Constrained Optimization

$$\text{Min } f(x), \quad x \in \mathbb{R}^n$$

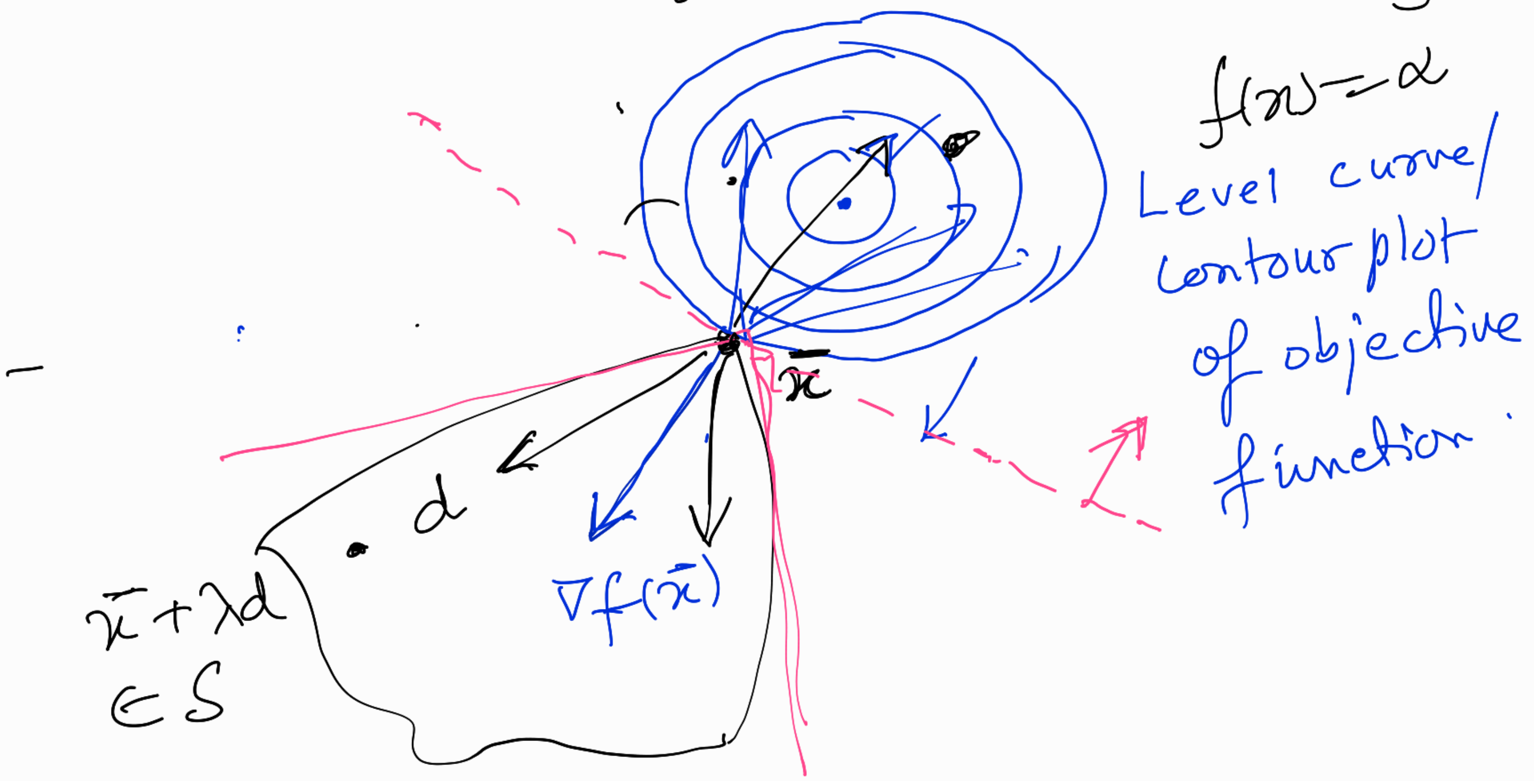
$$X \quad \text{s.t.} \begin{cases} g_i(x) \leq 0, & i = 1 \ldots m \text{ (convex)} \\ h_j(x) = 0, & j = 1 \ldots p \text{ (affine)} \end{cases}$$

A constrained problem in ML:

Soft margin SVM

$$\min_{\xi, w} \mathcal{L}(\xi, w) = \frac{1}{2}\|w\|_2^2$$

$$\text{s.t.} \left.\begin{aligned} & y_i(w^T x_i - b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned}\right\} \text{"}S\text{"}$$



S (feasible region)

$$D = \{ \vec{d} \mid \bar{x} + \lambda d \in S, \lambda > 0 \}$$

is the set of feasible directions

$$F_0 = \{ \vec{d} \mid \nabla f(\bar{x})^T d < 0 \}$$ "half space"

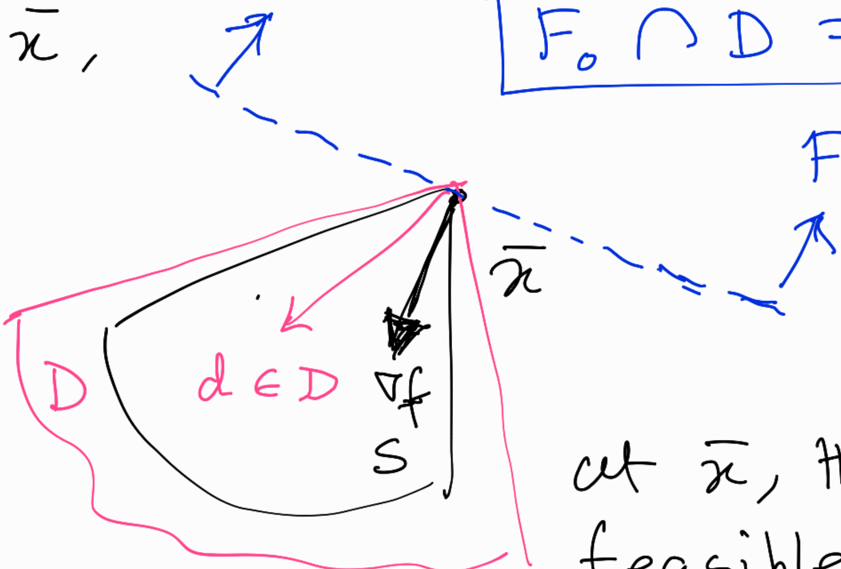
is the set of "improving" directions.

Since the objective is to minimize $f(x)$

$\nabla f(\bar{x})^T d < 0$ means function value reduces if we move away from $\bar{x}$ along the direction d, and hence objective is said to "improve".

At $\bar{x}$,

$$\boxed{F_0 \cap D = \phi}$$



This means at $\bar{x}$, there is no feasible improving direction, so $\bar{x}$ is optimal.

Thm 4.2.5 : If $\bar{x}$ is local minima, then $F_0 \cap D = \phi$

Example : $S = \{(x_1, x_2) \mid x_2 = x_1^2\}$

and $f(x) = x_2$

At $\bar{x} = (1,1)$, do we have $S \cap F_0 = \phi$

$F_0 = \{ d \mid \nabla f(\bar{x})^T d < 0 \}$

$\nabla f(x) = [0, 1]$, $\nabla f(\bar{x} = [1, 1]) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

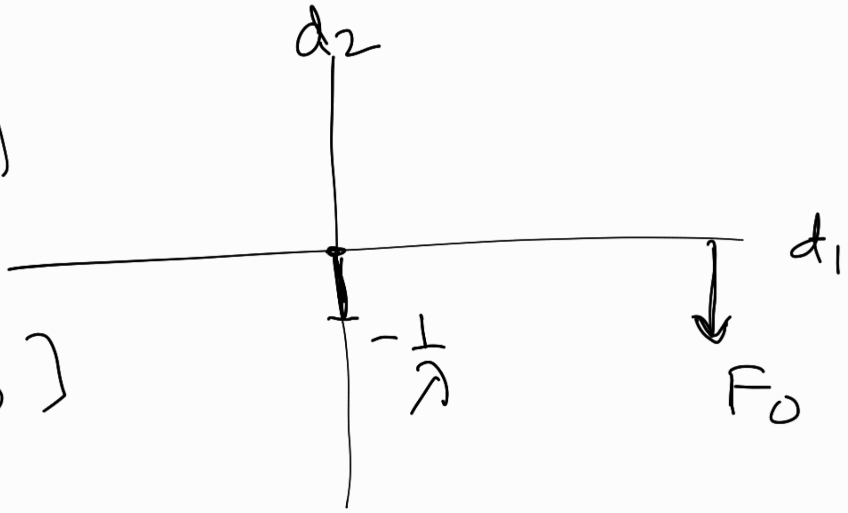
$F_0 = \{ (d_1, d_2) \mid (d_1, d_2)^T (0, 1) < 0 \}$

