Optimality Conditions

Let's consider an optimization problem given by (P) $\begin{pmatrix} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{pmatrix}$ where f is continuously differentiable (not necessarily convex) and C is a closed convex set.

Defn: $x^* \in C$ is called a stationary point of (P) if $\nabla f(x^*)^T(x - x^*) \geq 0$ for any $x \in C$.

Note that if $C = \mathbb{R}^n$ (hence (P) is an unconstrained problem), then a stationary point x^* satisfies

$$\nabla f(x^*)^T(x - x^*) \geq 0 \quad \text{for all } x \in \mathbb{R}^n. \quad (*)$$

In particular, $(*)$ has to hold for $x = x^* - \nabla f(x^*)$. This gives

$$\nabla f(x^*)^T\left(x^* - \nabla f(x^*) - x^*\right) = -\nabla f(x^*)^T \nabla f(x^*) \geq 0$$

$$\Longrightarrow \quad -\|\nabla f(x^*)\| \geq 0$$

$\Rightarrow \quad \nabla f(x^*) = 0$ has to hold.

Hence, we recover the stationary points for unconstrained optimization problems!

Theorem: (Necessary optimality condition)

Consider problem (P), where $C \subseteq \mathbb{R}^n$ is convex closed & f is continuously differentiable. If x^* is a local minimum point for (P), then it's a stationary point.

Proof: Assume x^* is a local minimum point, but not a stationary point. Then, $\exists\, x \in C$ such that

$$\nabla f(x^*)^T (x - x^*) < 0.$$

Recall that the directional derivative is $f'(x^*; d) = \nabla f(x^*)^T d$. Then, $d = x - x^*$ is a descent direction of f at x^* since

$$f'(x^*, x - x^*) = \nabla f(x^*)^T (x - x^*) < 0.$$

Then, there exists $\varepsilon > 0$ such that $f(x^* + td) < f(x^*)$ for all $0 < t < \varepsilon$.

Now, note that $x^* \in C$, $x \in C$ and C is convex. Then,

$$\underbrace{(1-t)x^* + tx}_{\in C} = x^* + t(x - x^*) = x^* + t.d \in C.$$

But this contradicts the local optimality of x^*. Hence, x^* has to be a stationary point. ∎

Theorem: (Sufficient condition for convex f)

Consider problem (P), $C \subseteq \mathbb{R}^n$ is closed & convex. Moreover, assume that f is a continuously differentiable convex function. Then, x^* is an optimal solution iff it's a stationary point.

Proof: If x^* is an optimal solution, then it's a stationary point by the previous theorem.

Assume x^* is a stationary point. Let $x \in C$ be arbitrary. By the first order characterization of convex functions we have

$$f(x) \geq f(x^*) + \underbrace{\nabla f(x^*)^T (x - x^*)}_{\geq 0 \quad \text{as } x^* \text{ is a stationary point.}}$$

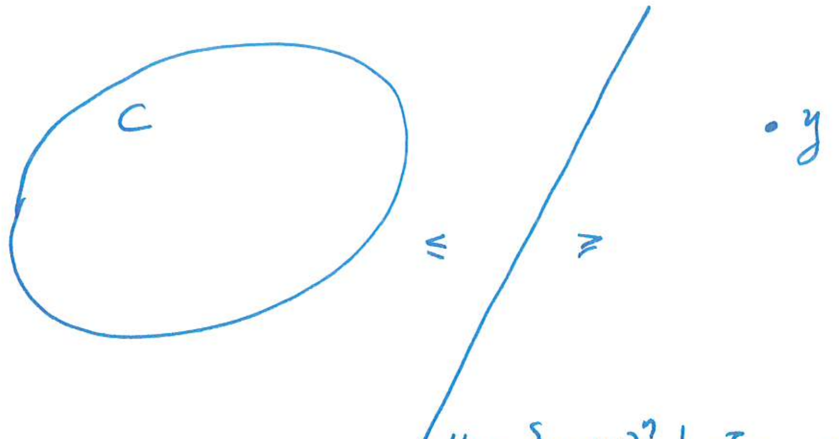
Hence, $f(x) \geq f(x^*)$ for all $x \in C$. This shows that x^* is an optimal solution for (P). ∎

Optimality Conditions for Linearly Constrained Problems

There's a strong yet intuitive result on convex sets which plays an important role for proving the optimality conditions later.

