↳ First order Necessary Conditions (FONC)

Karush Kuhn Tucker (KKT) Conditions.

As before $\mathcal{L}(x,\lambda) = f(x) - \sum_i \lambda_i c_i(x)$

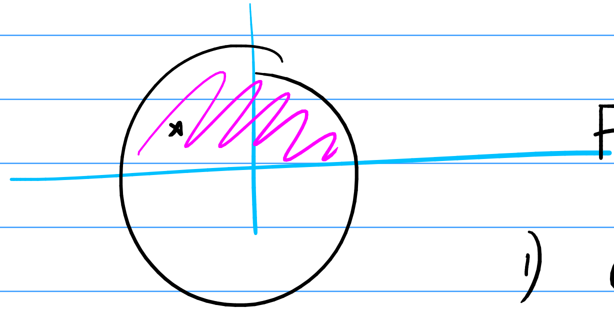
$i \in \mathcal{E} \cup \mathcal{I}$.

If	Then: There is a set of Lagrange multipliers s.t. the foll are satisfied at (x^*, λ^*):
① If x^* is a local soln to $\min_{x \in \Omega} f(x)$.	a) $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$
② f and c_i's are continuously differentiable	b) $c_i(x^*) = 0$, $i \in \mathcal{E}$
	c) $c_i(x^*) \geq 0$, $i \in \mathcal{I}$
③ LICQ holds at x^*	d) $\lambda_i \geq 0$, $i \in \mathcal{I}$
	e) $\lambda_i c_i = 0$, $i \in \mathcal{E} \cup \mathcal{I}$

There can be many pts where a–e hold. But if ③ holds λ^* is unique.

↳ Proof sketch. $f(x) = x_1 + x_2$ s.t. $2 - x_1^2 - x_2^2 \geq 0$

$x_2 \geq 0$



FONC for x^* to be a local min

1) $c_1(x^*) \geq 0$, $c_2(x^*) \geq 0$

2) a) If $c_1(x^*) = 0$, $c_2(x^*) = 0$, then

$$\nabla_x \mathcal{L} = \nabla f(x^*) - \lambda_1 \nabla c_1 - \lambda_2 \nabla c_2 = 0, \quad \lambda_1 \geq 0, \ \lambda_2 \geq 0$$

b) $c_1(x^*) > 0$, $c_2(x^*) = 0$, then $\Rightarrow \lambda_1 = 0$

$$\nabla_x \mathcal{L} = \nabla f(x^*) - \lambda_2 \nabla c_2(x^*) = 0, \quad \lambda_2 \geq 0.$$

c) $c_1(x^*) = 0$, $c_2(x^*) > 0$, then $\Rightarrow \lambda_2 = 0$

$$\nabla_x \mathcal{L} = \nabla f(x^*) - \lambda_1 \nabla c_1(x^*) = 0, \quad \lambda_1 \geq 0$$

d) $c_1(x^*) > 0$, $c_2(x^*) > 0$, then $\lambda_1 = 0$, $\lambda_2 = 0$

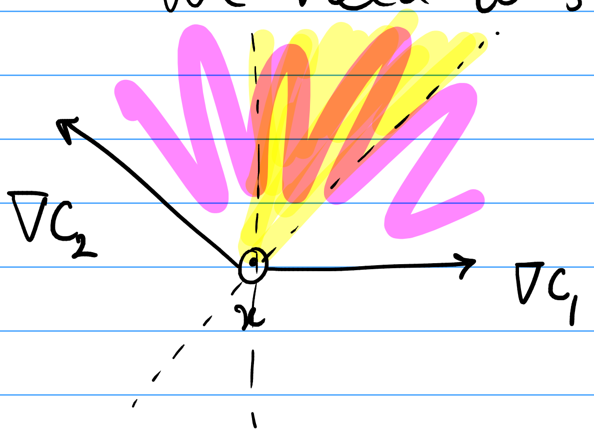
$$\nabla_x \mathcal{L} = \nabla f(x^*) = 0.$$

For (a) we need to show? $\nabla_x \mathcal{L}(x, \lambda) = 0$

$\Rightarrow \nabla f(x) = \lambda_1 \nabla c_1(x) + \lambda_2 \nabla c_2(x) \quad , \quad \lambda_1 \geq 0, \ \lambda_2 \geq 0$

Define $A = \left\{ u \in \mathbb{R}^2 \;\middle|\; u = \lambda_1 \nabla c_1 + \lambda_2 \nabla c_2 \text{ for } \lambda_1, \lambda_2 \geq 0 \right\}$

We need to show $\nabla f \in A$.