$\Rightarrow F_0 = \{ (d_1, d_2) \mid \underline{d_2 < 0} \}$

$D: \{d_1, d_2\}$

$\lambda d_2 + 1 \geq 0$

$d_2 \in [-\frac{1}{\lambda}, 0]$



(d_1, d_2):

$\underline{d_2 < 0}$

$S = \{ (x_1, x_2) \mid x_2 = x_1^2 \}$

$D = \{ d \mid (x_2 + \lambda d_2) = (x_1 + \lambda d_1)^2 \}$, $\underline{\lambda > 0}$

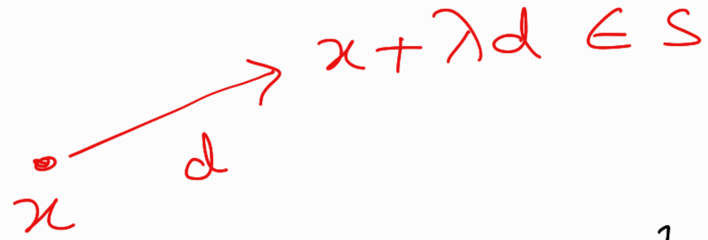
It is clear that $\underline{d_2 > 0}$ and d_1 can be both +ve or -ve.

So, there cannot be any cone D where $d_2 < 0$, i.e, $D = \phi$

Therefore at $\bar{x} = (1,1)$, we have

$$D \cap F_0 = \phi.$$

Suppose the constraints $g_i(x) \leq 0$ are differentiable and pseudo convex



Say $S = \{x \mid g_i(x) \leq 0, i = 1 \cdots m\}$

if $x + \lambda d \in S$, then $g_i(x + \lambda d) \leq 0, i = 1 \cdots$

Because d is improving direction,

$$g_i(x + \lambda d) - g_i(x) \leq 0$$

$$\Rightarrow \quad \frac{g_i(x + \lambda d) - g_i(x)}{\lambda} \leq 0$$

as $\lambda \to 0$, we get $\frac{g_i(x+\lambda d) - g_i(x)}{\lambda}$ becomes the directional derivative

$$\nabla g_i(x)^T d < 0$$

Define $G_0 = \{ d \mid \underbrace{\nabla g_i(\bar{x})^T d}_{\text{directional derivative of } g_i(\bar{x})} < 0, \forall i \}$

Example : 4.2.6 (Bazaraa)

$$\text{Min } (x_1-3)^2 + (x_2-2)^2 \equiv f$$

$$\text{s.t. } x_1^2 + x_2^2 \le 5 \quad (g_1)$$

$$x_1 + x_2 \le 3 \quad (g_2)$$

$$x_1, x_2 \geq 0$$

Is $x = (9/5, 6/5)$ local minima?

If $x = (9/5, 6/5)$ is minima then $F_0 \cap G_0 = \phi$

* If $F_0 \cap G_0 \neq \phi$ then x is not local minima

At $\bar{x} = \left(\frac{9}{5}, \frac{6}{5}\right)$, only $g_2(x)$ is binding

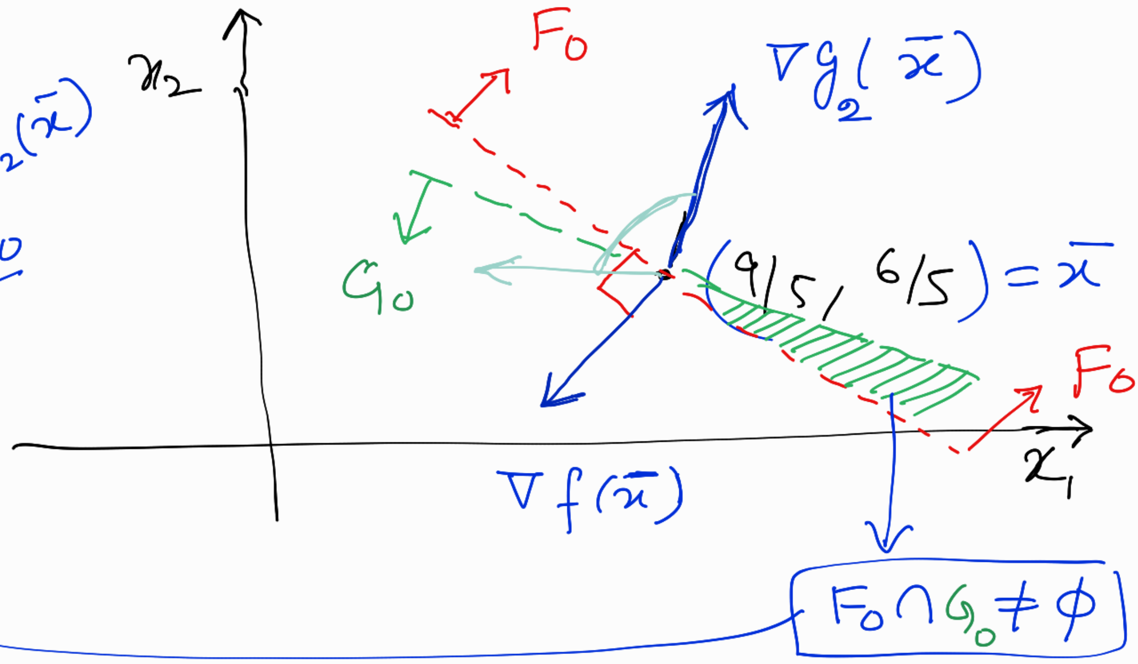
because $g_1(x) = 4.7 < 5$,

Find $\nabla f(x)$ at $(9/5, 6/5) = \bar{x}$

$$\nabla f(\bar{x}) = \left(-\frac{12}{5}, -\frac{8}{5}\right), \quad \nabla g_2(\bar{x}) = (1, 1)$$

$F_0 = \{ d \mid \nabla f(\bar{x})^T d < 0 \}$ half space

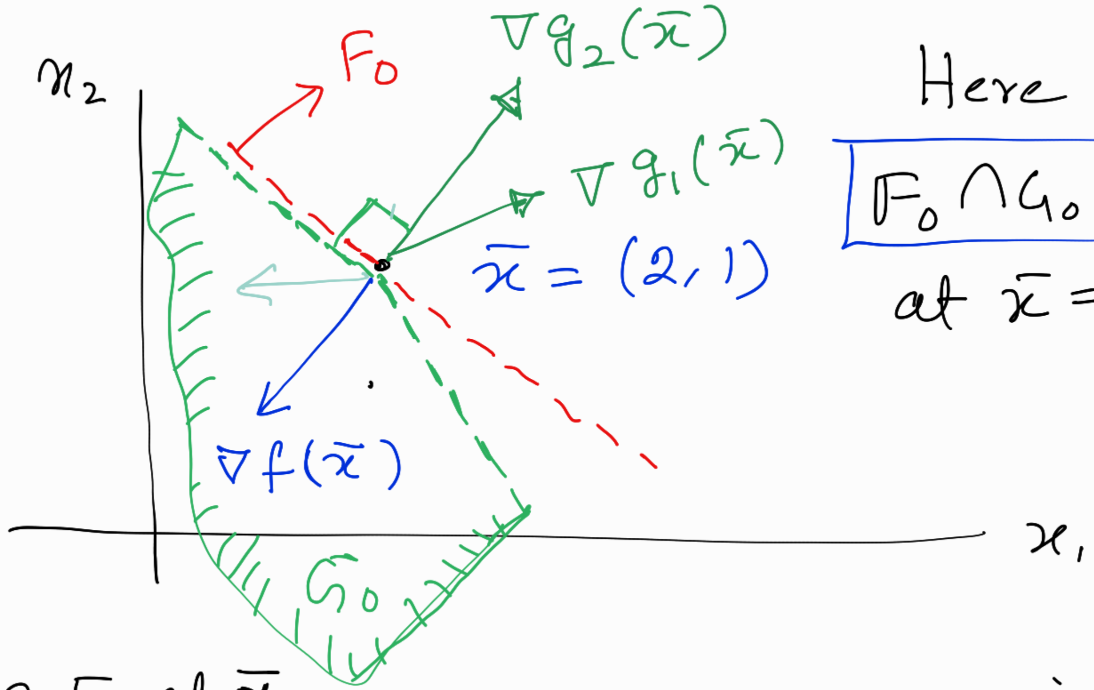
We cannot write $\nabla f(\bar{x}) = \alpha \nabla g_2(\bar{x})$, $\alpha > 0$



Now check at $\bar{x} = (2,1)$

Here $g_1(\bar{x}) = 5$ and $g_2(\bar{x}) = 3$, so both are binding.