Result: (Strict separation) Let $C \subseteq \mathbb{R}^n$ be a closed convex set and $y \notin C$. Then, $\exists\, p \in \mathbb{R}^n \setminus \{0\}$ and $\alpha \in \mathbb{R}$ s.t.

$$p^T y > \alpha \quad \text{and} \quad p^T x \leq \alpha \text{ for all } x \in C.$$



$H := \{x \in \mathbb{R}^n \mid p^T x = \alpha\}$

H is a "separating hyperplane" that strictly separates y from set C.

Farkas' lemma: Let $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$. Then, "exactly" one of the following systems has a solution.

I. $Ax \leq 0$, $c^T x > 0$

II. $A^T y = c$, $y \geq 0$.

This means that if $\exists\, x \in \mathbb{R}^n$ satisfying I, then there's no $y \in \mathbb{R}^m$ satisfying II. Conversely, if there's no $x \in \mathbb{R}^n$ satisfying I, then $\exists\, y \in \mathbb{R}^m$ satisfying II.

Proof of Farkas' lemma: Assume $\exists\, y \in \mathbb{R}^m$ satisfying II: $A^T y = c$, $y \geq 0$. We want to show that there's no $x \in \mathbb{R}^n$ satisfying I. (If there's no x with $Ax \leq 0$, then we are done.) If there's $x \in \mathbb{R}^n$ with $Ax \leq 0$, (indeed $x = 0$ would satisfy this!) then we observe that

$$Ax \leq 0 \Rightarrow y^T A x \leq 0 \Rightarrow \underbrace{(A^T y)}_{= c \text{ (from II)}}{}^T x \leq 0 \Rightarrow c^T x \leq 0 \text{ has to hold.}$$

$\Longrightarrow x$ does not satisfy I.

This shows that there's no x satisfying I if there's $y \in \mathbb{R}^m$ satisfying II.

For the reverse implication, assume that there's no $y \in \mathbb{R}^m$ satisfying II.

Consider the set

$$S = \{ x \in \mathbb{R}^n \mid x = A^T y \text{ for some } y \geq 0 \}$$

$$= \{ A^T y \mid y \in \mathbb{R}^m, y \geq 0 \}.$$

It can be shown that this is a closed convex set. Moreover, $c \notin S$.

As otherwise, it would mean that $A^T y = c, y \geq 0$ is satisfied.

By strict separation, $\exists\, p \in \mathbb{R}^n \setminus \{0\}, \alpha \in \mathbb{R}$ such that

① $p^T c > \alpha$ and ② $p^T x \leq \alpha$ for all $x \in S$.

Note that $0 \in S$, as for $y = 0$, $A^T y = 0$. Then, ② implies that

$p^T 0 = 0 \leq \alpha$ holds. ① implies $p^T c > \alpha \geq 0$.

Moreover, ② means that $p^T A^T y \leq \alpha$ for all $y \geq 0$.

$$\Rightarrow (Ap)^T y \leq \alpha \text{ for all } y \geq 0.$$

This implies that $Ap \leq 0$ has to hold. Otherwise, if there's a component that is positive, then $(Ap)_i y_i \longrightarrow +\infty$ as $y_i \longrightarrow +\infty$.

Note that p satisfies system I. $(Ap \leq 0,\ c^T p > 0)$

Theorem: (KKT Conditions for linearly constrained problems)

Let $f: \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^m$, $d \in \mathbb{R}^p$.

