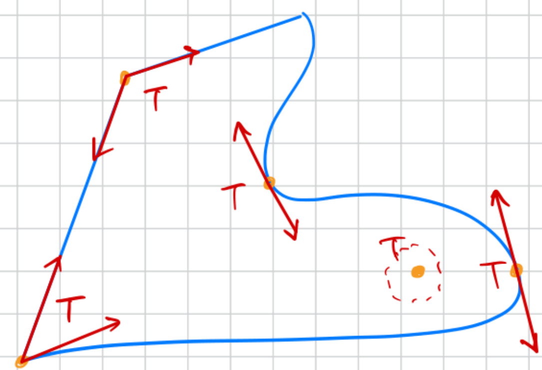
Def: A vector $d \in \mathbb{R}^n$ is a tangent vector to Ω at x if there is a feasible sequence $\{z_k\}$ approaching x and a sequence of positive scalars $\{t_k\}$ with $t_k \to 0$ such that

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

The set of all tangent vectors to Ω at x is called the tangent cone $T_\Omega(x)$.

Def: The set of linearized feasible directions of Ω at x is

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

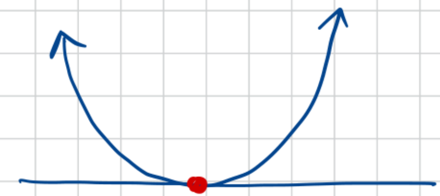


Theorem. If x is a local minimizer then $\nabla f(x)^T d \geq 0$ for all $d \in T_\Omega(x)$.

This is a beautiful result. It also emphasizes the potential difference between $T_\Omega(x)$ and $F_\Omega(x)$.

Example:

$$\Omega = \{ x \in \mathbb{R}^n \mid x_2 \geq x_1^2 ,\ x_2 \leq 0 \}$$



$x = (0,0)$ is the only feasible point

$$\boxed{T_\Omega(0,0) = \vec{0}}$$

$c_1(x) = -x_1^2 + x_2 \qquad \nabla c_1(x) = \begin{pmatrix} -2x_1 \\ 1 \end{pmatrix} \qquad \nabla c_1(0,0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$c_2(x) = -x_2 \qquad \nabla c_2(x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \qquad \nabla c_2(0,0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

linearized feasible directions d satisfy $d^T c_i(0,0) \geq 0$

$$\Rightarrow \quad F_\Omega(x) = \{ d = (d_1, 0) \in \mathbb{R}^2 \mid d_1 \in \mathbb{R} \}.$$

Def: Given $x \in \Omega$ and active set $A(x)$.

we say that the linear independence constraint qualification (LICQ) holds if $\{ \nabla c_i(x) \mid i \in A(x) \}$ is linearly independent.

$\boxed{\begin{array}{l} \textbf{Fact: } \text{Let } x \in \Omega \text{ then} \\ \quad \text{(a) } T_\Omega(x) \subseteq F_\Omega(x) \\ \quad \text{(b) If LICQ holds then } T_\Omega(x) = F_\Omega(x). \end{array}}$

First Order Necessary Conditions

Theorem (KKT) Suppose x is a local optimum, f and all c_i are continuously differentiable, and LICQ holds at x. Then there exists λ such that the following hold.

or, more generally $T_\Omega(x) = F_\Omega(x)$

(a) $\nabla_x L(x, \lambda) = 0$

(b) $c_i(x) = 0, \quad i \in \mathcal{E}$

(c) $c_i(x) \geq 0, \quad i \in \mathcal{I}$

(d) $\lambda_i \geq 0, \quad i \in \mathcal{I}$

(e) $\lambda_i c_i(x) = 0, \quad i \in \mathcal{E} \cup \mathcal{I}$

Another geometric connection

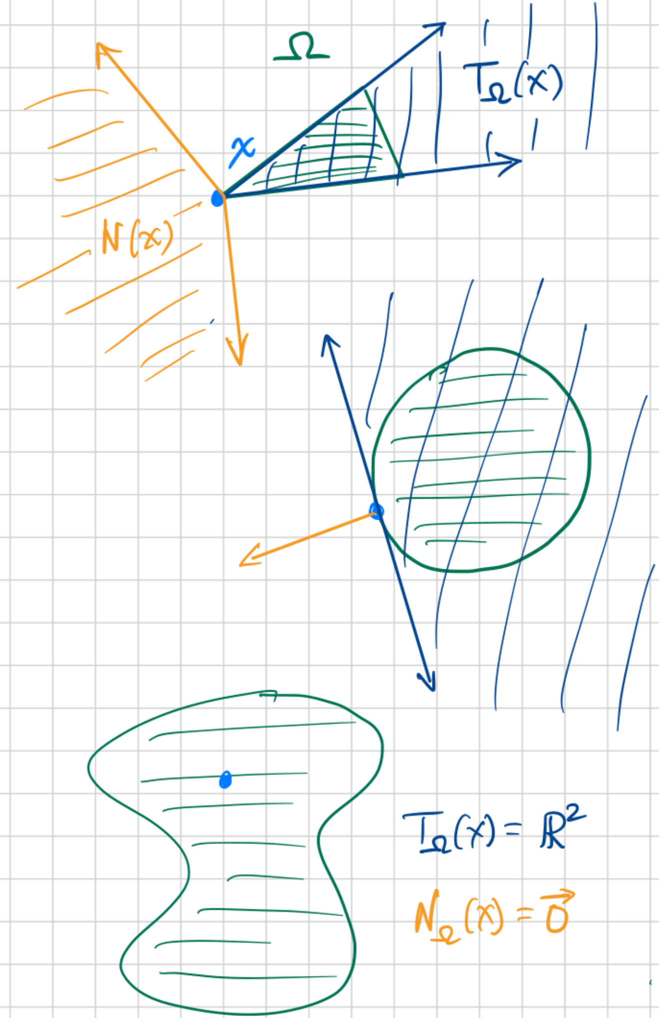
Def: The normal cone $N_\Omega(x)$ to Ω at $x \in \Omega$ is

$$N_\Omega(x) = \{ u \mid u^T w \leq 0 \ \forall\, w \in T_\Omega(x) \}$$

Each such u is called a normal vector.

> **Theorem**: Suppose x is a local minimizer. Then $-\nabla f(x) \in N_\Omega(x)$.

> **Lemma**: Suppose LICQ holds at x. Then $N_\Omega(x)$ is the cone given by the set of all conic combinations of $-\nabla c_i(x)$, $i \in \mathcal{A}(x)$.



More on Constraint Qualifications.

Remember the goal is to find conditions under which $T_\Omega(x) = F_\Omega(x)$.

Here are three sufficient conditions:
(If any one holds then $T_\Omega(x) = F_\Omega(x)$)

(a) LICQ

(b) All active constraints are linear

(c) MFCQ

Def: We say that MFCQ holds at $x \in \Omega$ if there exists w satisfying the following three conditions:

(1) $\nabla c_i(x)^T w > 0$, $i \in \mathcal{A}(x) \cap \mathcal{I}$

(2) $\nabla c_i(x)^T w = 0$, $i \in \mathcal{E}$

(3) $\{\nabla c_i(x), i \in \mathcal{E}\}$ is L.I.