$\nabla f(\bar{x}) = (-2,-2)$, $\nabla g_1(\bar{x}) = (4,2)$ and $\nabla g_2(\bar{x}) = (1,1)$



Here $\boxed{F_0 \cap G_0 = \phi}$ at $\bar{x} = (2,1)$

Theorem 4.2.5 at $\bar{x}$

If $F_0 \cap G_0 = \phi$ and f is pseudoconvex at $\bar{x}$ and $g_i(\bar{x})$, $i = 1, 2 \ldots m$ are strictly pseudoconvex then $\bar{x}$ is local minima

Example 4.2.7 (Bazaraa)

$$\text{Min } f(x) = (x_1-1)^2 + (x_2-1)^2$$
$$\text{s.t. } (x_1+x_2-1)^3 \le 0, \quad x_1, x_2 \ge 0$$

Let $\bar{x} = \left(\frac{1}{2}, \frac{1}{2}\right)$, check $F_0 \cap G_0$

Let $g_1(x) \equiv (x_1+x_2-1)^3 \le 0$

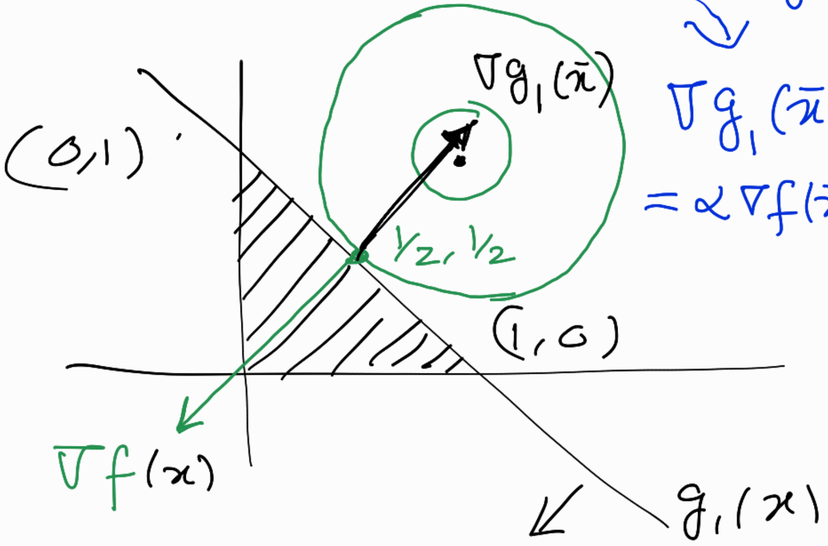
$$\Rightarrow g_1(x) \equiv (x_1+x_2) \le 1$$

$\nabla f(\bar{x}) = (-1,-1)$, $\nabla g_1(\bar{x}) = (1,1)$

we can say $\nabla g_1(\bar{x}) = \alpha \nabla f(\bar{x})$

So, $\boxed{F_0 \cap G_0 = \phi}$

Is $\bar{x}$ minima?



f is quadratic hence, convex, hence, pseudoconvex

$g_1(x) \equiv x_1+x_2 \le 1$ is affine, hence convex, hence st. pseudoconvex.
So, all conditions of Thm 4.2.5.

Observations from the above examples

① If $F_0 \cap G_0 \ne \phi$ we cant have ∇f and ∇g collinear.

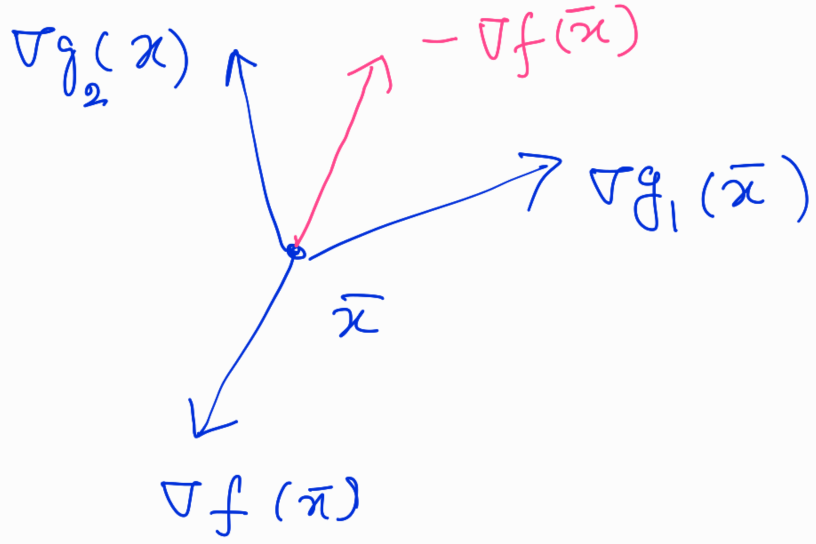
① If $F_0 \cap G_0 = \phi$, then we can express $\nabla f(x)$ in terms of a linear combination of $\nabla g_i(x)$, $i = 1, 2 \ldots m$.

Fritz-John (Necessary) conditions (Thm 4.2.8)

If $\bar{x}$ is a local minima, then

$$\boxed{u_0 \nabla f(\bar{x}) + \sum_{i \in I} u_i \nabla g_i(\bar{x}) = 0}$$

$u_0, u_i \geq 0$, (not all zeros)



equivalently

$$-\nabla f(\bar{x}) = \sum_{i \in I} \frac{u_i}{u_0} \nabla g_i(\bar{x})$$

negative gradient is a linear combination of gradients of binding constraints $(i \in I)$

Lagrange Multipliers

$$u_0 \nabla f(\bar{x}) + \sum_{i \in I} u_i \nabla g_i(\bar{x}) = 0$$

$$u_0, u_i \geq 0, \; i \in I \quad (\text{not all zero})$$

The multipliers $u_0, u_i, i \in I$ .. are called Lagrange multipliers. and the Fritz-John condition stated above is called Dual Feasibility condition.

Now, note that I is the set of binding constraints ie, $g_i(\bar{x}) = 0$

for any $k \notin I$, we have $g_k(\bar{x}) < 0$

Now we can write the Fritz-John condition as

$$\underbrace{u_0 \nabla f(\bar{x}) + \sum_{i \in I} u_i \nabla g_i(\bar{x})}_{=0} + \sum_{k \notin I} u_k \nabla g_k(\bar{x}) = 0$$

only when $u_k = 0, \; k \notin I$

So,

$$u_i = \begin{cases} u_i & \text{if } i \in I \text{ (binding) } \; g_i(\bar{x}) = 0 \\ 0 & \text{if } i \notin I \text{ (nonbinding) } \; g_i(\bar{x}) < 0 \end{cases}$$

So, $\quad u_i g_i(\bar{x}) = 0, \quad i = 1, 2, \ldots m$

So, now we can include all constraints in the Fritz-John conditions as

$$\boxed{\begin{aligned} & u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) = 0 \\ & u_i g_i(\bar{x}) = 0, \quad i = 1, 2, \ldots m \\ & u_0, u_i \geq 0, \quad i = 1, 2 \ldots m \\ & (u_0, u) \neq 0 \quad (\text{not all are zero}) \end{aligned}}$$

FRITZ – JOHN CONDITIONS

Example 4.2.9 $\quad \text{Min } f(x) = (x_1 - 3)^2 + (x_2 - 2)^2$

$$\begin{aligned} \text{s.t } \; & x_1^2 + x_2^2 \leq 5 \quad (g_1) \\ & x_1 + 2x_2 \leq 4 \quad (g_2) \\ & x_1, x_2 \geq 0 \end{aligned}$$

Check if the Fritz-John conditions hold at $\bar{x} = (2, 1)$

At $\bar{x}$, only $\underbrace{g_1(\bar{x}) = 0 \text{ and } g_2(\bar{x}) = 0}_{\text{Binding}}$

$\nabla f(\bar{x}) = (-2,-2)$

$\nabla g_1(\bar{x}) = (4,2)$, $\nabla g_2(\bar{x}) = (1,2)$

$$u_0\begin{pmatrix}-2\\-2\end{pmatrix} + u_1\begin{pmatrix}4\\2\end{pmatrix} + u_2\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

$\Rightarrow u_1 = u_0/3$, $u_2 = 2u_0/3$

So, for any $u_0 \geq 0$, $u_1, u_2 \geq 0$, so the conditions are satisfied.