Consider problem

$$\begin{aligned} (P) \quad & \text{minimize} \quad f(x) \\ & \text{subject to} \quad Ax \leq b \\ & \qquad\qquad\quad\; Cx = d. \end{aligned}$$

(a) If x^* is a local minimum point of (P), then $\exists\, \lambda \in \mathbb{R}^m$, $\underline{\lambda \geq 0}$ and $\mu \in \mathbb{R}^p$ such that

(1) $\nabla f(x^*) + A^T\lambda + C^T\mu = 0$

(2) $\lambda_i (a_i^T x^* - b_i) = 0, \quad i = 1, \dots, m$ $\quad$ (a_i^T is the ith row of A)

(b) Assume that f is convex. If x^* is feasible for (P) and $\exists\, \lambda \in \mathbb{R}^m$, $\lambda \geq 0$, $\mu \in \mathbb{R}^p$ such that (1), (2) hold true, then x^* is an optimal solution for (P).

Later, we will see similar results for more general problems. For now, let us write the conditions (1), (2) in a form which will be used later in general setting.

Let's rewrite problem (P) as follows:

$$\begin{aligned} & \text{minimize} \quad f(x) \\ & \text{subject to} \quad a_i^T x - b_i \leq 0, \quad i = 1, \dots, m \\ & \qquad\qquad\quad\; c_j^T x - d_j = 0, \quad j = 1, \dots, p. \end{aligned}$$

The Lagrangian associated with this problem is a function of the form:

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i (a_i^T x - b_i) + \sum_{j=1}^{p} \mu_j (c_j^T x - d_j),$$

where λ_i, μ_j are the "Lagrange multipliers" corresponding to constraints $a_i^T x - b_i \leq 0$ and $c_j^T x - d_j = 0$, respectively. Note that condition (2) is written for each inequality constraint: $\lambda_i (a_i^T x - b_i) = 0 \ \forall i$.

On the other hand, condition (1) corresponds to writing $\nabla_x L(x,\lambda,\mu) = 0$.

Indeed, we can write the Lagrangian using the matrix notation as:

$$\begin{aligned} L(x,\lambda,\mu) &= f(x) + \lambda^T (Ax - b) + \mu^T (Cx - d) \\ &= f(x) + (A^T\lambda)^T x - \lambda^T b + (C^T\mu)^T x - \mu^T d \end{aligned}$$

Taking the partial derivative with respect to x yields:

$$\nabla_x L(x,\lambda,\mu) = \nabla f(x) + A^T\lambda + C^T\mu .$$

Hence $(2) \iff \nabla_x L(x,\lambda,\mu) = 0$.

If problem (P) has only inequality constraint, i.e., $p=0$, then we have $L(x,\lambda) = f(x) + \lambda^T (Ax - b)$. Similarly, if $m=0$, then we have $L(x,\mu) = f(x) + \mu^T (Cx - d)$. In both cases, condition (1) is $(\nabla_x L_x = 0)$ and the theorem holds true.

Example: minimize $\frac{1}{2}(x_1^2 + x_2^2 + x_3^2)$

subject to $x_1 + x_2 + x_3 = 3$.

Note that $f(x) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) = \frac{1}{2}\|x\|_2^2$ is a convex function. Then, conditions (1) & (2) are both necessary & sufficient for optimality.

If there's a feasible x satisfying (1) & (2), it's optimal.

Lagrangian: $L(x, \mu) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \mu(x_1 + x_2 + x_3 - 3)$

Since there's no inequality constraint, condition (2) is not relevant.

$$\nabla_x L(x,\mu) = 0 \iff \left.\begin{aligned} \frac{\partial L}{\partial x_1} &= x_1 + \mu = 0 \\ \frac{\partial L}{\partial x_2} &= x_2 + \mu = 0 \\ \frac{\partial L}{\partial x_3} &= x_3 + \mu = 0 \end{aligned}\right\} \quad x_1 = x_2 = x_3 = -\mu \text{ has to hold.}$$

Together with feasibility, we obtain $x_1 + x_2 + x_3 = -3\mu = 3 \rightarrow \mu = -1$.

Then, the unique solution of this system is $x_1 = x_2 = x_3 = 1$.

$x^* = (1, 1, 1)$ is the unique optimal solution with value: $f(x^*) = \frac{3}{2}$.