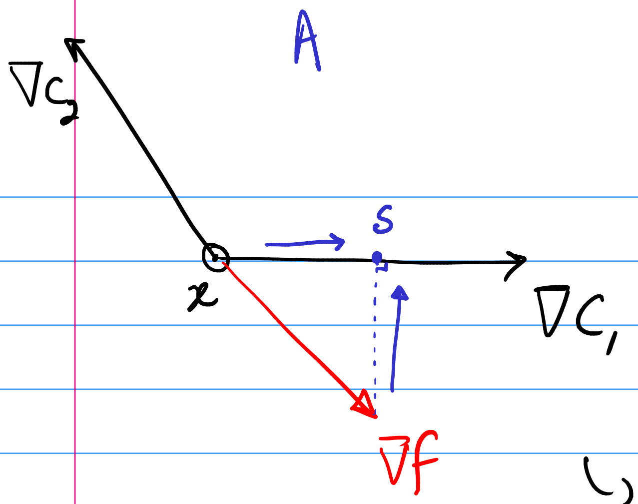


feasible directions:

$\nabla c_1^T d \geq 0 \quad , \quad \nabla c_2^T d \geq 0$

Proof by contradiction:

i.e. $\nabla f \notin A$.



$$s = \underset{w \in A}{\arg\min} \| w - \nabla f \|$$

(Claim: $(s - \nabla f) \longrightarrow$ feasible & descent direction)

↳ If this is true, then x is not a local min of the problem, since f can be reduced further by going along $s - \nabla f$.

a) $\|s\| < \|\nabla f\|$ since A is convex AND s is the proj of ∇f.

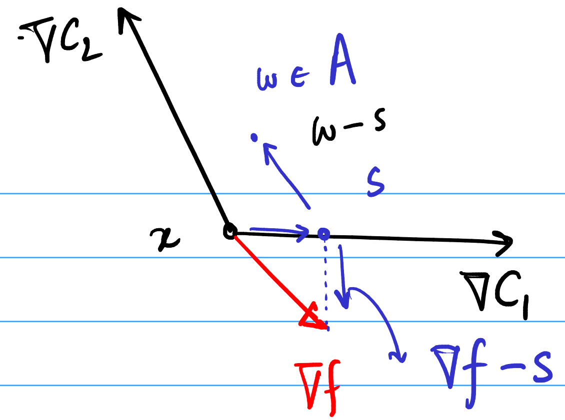
$$\|\nabla f\| \|s\| < \|\nabla f\|^2 = \nabla f^T \nabla f$$

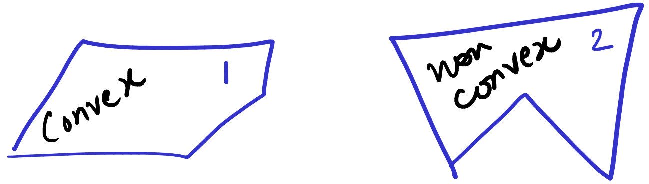
$$\|\nabla f\| \|s\| \cos\theta \leq \|\nabla f\| \|s\|$$

$$s^T \nabla f < \nabla f^T \nabla f \longrightarrow (s - \nabla f)^T \nabla f < 0$$

$\Rightarrow (s - \nabla f)$ is a descent direction!

(b)





By property of convex sets:

$$(w-s)^T(\nabla f - s) \leq 0$$

$$(w-s)^T(s - \nabla f) \geq 0$$

Say we choose $w = s + \nabla C_1 \in A$

$$\Rightarrow \nabla C_1^T (s - \nabla f) \geq 0$$

Say we choose $w = s + \nabla C_2 \in A$

$$\Rightarrow \nabla C_2^T (s - \nabla f) \geq 0$$

$\Rightarrow s - \nabla f$ is feasible!

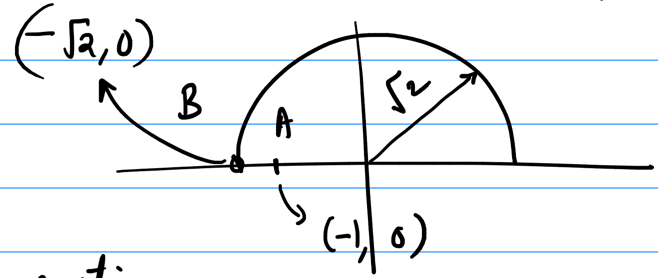
$\Rightarrow$ Going along $s - \nabla f$ will give me a better point that x. $\Rightarrow x$ is not a local min

$\Rightarrow$ Contradicts pt 1 of KKT thm.

$\Rightarrow \nabla f \in A$, Rest of the case can be done similarly. (try b or c as exercise).

↳ $f(x) = x_1 + x_2$, $\quad c_1(x) = 2 - x_1^2 - x_2^2 \geq 0$, $\quad c_2(x) = x_2 \geq 0$

$\nabla c_1 = -2\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\nabla c_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$



Pt. A

$c_1(x_A) = 1 > 0$ inactive.