Now check at $\bar{x} = (0,0)$. Here $g_1(x)$ and $g_2(x)$ are not binding, but $g_3(x) \equiv -x_1 \leq 0$ and $g_4(x) \equiv -x_2 \leq 0$ are both binding.

At $\bar{x} = (0,0)$, $\nabla f(\bar{x}) = (-6,-4)$

So, the condition is

$$u_0\begin{pmatrix}-6\\-4\end{pmatrix} + u_3\begin{pmatrix}-1\\0\end{pmatrix} + u_4\begin{pmatrix}0\\-1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

$\Rightarrow u_3 = -6u_0$, $u_4 = -4u_0$

So, for any $u_0 > 0$ we cannot get $u_i \geq 0 \ \forall i$

If $u_0 = 0$, $u_3 = u_4 = 0$, but all multipliers cannot be zero.

Now if we add equality constraints

COP

$$\boxed{\begin{aligned} &\text{Min } f(x) \\ &\text{s.t } g_i(x) \le 0, \quad i=1 \dots m \\ (\text{affine}) \quad &h_j(x) = 0, \quad j = 1 \dots l \end{aligned}}$$

How will the Fritz-John conditions change?

At $\bar{x}$,

$$\begin{cases} u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{j=1}^{l} v_j \nabla h_j(\bar{x}) = 0 \\ u_i g_i(\bar{x}) = 0, \quad i = 1 \dots m \\ u_0, u_i \neq 0 \end{cases}$$

Redundant $\times$

$$\boxed{v_j \underbrace{h_j(\bar{x})}_{=0} = 0, \quad v_j \in \mathbb{R}}$$

need not be positive always.

Always satisfied.

Note: $h_j(\bar{x}) = 0$ can be written as

$$h_j(\bar{x}) \le 0$$

$$-h_j(\bar{x}) \le 0$$

KARUSH-KUHN-TUCKER CONDITIONS
(KKT Conditions)

$$\boxed{A \Rightarrow B} \Leftrightarrow \boxed{\neg B \Rightarrow \neg A}$$

$B \not\Rightarrow A$ $\qquad$ $B \not\Rightarrow A$

Sufficient $\qquad$ Necessary

KKT Sufficient Conditions

Thm 4.3.8

$$\begin{aligned} &\text{Min } f(x) \\ \text{s.t. } &g_i(x) \le 0, \quad i = 1 \dots m \\ &h_j(x) = 0, \quad j = 1 \dots l \end{aligned}$$

If f is pseudoconvex, g_i and h_j's are all quasiconvex, and $\bar{x}$ satisfies the KKT conditions, i.e., there exists scalars $u_i \geq 0$, such that

$$\boxed{\bar{X} \text{ is KKT}} \xrightarrow[g_i,\, h_j \text{ quasiconvex}]{f \text{ pseudoconvex}} \boxed{\bar{X} \text{ is optimal}}$$

(A) $\left[\begin{array}{ll} g_i(\bar{x}) \leq 0, & i = 1 \cdots m \\ h_j(\bar{x}) = 0, & j = 1 \cdots \ell \end{array}\right]$ Primal feasibility

(B) $\left[\nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{j=1}^{\ell} v_j \nabla h_j(\bar{x}) = 0\right]$ Stationarity

(C) $[\; u_i g_i(\bar{x}) = 0 \;]$ Complementary slackness

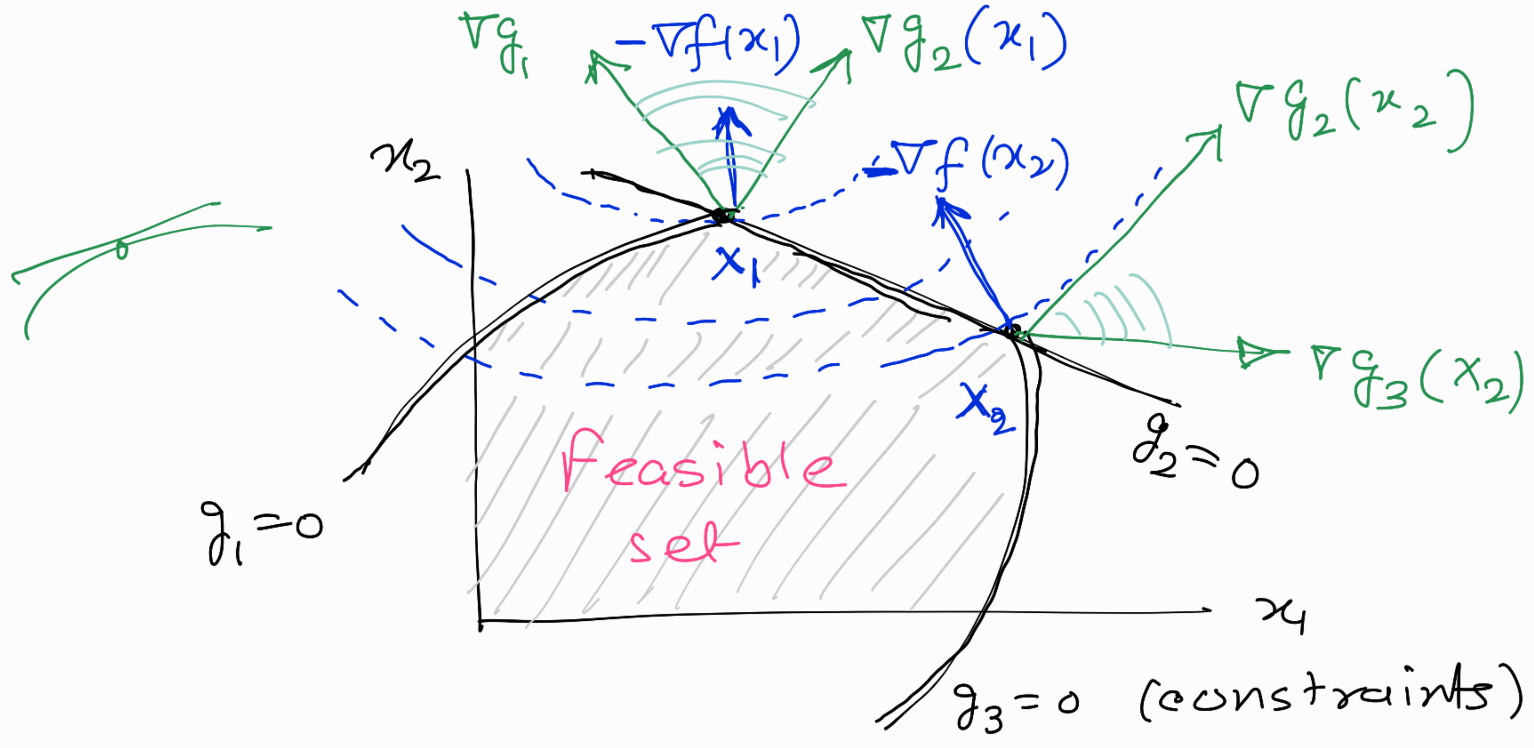
(D) $[\; u_i \geq 0 \;]$ Dual feasibility

Multipliers

Then $\bar{x}$ is a global optimal solution

Any point x that satisfies (A)-(D) is called a <u>KKT point.</u>

Geometric interpretation of KKT conditions



At x_1, $-\nabla f(x_1)$ lies inside the cone created by $\nabla g_1(x_1)$ and $\nabla g_2(x_1)$.

So, we can write
$$-\nabla f(x_1) = u_1 \nabla g_1(x_1) + u_2 \nabla g_2(x_1)$$
$$u_1, u_2 \geq 0$$

So, x_1 is a KKT point.

At x_2, $-\nabla f(x_2)$ is outside the cone generated by the rays $\nabla g_2(x_2)$ and $\nabla g_3(x_2)$, so, we cannot write
$$-\nabla f(x_2) = u_2 \nabla g_2(x_2) + u_3 \nabla g_3(x_2)$$
where $u_2 \geq 0$ and $u_3 \geq 0$

Conclusion: For any convex optimization problem (C.O.P),

if we can find a KKT point for such a problem, we have found the solution to the problem.

Are these conditions also necessary?