Example: minimize $x_1^2 + x_2^2 + 4x_1x_2$

subject to $x_1 + x_2 = 1$

$x_1, x_2 \geq 0$ $\quad (-x_1 \leq 0,\ -x_2 \leq 0)$

Is this a convex program? Note that $f(x) = x_1^2 + x_2^2 + 4x_1x_2$.

$$\nabla f(x) = \begin{bmatrix} 2x_1 + 4x_2 \\ 2x_2 + 4x_1 \end{bmatrix} \quad \text{and} \quad \nabla^2 f(x) = \begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix}.$$

$\det \nabla^2 f(x) = 4 - 16 = -12 < 0$. This implies that one of the eigenvalues is positive and the other one is negative. $\nabla^2 f(x)$ is indefinite, hence f is not convex. Thus, KKT conditions are not sufficient for optimality. However, they're necessary conditions.

Let's write the Lagrangian:

$$L(x, \lambda, \mu) = x_1^2 + x_2^2 + 4x_1x_2 + \lambda_1(-x_1) + \lambda_2(-x_2) + \mu(x_1 + x_2 - 1)$$

KKT conditions are:

(1) $\nabla_x L(x, \lambda, \mu) = 0 \iff \dfrac{\partial L}{\partial x_1} = 2x_1 + 4x_2 - \lambda_1 + \mu = 0$

$$\frac{\partial L}{\partial x_2} = 2x_2 + 4x_1 - \lambda_2 + \mu = 0$$

(2) $\lambda_1(-x_1) = 0$

$\lambda_2(-x_2) = 0$

(Primal feasibility:) $x_1 + x_2 = 1$, $x_1, x_2 \geq 0$.

(Dual feasibility:) $\lambda_1, \lambda_2 \geq 0$.

$$2x_1 + 4x_2 - \lambda_1 + \mu = 0$$
$$2x_2 + 4x_1 - \lambda_2 + \mu = 0$$
$$\lambda_1 x_1 = 0$$
$$\lambda_2 x_2 = 0$$

$$x_1 + x_2 = 1$$
$$x_1, x_2 \geqslant 0$$
$$\lambda_1, \lambda_2 \geqslant 0.$$

We know that if x^* is an optimal solution, then it's a solution of this system (necessary condition). So, we'll find all x, λ, μ satisfying the KKT conditions. <u>If</u> there exists a solution to the problem, then it would be among these points!

Indeed, note that $S = \{x \in \mathbb{R}^2 \mid x_1 + x_2 = 1, x_1 \geqslant 0, x_2 \geqslant 0\}$ is a closed and bounded (hence compact) set. Moreover f is continuous. Then, by the Weierstrass theorem, there exists a solution to this problem and it has to be a KKT point.

Let's find the KKT points by considering different cases which satisfy $\lambda_1 x_1 = 0$, $\lambda_2 x_2 = 0$.

<u>Case 1</u>: $\lambda_1 = \lambda_2 = 0 \Rightarrow$
$$\left.\begin{aligned} 2x_1 + 4x_2 + \mu = 0 \\ 2x_2 + 4x_1 + \mu = 0 \end{aligned}\right\}$$
$$x_1 + x_2 = 1$$

Adding these:
$$6x_1 + 6x_2 + 2\mu = 0$$
$$\Downarrow$$
$$6(x_1 + x_2) = -2\mu$$
$$\Downarrow$$
$$\mu = -3$$

Then, we have
$$2x_1 + 4x_2 = 3$$
$$2x_2 + 4x_1 = 3$$

$\Rightarrow x_1 = x_2 = \frac{1}{2}$. Then, $x = \left(\frac{1}{2}, \frac{1}{2}\right)$ is a KKT point.

Case 2: $\lambda_1 > 0$, $\lambda_2 > 0$. This would imply $x_1 = 0$, $x_2 = 0$ by complementary slackness.

But $x = (0,0)$ is not feasible! No KKT points found from here.

Case 3: $\lambda_1 > 0$, $\lambda_2 = 0 \Rightarrow \lambda_1 > 0$ implies $x_1 = 0$.