$c_2(x_A) = 0 = 0$ active

$\mathcal{L}(x,\lambda) = f(x) - \lambda_2 c_2(x)$, $\lambda_1 = 0$

If x_A is optimal $\Rightarrow \nabla_x \mathcal{L} = 0$

$$\nabla f = \lambda_2 \nabla c_2$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda_2 \end{pmatrix} \Rightarrow \text{No value of } \lambda_2 \text{ works}$$

$\Rightarrow x_A$ is not optimal.

pt B.

$\left.\begin{array}{l} c_1(x_B) = 0 \\ c_2(x_B) = 0 \end{array}\right\}$ active

$\nabla f = \lambda_1 \nabla c_1 + \lambda_2 \nabla c_2 \rightarrow \nabla c^T \lambda$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

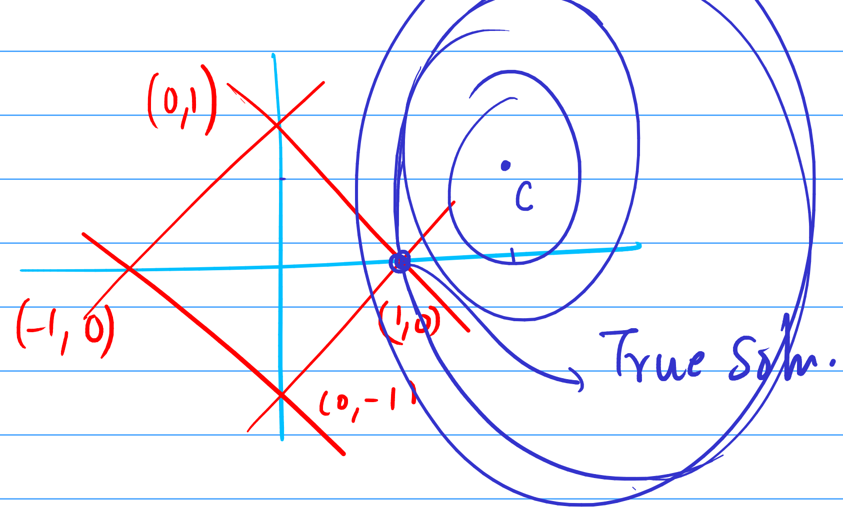
$\Rightarrow \underline{\lambda_1 = 1/2\sqrt{2}, \ \lambda_2 = 1}$

$\lambda_1, \lambda_2 \geq 0$

∴ Satisfies KKT

∴ optimal pt.

↳ Another example: $f(x) = \left(x_1 - 3/2\right)^2 + \left(x_2 - 1/2\right)^4$



$$\text{s.t.}\quad \left.\begin{aligned} c_1 &= 1 - x_1 - x_2 \\ c_2 &= 1 - x_1 + x_2 \\ c_3 &= 1 + x_1 - x_2 \\ c_4 &= 1 + x_1 + x_2 \end{aligned}\right\} \geq 0$$

Pt to test is $(1, 0)$

① Which constraints active $c_1 = 0$, $c_2 = 0$, $\cancel{c_3}$, $\cancel{c_4}$.

② L. Mult $\rightarrow$ $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\lambda_3 = 0$, $\lambda_4 = 0$.

③ $\nabla_x \mathcal{L} = 0 \rightarrow \mathcal{L} = f - \lambda_1 c_1 - \lambda_2 c_2 \Rightarrow \nabla f = \lambda_1 \nabla c_1 + \lambda_2 \nabla c_2$

$(1,0)$

$$\begin{pmatrix} -1 \\ -1/2 \end{pmatrix} = \begin{pmatrix} 2(x_1 - 3/2) \\ 4(x_2 - 1/2)^3 \end{pmatrix} = \lambda_1 \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$$

$$\left.\lambda_1 = \frac{3}{4}, \quad \lambda_2 = \frac{1}{4}\right) \geq 0 \ \checkmark$$

Projected Gradient Descent (PGD)

↳ Problem: $\min_{x \in \Omega} f(x)$, e.g.: $\min_{\|x\|_2 \leq 1} \|Ax - b\|^2$

↳ Recap of GD: ① Pick a starting pt $x_0 \in \mathbb{R}^n$

② loop till satisfied

↳ find $-\nabla f$, find α_k

↳ update $x_{k+1} = x_k - \alpha_k \nabla f_k$

↳ PGD is a small modification. ⟶ $x_{k+1} = P_\Omega(x_k - \alpha_k \nabla f_k)$

$P_\Omega(\,)$ is a projection operator:

$$P_\Omega(x^\circ) = \underset{x \in \Omega}{\operatorname{argmin}} \frac{1}{2} \|x - x^\circ\|_2^2 ,$$

$$P_\Omega : \mathbb{R}^n \longrightarrow \mathbb{R}^n .$$