Consider this convex programming problem

$$\text{Min } x_1$$

$$\text{s.t. } (x_1-1)^2 + (x_2-1)^2 \leq 1 \quad (= g_1(x))$$

$$(x_1-1)^2 + (x_2+1)^2 \leq 1 \quad (= g_2(x))$$

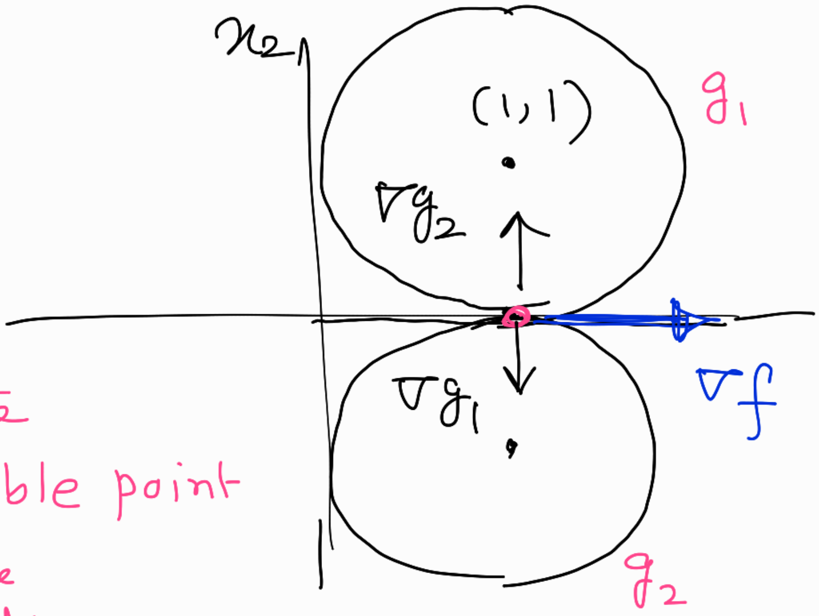


Fig 4.11

$\bar{x} = (1,0)$ is the only feasible point and hence the optimal solution

Note that the point $\bar{x} = (1,0)$ is not KKT because $\nabla g_2(\bar{x}) = -\nabla g_1(\bar{x})$, the cone cannot be formed by the rays $\nabla g_1(\bar{x})$ and $\nabla g_2(\bar{x})$.

This shows KKT-sufficient conditions are not necessary!

Exercise 4.7 (Bazaraa)

$$\text{Min } \left(x_1 - 9/4\right)^2 + (x_2 - 2)^2$$
$$\text{s.t. } g_1(x) \equiv x_2 - x_1^2 \geq 0$$
$$g_2(x) \equiv x_1 + x_2 \leq 6, \quad x_1, x_2 \geq 0$$

a) Write KKT conditions and verify it holds true at $\bar{x} = (3/2, 9/4)$

b) Interpret KKT conditions at $\bar{x}$ graphically

c) Show $\bar{x}$ is unique global optima.

Solution

At $\bar{x}$:

$$\nabla f(\bar{x}) + u_1 \nabla g_1(\bar{x}) + u_2 \nabla g_2(\bar{x}) - u_3 - u_4 = 0$$
(stationarity)

$$\left.\begin{aligned} u_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ u_2(6 - \bar{x}_1 - \bar{x}_2) &= 0 \\ u_3 \bar{x}_1 &= 0 \\ u_4 \bar{x}_2 &= 0 \end{aligned}\right\} \text{ complementary slackness (CS)}$$

$u_i \geq 0, \ i = 1, 2, 3, 4 \Rightarrow$ Dual feasibility

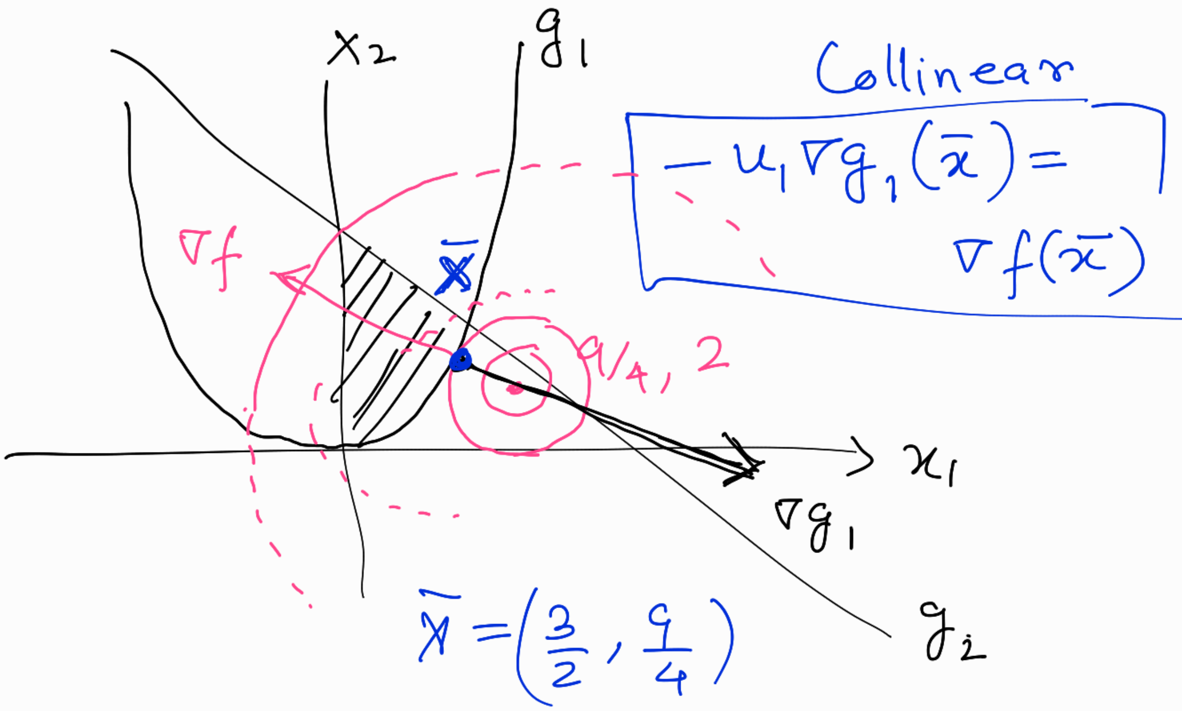
$$\left.\begin{aligned} \bar{x}_1^2 - \bar{x}_2 &\leq 0 \\ \bar{x}_1 + \bar{x}_2 &\leq 6 \end{aligned}\right\} \text{ Primal feasibility}$$

Using C.S, we get $u_3 = u_4 = u_2 = 0$, By stationarity we get

$$\begin{pmatrix} 2\bar{x}_1 - 9/2 \\ 2\bar{x}_2 - 4 \end{pmatrix} + u_1 \begin{pmatrix} 2\bar{x}_1 \\ -1 \end{pmatrix} + u_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + u_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + u_4 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0$$

put $u_2 = u_3 = u_4 = 0$, and we get

$\boxed{u_1 = 1/2}$, so, KKT conditions are satisfied at $\bar{x} = \left(\frac{3}{2}, 9/4\right)$



c) Is $\bar{x}$ unique global minima?

Ans: Graphically we see that all other feasible points lie on a higher level curve.

Also, the problem is convex optimization problem, (f is convex, st feasible region is closed convex set)

Exercise: 4.5

$$\text{Min } f(x) = x_1^4 + x_2^4 + 12x_1^2 + 6x_2^2 - x_1 x_2 - x_1 - x_2$$

s.t. $x_1 + x_2 \geq 6 \Rightarrow -x_1 - x_2 \leq -6$

$2x_1 - x_2 \geq 3 \Rightarrow -2x_1 + x_2 \leq -3$

$x_1, x_2 \geq 0$

Write KKT conditions and show $\bar{X} = (3,3)$ is unique optimal solution.

From C.S condition $u_3 x_1 = u_4 x_2 = 0$ and $x_1 = x_2 = 3 \Rightarrow u_3 + u_4 = 0$

Stationary condition is

$$\begin{bmatrix} 4x_1^3 + 24x_1 - x_2 - 1 \\ 4x_2^3 + 12x_2 - x_1 - 1 \end{bmatrix} + u_1 \begin{pmatrix} -1 \\ -1 \end{pmatrix} + u_2 \begin{pmatrix} -2 \\ 1 \end{pmatrix} - \overbrace{(u_3 + u_4)}^{=0} = 0$$

Thus there are 2 unknowns u_1 and u_2 now.

put $x_1 = x_2 = 3$, we get $u_1 = 152$, $u_2 = 12$

Thus all KKT conditions are satisfied.