We obtain:

$$4x_2 - \lambda_1 + \mu = 0$$
$$2x_2 + \mu = 0 \Rightarrow \mu = -2x_2 = -2$$
$$x_1 + x_2 = 1 \Rightarrow x_2 = 1$$

$$4 - \lambda_1 - 2 = 0 \Rightarrow \lambda_1 = 2 > 0.$$

So, $x = (0,1)$ is a KKT point (where $\lambda = (2,0) \geq 0$, $\mu = -2$).

Case 4: $\lambda_1 = 0$, $\lambda_2 > 0 \Rightarrow x_2 = 0 \Rightarrow x_1 = 1$ from feasibility.

We obtain:

$$2 + \mu = 0 \Rightarrow \mu = -2$$
$$4 - \lambda_2 + \mu = 0 \Rightarrow \lambda_2 = 4 - 2 = 2 > 0.$$

Hence, $x = (1,0)$ is a KKT point (where $\lambda = (0,2) \geq 0$, $\mu = -2$.).

We have 3 KKT points. $f(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4} + \frac{1}{4} + 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{2}$

$$f(0,1) = 1$$
$$f(1,0) = 1$$

Since $(0,1)$, $(1,0)$ yield the minimum objective function value, they are the optimal solutions for this problem.

Next we will prove the theorem (KKT for linearly constrained problems).

For simplicity, we consider the problem only with inequality constraints:

$$(P) \quad \text{minimize} \quad f(x)$$
$$\text{subject to} \quad a_i^T x - b_i \leq 0, \quad i = 1, \dots, m.$$

If x^* is a local minimum point, then $\exists \lambda_1, \dots, \lambda_m \geq 0$ s.t.

$$(1) \quad \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i a_i = 0$$

$$(2) \quad \lambda_i (a_i^T x - b_i) = 0, \quad i = 1, \dots, m$$

Proof: Since x^* is a local minimum, it's a stationary point:

$$(*) \quad \nabla f(x^*)^T (x - x^*) \geq 0 \quad \text{for all } x \in \mathbb{R}^n \text{ s.t. } a_i^T x - b_i \leq 0, \ i = 1, \dots, m.$$

Note that as x^* is feasible, we have $a_i^T x^* - b_i \leq 0$ for all i.

Moreover, it may be true that for some i the inequality is strict, whereas for the remaining indices it may hold as equality. We consider these constraints separately:

let $I(x^*) = \{ i \mid a_i^T x^* = b_i \} \subseteq \{1, \dots, m\}$ be the set of active constraints. So for all

$$i \in I(x^*) : \quad a_i^T x^* = b_i$$
$$i \notin I(x^*) : \quad a_i^T x < b_i.$$

To satisfy (2), we let $\lambda_i = 0$ for all $i \notin I(x^*)$. Then, (1) simplifies to

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i a_i = 0. \quad (**)$$

The aim is to show that $\exists\ \lambda_i \geq 0$ for $i \in I(x^*)$ (so that $a_i^T x^* = b_i$)

such that (**) is satisfied.

From now on assume without loss of generality that $I(x^*) = \{1, \dots, k\}$

for some $k \leq m$, that is, first k constraints hold as equality.

Now, let's write (*) by making a change of variable: $y := x - x^*$.

(*) $\nabla^T f(x^*)(x - x^*) \geq 0$ for all x : $a_i^T x - b_i \leq 0$ for $i = 1, \dots, m$.

$$\Rightarrow \nabla f(x^*)^T y \geq 0 \quad \text{for all } y : \quad a_i^T y \ \Big(= \underbrace{a_i^T x}_{\leq b_i} - a_i^T x^* \Big) \leq \underbrace{b_i - a_i^T x^*}_{= \begin{cases} 0 & \text{if } i = 1, \dots, k \\ > 0 & \text{for } i > k. \end{cases}}$$

In particular, we observe that if $a_i^T y \leq 0$ for all $i = 1, \dots, k$

and $a_i^T y \leq b_i - a_i^T x^*$ for all $i = k+1, \dots, m$

then we have $\nabla f(x^*)^T y \geq 0$.