KKT Conditions

Consider the constrained problem.

$$\text{Min } f(x)$$

$$\text{s.t } g_i(x) \leq 0, \quad i = 1 \cdots m$$

$$h_j(x) = 0, \quad j = 1 \cdots l$$

The KKT system is

(1) $\nabla f(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) + \sum_{j=1}^{l} v_j \nabla h_j(x) = 0$ (stationarity)

(2) $u_i g_i(x) = 0$ (Complementary slackness)

(3) $u_i \geq 0$ (v_j unrestricted) dual feasibility

(4) $g_i(x) \leq 0, \ h_j(x) = 0$ primal feasibility

Let $\boxed{\mathcal{L}(x,u,v) = f(x) + \sum u_i g_i(x) + \sum v_j h_j(x)}$ LAGRANGIAN

Then $\nabla_x \mathcal{L}(x,u,v) = 0$ is the stationarity

$\mathcal{L}(x,v,u)$ is called the Lagrangian function and $\nabla_x \mathcal{L}(x,u,v) = 0$ gives the stationary point $x^*(u,v)$ for given (u,v)

Note that if $\boxed{\mathcal{L}(x, v, u)}$ is convex in x

$\nabla_x \mathcal{L}(\ldots\ldots) = 0$ gives the minima of the Lagrangian.

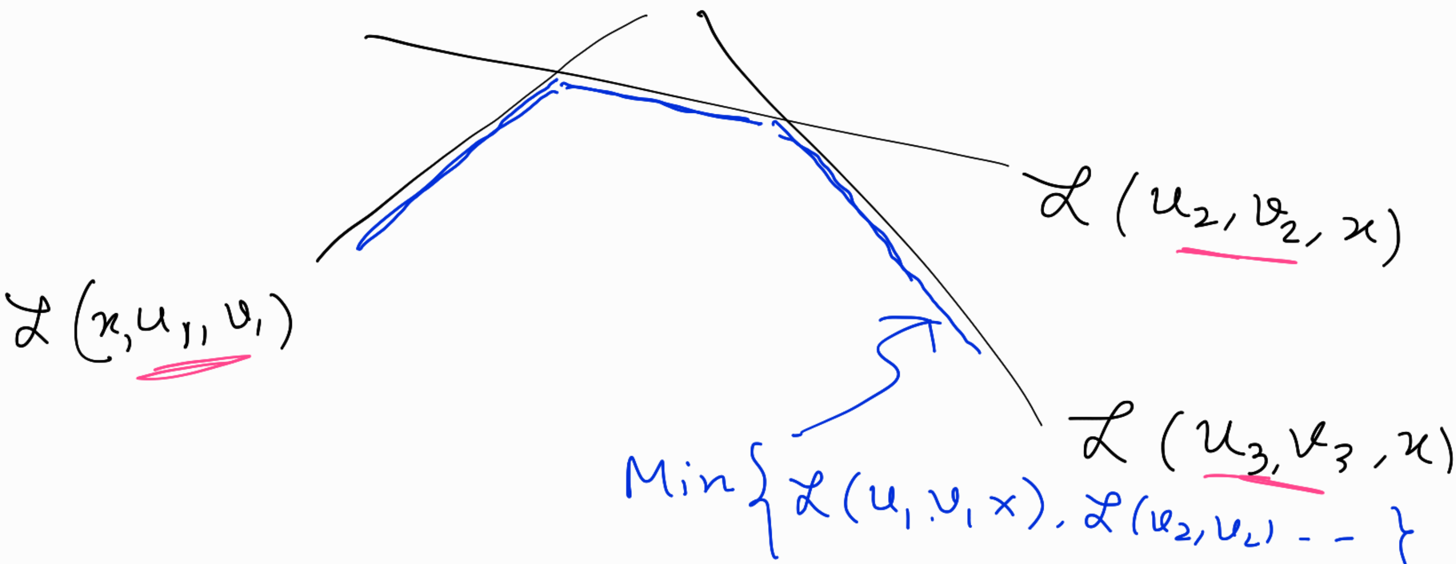
and $\mathcal{L}(x, u, v)$ is convex in x $\underline{\text{iff}}$ $\boxed{f, g, h \text{ are convex}}$

So, if we minimize $\mathcal{L}(x, u, v)$ on x we will get always affine in (u, v)

$$\theta(u, v) = \min_x \mathcal{L}(x, v, u)$$

(LAGRANGIAN DUAL) (LAGRANGIAN)

Also, $\mathcal{L}(x, u, v)$ is affine in $\underline{(u, v)}$



$\theta(u, v)$ is poinwiste minimum of affine functions $\mathcal{L}(u, v\, x)$, so, it is concave (and piece wise linear)

$\theta(u, v)$ is called the LAGRANGIAN DUAL

Since $\mathcal{L}(x,u,v)$ is always affine in (u,v) irrespective of whether f, g, h are convex in x, the pointwise minimum

$$\min_{x} \mathcal{L}(x, v, u) = \theta(u,v)$$

$\theta(u,v)$ is always concave in (u,v)

LAGRANGIAN DUAL $\Rightarrow$ Concave in (u,v) ALWAYS

LAGRANGIAN $\Rightarrow$ Convex in x, iff f, g, h are convex.

So the maximization will yield (u^*, v^*)

$$\max_{(u,v)} \theta(u,u) \Rightarrow u^*, v^*$$

which are called 'OPTIMAL DUAL VARS

$$u^*, v^* \Leftarrow \max_{(u,v)} \theta(u,v) = \min_{x} \mathcal{L}(x,u,v) \Rightarrow x^* = x(u,v)$$

plug (u^*, v^*)

Is this going to give optimal sol^n?

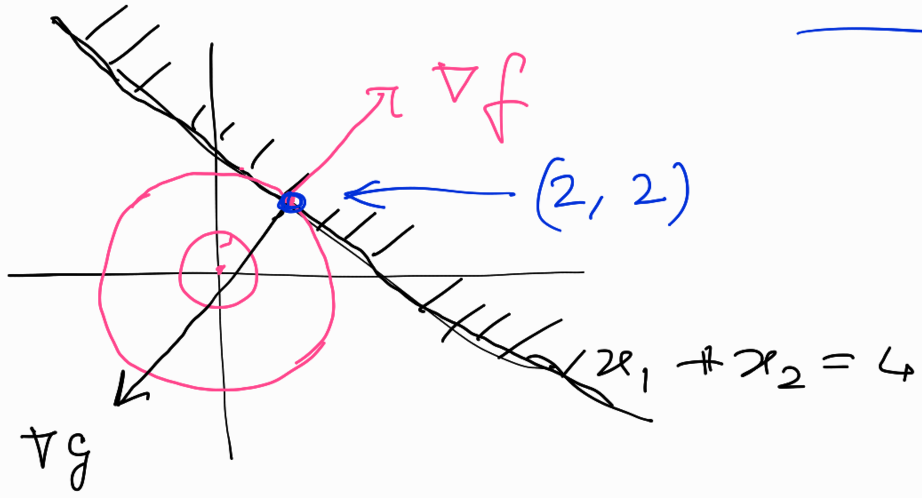
Example : Primal minimization

$$\text{Min} \quad x_1^2 + x_2^2$$
$$\text{s.t.} \quad -x_1 - x_2 + 4 \leq 0 , \quad x_1, x_2 \geq 0$$

graphically we get the solution as

$x^* = (2, 2)$

so. $f(x^*) = 8$



Lagrangian f^n:

$$\mathcal{L}(x_1, x_2, u) = x_1^2 + x_2^2 + u(-x_1 - x_2 + 4)$$

$$\underset{(x_1, x_2)}{\text{Min}} \mathcal{L}(x_1, x_2, u) = \left[\underset{x_1 \geq 0}{\text{Min}} \; x_1^2 - u x_1\right] + \left[\underset{x_2 \geq 0}{\text{Min}} \; x_2^2 - u x_2\right] + 4u$$

($\mathcal{L}$ is separable)

If $u \geq 0$, $\quad x_1^*(u) = x_2^*(u) = u/2$

If $u < 0$, $\quad x_1^*(u) = x_2^*(u) = 0$

We know,

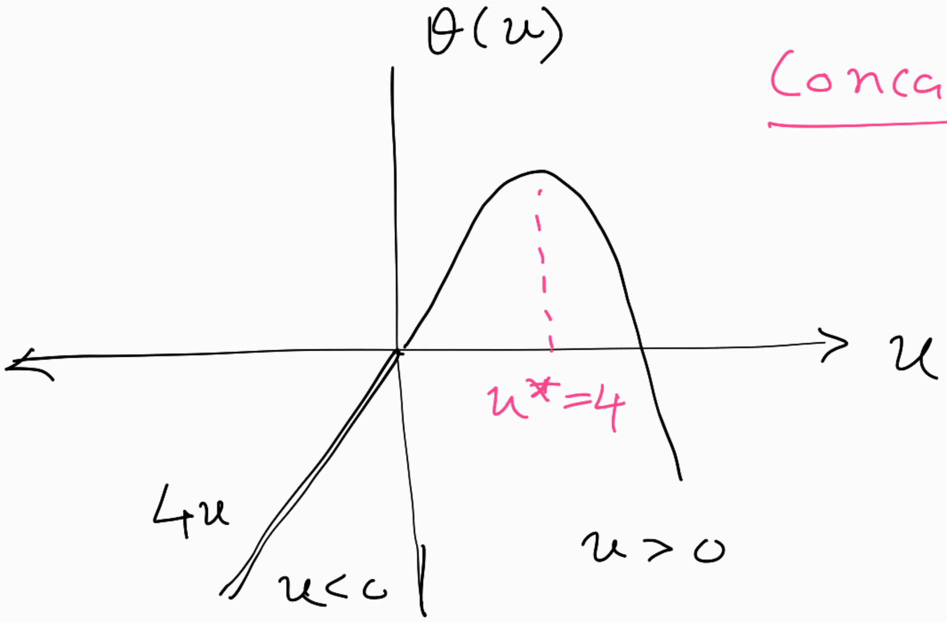
$$\theta(u) = \mathcal{L}(x_1^*(u), x_2^*(u)) = \underset{x}{\text{Min}} \; \mathcal{L}(x_1, x_2, u)$$

$$\theta(u) = \begin{cases} 4u - \dfrac{u^2}{2} & u \geq 0 \\ 4u & u < 0 \end{cases}$$

(Lagrangian dual)

$$\underset{u}{\text{Max}} \; \theta(u) \Leftarrow \text{Dual Maximization.}$$

Concave



$$\theta(u^*) = 8 = f(x^*)$$

optimal dual value

$$\underbrace{\theta(u^*) - f(x^*)}_{\text{duality gap}}$$

optimal primal value

dual solution

$$\boxed{u^* = 4 \leftarrow \theta(u)} = \min_{x} \mathcal{L}(x,u) \longrightarrow \boxed{x^*(u^*) = \begin{bmatrix}2\\2\end{bmatrix}}$$

Dual maximization.

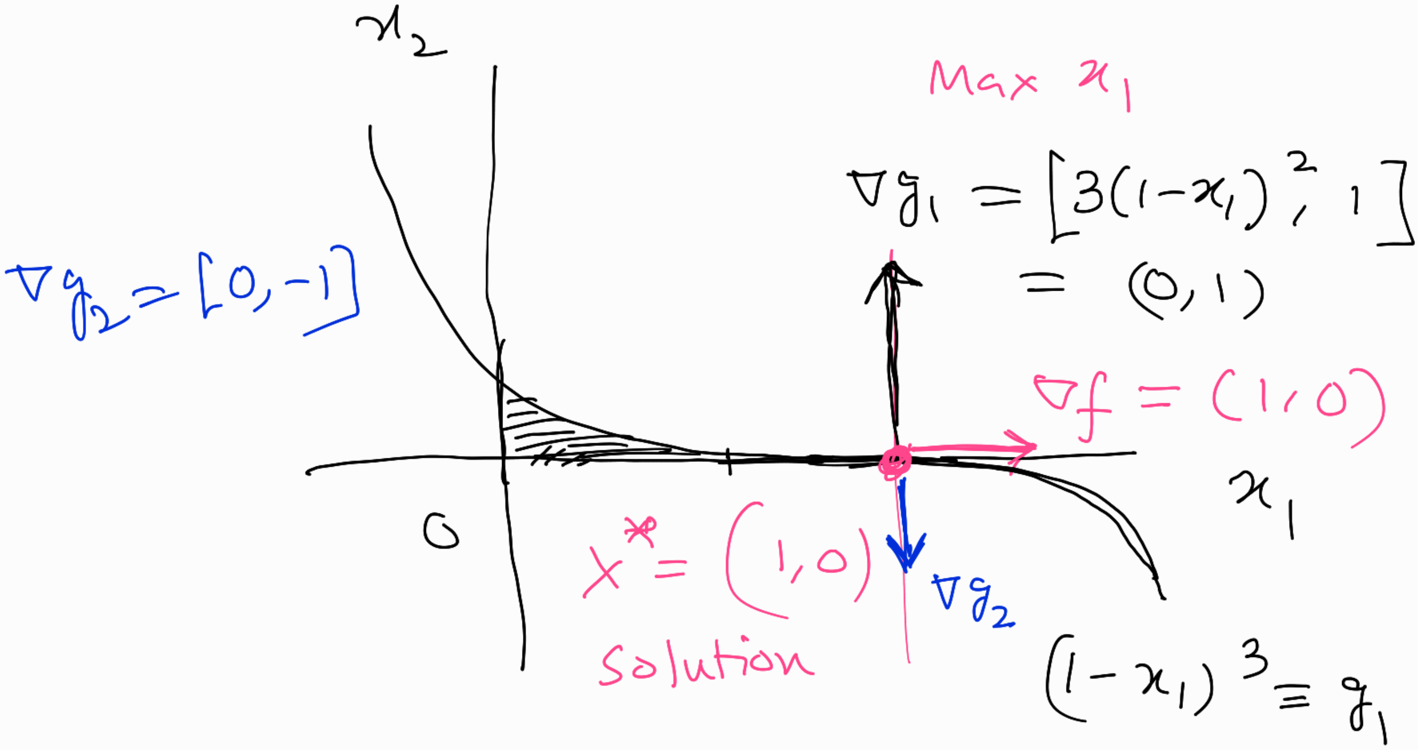
Primal solution

Example:

$$\text{Max } x_1$$

$$\text{s.t. } x_2 \le (1-x_1)^3$$

$$x_2 \ge 0$$



So, $\bar{x} = (1,0)$ is optimal solution

but $\nabla f(\bar{x}) = \sum_{i=1}^{2} u_i \nabla g_i(x)$

with $u_i \geq 0$ cannot be written.

So, $\bar{x}$ is not KKT.

Here $\nabla g_1(\bar{x})$ and $\nabla g_2(\bar{x})$ are not linearly independent, and feasible region is not convex.

Example: Max $-(x_1-2)^2 - 2(x_2-1)^2$

s.t. $\left[\begin{array}{l} x_1 + 4x_2 \leq 3 \\ \boxed{x_1 \geq x_2} \end{array}\right]$ convex set

or, $-x_1 + x_2 \leq 0$

Find solution graphically then analytically.

First note that feasible set is convex and objective f is convex

Therefore let us find a KKT point.

The Lagrangian is

$$L(x_1, x_2, u_1, u_2) = (x_1-2)^2 + 2(x_2-1)^2 + \mu_1(-3 + x_1 + 4x_2) + \mu_2(0 - x_1 + x_2)$$

Set

$$\nabla_x \mathcal{L}(x, u_1, u_2) = 0, \quad x = (x_1, x_2)$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = 2(x_1-2) + \mu_1 - \mu_2 = 0$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 4(x_2-1) + 4\mu_1 + \mu_2 = 0$$

Stationarity

Complementary slackness

$$\mu_1(4x_2 + x_1 - 3) = 0$$

$$\mu_2(x_2 - x_1) = 0$$

From this we get 4 cases,

✗ a) $\underline{\mu_1 = \mu_2 = 0} \Rightarrow x_1 = 2 \quad x_2 = 1$

✗ b) $\mu_1 = 0,\ x_2 - x_1 = 0 \Rightarrow x_1 = \frac{4}{3},\ \boxed{\mu_2 = -\frac{4}{3}}$

✓ c) $3 - x_1 - 4x_2 = 0,\ \underline{\underline{\mu_2}} = 0 \Rightarrow x_1 = \frac{5}{3},\ \mu_1 \underline{\underline{=}} \frac{2}{3}$

$x_2 = 1/3$

✗ d) $3 - x_1 - 4x_2 = 0,\ x_1 - x_2 = 0 \Rightarrow$

$x_1 = 3/5$

$x_2 = 3/5$

$\mu_1 = 22/25$

$\boxed{\mu_2 = -48/25}$

So, the solution is $x_1^* = \frac{5}{3},\ x_2^* = \frac{1}{3}$

Is the solution unique? YES

At $x^* = \left(\frac{5}{3}, \frac{1}{3}\right)$, $\nabla g_1(x) = [1, 4]$

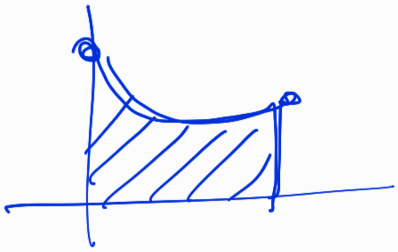
$\nabla g_2(x) = [-1, 1]$

∇g_1 and ∇g_2 are <u>Linearly</u> <u>independent</u>

(This property is required on the constraints. This is also called "Constraint Qualification")