(Indeed it's sufficient $a_i^T y \leq 0$ for all $i = 1, \dots, k$. Even if $a_i^T y \leq b_i - a_i^T x^*$ is not satisfied for $i = k+1, \dots, m$, we can find some small positive number $\alpha > 0$ such that $\alpha a_i^T y \leq b_i - a_i^T x^*$ since $b_i - a_i^T x^*$ is positive)

Then (αy) would satisfy both conditions: $a_i^T(\alpha y) \leq 0 \quad i = 1, \dots, k$

$a_i^T(\alpha y) \leq b_i - a_i^T x^*, \quad i = k+1, \dots, m.$

This implies $\nabla f(x^*)^T(\alpha y) \geq 0. \Rightarrow \nabla f(x^*)^T y \geq 0.$)

Overall, we see that

if $a_i^T y \leq 0$ for $i=1,\dots,k$, then $\nabla f(x^*)^T y \geq 0$.

In other words, there's no solution to the system:

$$\underbrace{a_i^T y \leq 0 \text{ for } i=1,\dots,k} \quad \text{and} \quad \underbrace{-\nabla f(x^*)}_{c}{}^T y > 0 .$$

Let $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_k^T \end{bmatrix} \in \mathbb{R}^{k\times n} \Rightarrow \left(Ay \leq 0, \ c^T y > 0 \right)$ has no solution.

By Farkas Lemma, there exists $\lambda \geq 0$, $A^T \lambda = c$, where $\lambda \in \mathbb{R}^k$.

That is, $\lambda_1, \dots, \lambda_k \geq 0$ such that $\begin{bmatrix} a_1, \dots, a_k \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix} = -\nabla f(x^*)$

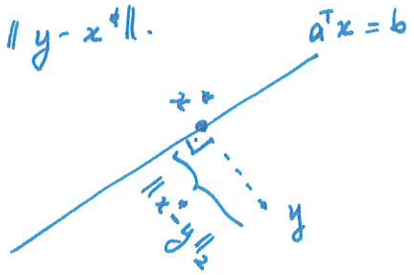
$$\Rightarrow \exists \ \lambda_1, \dots, \lambda_k \geq 0 \ \text{ s.t. } \ \nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i a_i = 0 .$$

■

Example: (Orthogonal projection onto a hyperplane)

Consider the hyperplane $H = \{ x \in \mathbb{R}^n \mid a^T x = b \}$, where $a \in \mathbb{R}^n$, $a \neq 0$, $b \in \mathbb{R}$.

Let $y \in \mathbb{R}^n$ be a given point. We want to compute the "distance" from y to H, that is, the closest point $x^* \in H$ and $\|y - x^*\|$.



This can be done by solving:

$$\left(\begin{array}{ll} \text{minimize} & \|y - x\|_2^2 \\ \text{subject to} & a^T x = b \end{array} \right)$$

Let's solve this problem by KKT conditions, as we only have linear constraints they're necessary conditions for local optimality.

Moreover, since $f(x) = \|y-x\|_2^2$ is a convex function KKT conditions are also sufficient in this case.

Lagrangian: $L(x,\mu) = \underbrace{\|x-y\|_2^2}_{(x-y)^T(x-y)} + \mu\,(a^Tx - b)$, where $\mu \in \mathbb{R}$ is a free variable. (we have equality constr. only)

KKT conditions are:

① $\nabla_x L(x,\mu) = 0$

② (Primal feasibility) $a^Tx = b$.

$$\nabla_x L(x,\mu) = 2(x-y) + \mu\cdot a = 0 \quad\Rightarrow\quad x-y = \frac{\mu\cdot a}{2}$$

$$\Rightarrow\quad x = y + \frac{\mu\cdot a}{2}$$

Then, $a^Tx = a^T\left(y + \frac{\mu a}{2}\right) = a^Ty + \mu\cdot\frac{a^Ta}{2} = b$ implies

$$\mu\cdot\frac{\|a\|_2^2}{2} = b - a^Ty \quad\Rightarrow\quad \hat{\mu} = \frac{2(b-a^Ty)}{\|a\|_2^2} \qquad (\|a\|_2 > 0 \text{ as } a \neq 0)$$

Then, $\hat{x} = y + \frac{(b-a^Ty)}{\|a\|_2^2}\cdot a$ is the only KKT point, hence the global minimum.

The optimal objective function value is $\|\hat{x}-y\|_2^2 = \left\|\frac{b-a^Ty}{\|a\|_2^2}\cdot a\right\|_2^2 = \frac{(b-a^Ty)^2}{\|a\|_2^4}\cdot\|a\|_2^2$

$= \frac{(b-a^Ty)^2}{\|a\|_2^2}$.

Distance from y to H is then $\frac{|b-a^Ty|}{\|a\|_2}$.

The KKT Conditions

Until now, we have seen optimality conditions for linearly constrained problems. Now, we will consider general constraints & discuss Karush-Kuhn-Tucker (KKT) conditions for the general case.

Let us first state Fritz-John conditions for inequality constrained problems:

Theorem: Consider problem: $$\begin{pmatrix} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \ i=1,\dots,m \end{pmatrix} \quad (P)$$

where $f_0, f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable. If x^* is a local minimum of (P), then there exists multipliers $\lambda_0, \dots, \lambda_m \geq 0$ (not all zero) such that

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) = 0$$

$$\lambda_i f_i(x^*) = 0, \quad i=1,\dots,m. \quad \text{(Complementary slackness)}$$

This result is quite general, however it has a major drawback. According to this, λ_0 can be zero, which leads the first condition to simplify as

$$\sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) = 0. \quad (*)$$

Indeed, $\lambda_i = 0$ for the constraints satisfying $f_i(x^*) < 0$. Hence, it further simplifies $\sum_{i \in I(x^*)} \lambda_i \nabla f_i(x^*) = 0$, where $I(x^*)$ is the set of indices for which we have $f_i(x^*) = 0$, i.e., the indices of the active constraints.

Condition (*) does not depend on the objective function & there may be many solutions satisfying it. Finding all such points & comparing their objective function values may not be tractable for many problems.

On the other hand, simplified version $\sum_{i \in I(x^*)} \lambda_i \nabla f_i(x^*) = 0$ means that the gradients of the active constraints are linearly dependent.

The next result is based on this observation:

Theorem: (KKT conditions for inequality constrained problems)

Consider problem:
$$\begin{pmatrix} \text{minimize } f_0(x) \\ \text{subject to } f_i(x) \leq 0, \quad i = 1, \dots, m \end{pmatrix} \quad (P)$$

where $f_0, f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable. Let x^* be a local minimum for (P) and let $I(x^*) = \{ i \mid f_i(x^*) = 0 \} \subseteq \{1, \dots, m\}$ be the set of active constraints. Assume that the vectors $\{ \nabla f_i(x^*) \mid i \in I(x^*) \}$ are linearly independent. Then, there exists $\lambda_1, \dots, \lambda_m \geq 0$ such that

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) = 0 \qquad (\nabla_x L(x, \lambda) = 0)$$

$$\lambda_i f_i(x^*) = 0, \quad i = 1, \dots, m. \qquad \text{(Compl. slackness)}$$

Proof: By the previous result, we know that there exists $\tilde{\lambda}_0, \dots, \tilde{\lambda}_m \geq 0$ not all zero, such that

$$\tilde{\lambda}_0 \nabla f_0(x^*) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(x^*) = 0$$

$$\tilde{\lambda}_i f_i(x^*) = 0 \quad , \quad i=1,\dots,m.$$

If $\tilde{\lambda}_0 = 0$, then the first condition yields: $\sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(x^*) = 0$.

For $i \in I(x^*)$, we have $f_i(\tilde{x}^*) = 0$, hence $\tilde{\lambda}_i$ may be positive; while for $i \notin I(x^*)$, $\tilde{\lambda}_i = 0$.

Hence, $\sum_{i \in I(x^*)} \tilde{\lambda}_i \nabla f_i(x^*) = 0$ has to hold. Since $\tilde{\lambda}_i$ are not all zero, there is at least one $i \in I(x^*)$ with $\tilde{\lambda}_i \neq 0$. But then $\{ \nabla f_i(x^*) \mid i \in I(x^*) \}$ are linearly dependent. Since this contradicts the assumption of the theorem, $\tilde{\lambda}_0 \neq 0$ has to hold.