Hometask :

① Solve graphically and verify if KKT holds at the optimal point



$$\text{Max} \quad 3x_1 + 5x_2$$

$$\text{s.t} \quad x_1 \leq 4, \quad 2x_2 \leq 14.$$

$$8x_1 - x_1^2 + 14x_2 - x_2^2 \leq 49$$

② $$\text{Min} \quad x_1^2 + x_2^2$$

$$\text{s.t.} \quad x_1^2 + x_2^2 \leq 5$$

$$x_1 + 2x_2 = 4, \quad x_1, x_2 \geq 0$$

③ $$\text{Max} \quad \ln(x_1 + 1) + x_2$$

$$\text{s.t} \quad 2x_1 + x_2 \leq 3, \quad x_1, x_2 \geq 0$$

The BIG picture

$\boxed{\bar{x} \text{ KKT point}} \Longleftrightarrow \boxed{\bar{x} \text{ optimal point}}$

Convex problem

+

Constraint qualification

Theorem 5.3.1 (KKT necessary condition)

- - - - - - -

If $\bar{x}$ is a solution to the problem and "a constraint qualification" holds, then $\bar{x}$ is also KKT.

$$\boxed{\bar{x} \text{ local optimal}} \xrightarrow{\text{via constraint qualification}} \boxed{\bar{x} \text{ is KKT}}$$

§. Examples of KKT conditions for Some ML Models

(1)

$$\text{Min} \quad \frac{1}{2} x^T Q x + c^T x \quad (\text{quadratic})$$

$$\text{s.t} \quad Ax = b \quad (\text{Linear})$$

KKT condition:

Lagrangian function.

$$L(x, u) = \frac{1}{2} x^T Q x + c^T x + u^T (Ax - b)$$

$$\nabla_x \mathcal{L}(x,u) = 0 = Qx + c + A^T u$$ (Stationarity)

$Ax = b$ (primal feasibility)

$u \gtrless 0$ (unrestricted)

Complementary slackness is always satisfied

KKT system

(x,u) satisfies $\Rightarrow \begin{bmatrix} Q\mathbf{x} + c + A^T\mathbf{u} = 0 \\ A\mathbf{x} = b \end{bmatrix}$

Linear system.

$$\underbrace{\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}}_{\text{KKT Matrix}} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

(2) $\text{Min} \quad \frac{1}{2}x^T Q x + c^T x$ (e.g Soft Margin SVM)

s.t $\quad Ax = b$

$\quad Gx \preceq h$ (system of inequalities) which are convex

Lagrangian:

$$L(x,u,v) = \frac{1}{2}x^T Q x + c^T x + u^T(Gx - h) + v^T(Ax - b)$$

(i) $\nabla_x \mathcal{L}(x,u,v) = Qx + c + u^T G + v^T A = 0$ (stationarity)

(ii) $u^T(Gx - h) = 0$ (complementary slackness)

(iii) $u \geq 0$ (dual feasibility)

$\left.\begin{array}{l}\text{(iv) } Gx - h + \delta = 0 \\ \text{(v) } Ax = b\end{array}\right\}$ primal feasibility

slack variables

$$\begin{cases} Qx + c + u^T G + v^T A = 0 \\ Gx - h + \delta = 0 \\ Ax - b = 0 \end{cases}$$

primal

$$\begin{bmatrix} Q & G & A & 0 \\ G^T & 0 & 0 & 1 \\ A^T & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ u^T \\ v^T \\ \delta \end{bmatrix} = \begin{bmatrix} -c \\ h \\ b \end{bmatrix}$$

dual variables

$u \geq 0$, $u^T(Gx - h) = 0$

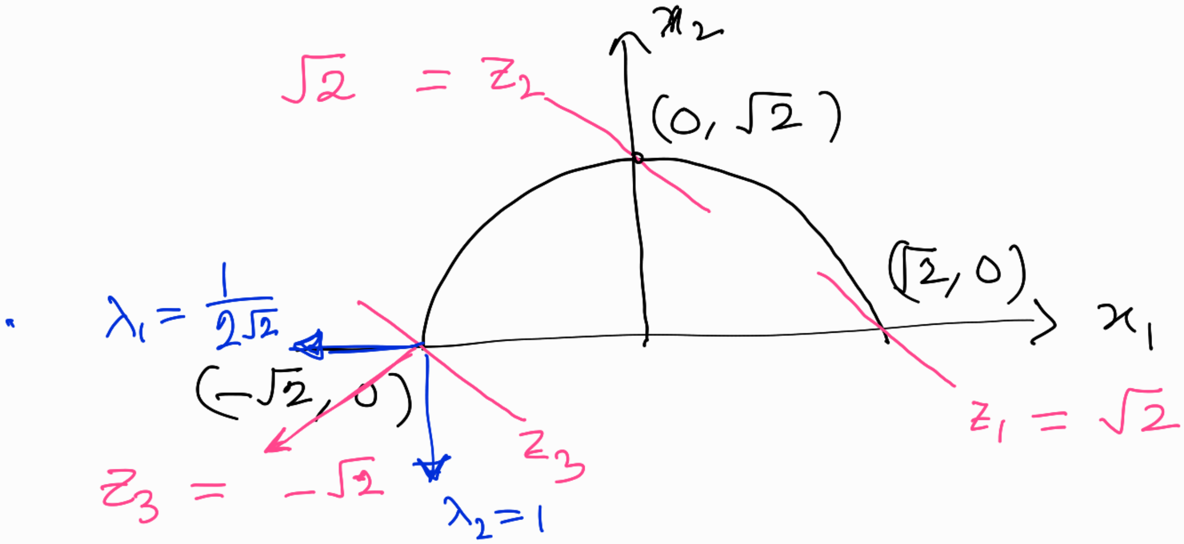
artificial / slack

SVM

Example : Min $x_1 + x_2 = z$

s.t $x_1^2 + x_2^2 \leq 2$

$x_2 \geq 0$



Lagrangian $\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2) = x_1 + x_2$

$$+\lambda_1 (x_1^2 + x_2^2 - 2)$$

$$-\lambda_2 x_2$$

$$\left.\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= 1 + 2\lambda_1 x_1 = 0 \\ \frac{\partial \mathcal{L}}{\partial x_2} &= 1 + 2\lambda_1 x_2 - \lambda_2 = 0 \end{aligned}\right\} \text{Stationarity}$$

$$\lambda_1, \lambda_2 \geq 0 \qquad \text{Dual feasibility}$$

$$\left.\begin{aligned} \lambda_1 (2 - x_1^2 - x_2^2) &= 0 \\ \lambda_2 x_2 &= 0 \end{aligned}\right\} \text{complementary slackness}$$

$$\left.\begin{aligned} -x_2 &\leq 0 \\ x_1^2 + x_2^2 - 2 &\leq 0 \end{aligned}\right\} \text{Primal feasibility}$$

After solving for $x_1, x_2, \lambda_1, \lambda_2$ we get

$$x_1 = -\sqrt{2} \;,\; x_2 = 0 \;,\; \lambda_1 = \frac{1}{2\sqrt{2}} \;,\; \lambda_2 = 1$$

Example :

Find the point on $y = \frac{1}{5}(x-1)^2$ that is closest to the point $(1,2)$

Soln.

$$\text{Min} \quad (x-1)^2 + (y-2)^2$$

$$\text{s.t.} \quad (x-1)^2 = 5y \quad \text{(equality)}$$

$$\mathcal{L}(x,y,\lambda) = (x-1)^2 + (y-2)^2 + \lambda\left[(x-1)^2 - 5y\right]$$

$\lambda \gtrless 0$ because there is equality const. only.

$$\frac{\partial \mathcal{L}}{\partial x} = 0 \Rightarrow 2(x-1) + 2\lambda(x-1) = 0$$

$$\frac{\partial \mathcal{L}}{\partial y} = 0 \Rightarrow 2(y-2) - 5\lambda = 0$$

and $\quad (x-1)^2 = 5y \Rightarrow y \geq 0$

$$(2+2\lambda)(x-1)=0$$

$\Rightarrow \quad x=1 \qquad$ or $\qquad \lambda=-1$

$\downarrow \qquad\qquad\qquad\qquad \downarrow$

From $5y=(x-1)^2 \qquad\qquad y=-\frac{1}{2}$ X

$y=0$

$\downarrow$

$\lambda=-4/5$

Solution is $(1,0)$ and $\lambda=-4/5$