Let's define $\lambda_i := \dfrac{\tilde{\lambda}_i}{\tilde{\lambda}_0} \Rightarrow \lambda_0 = 1, \ \lambda_i = \dfrac{\tilde{\lambda}_i}{\tilde{\lambda}_0}, \ i=1,\dots,m.$

Rewriting the conditions by multiplying them with $\frac{1}{\tilde{\lambda}_0}$, we obtain:

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) = 0$$

$$\lambda_i f_i(x^*) = 0 \, , \quad i=1,\dots,m$$

$$\lambda_1, \dots, \lambda_m \geq 0$$

$\blacksquare$

Inequality & equality constrained problems

Consider the problem:
$$\left(\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le 0, & i=1,\dots,m \\ & h_j(x) = 0, & j=1,\dots,p. \end{array}\right) \quad (P)$$

where $f_0, f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$, $h_1, \dots, h_p : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable.

Definition: A feasible point x^* (i.e., $f_i(x^*) \le 0 \ \forall i$, $h_j(x^*) = 0, \ \forall j$) is called <u>regular</u> if the gradients of "the active constraints" are linearly independent, that is,

$$\{\nabla f_i(x^*) \mid i \in I(x^*)\} \cup \{\nabla h_j(x^*) \mid j = 1, \dots, p\}$$

are linearly independent vectors.

For problem (P), we write the Lagrangian as follows:

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x).$$

Taking the derivative with respect to x, we have:

$$\nabla_x L(x, \lambda, \mu) = \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{j=1}^{p} \mu_j \nabla h_j(x).$$

The KKT conditions for problem (P) are as follows:

(1) Primal feasibility: $f_i(x^*) \leq 0$, $h_j(x^*) = 0$ for all i, j.

(2) (Dual feasibility): $\lambda_i \geqslant 0$ for all $i = 1, \dots, m$

(3) Complementary slackness: $\lambda_i f_i(x^*) = 0$ for all $i = 1, \dots, m$.

(4) $\nabla_x L(x, \lambda, \mu) = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) = 0$.

Definition: A point $x^* \in \mathbb{R}^n$ is called a KKT point for problem (P) if there exists $\lambda \in \mathbb{R}^m$ $(\lambda_1, \dots, \lambda_m \in \mathbb{R})$, $\mu \in \mathbb{R}^p$ $(\mu_1, \dots, \mu_p \in \mathbb{R})$ satisfying the KKT conditions (1)-(4).

Theorem: (necessary condition for local optimality)
Consider problem (P). Let x^* be a local minimum. Suppose that x^* is a regular point for (P). Then, x^* is a KKT point.

This theorem states that KKT conditions are necessary conditions for regular points of the problem. To use this result for solving problems, one needs to check irregular points (if any), separately.

As usual, one has to check if there exists a solution to the problem. If that is the case, then the solution has to be among the KKT points or the irregular points (if any).

Example: minimize $x_1 + x_2$

subject to $x_1^2 + x_2^2 = 1$.

Note that the feasible region of the problem $\{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}$ is closed and bounded. Moreover, the objective function is continuous. Then, by Weierstrass theorem, we know that there exists a solution. From the previous result, it's either a KKT point or an irregular point.

Let's check for irregular points first. The only constraint is an equality constraint, hence it's active: $h_1(x) = x_1^2 + x_2^2 - 1$.

Then, a point is irregular if $\nabla h(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ is linearly dependent.

A vector in $\mathbb{R}^2$ is linearly dependent only if it's the zero vector, that is, $\nabla h(x) = 0 \in \mathbb{R}^2$. This would yield $2\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Rightarrow x_1 = x_2 = 0$.

Hence, $0 \in \mathbb{R}^2$ is the only irregular point. Note that $0 \in \mathbb{R}^2$ is not feasible for the problem. We conclude that all feasible points for (P) are regular. Hence, an optimal solution has to be among KKT points.

$L(x,\mu) = x_1 + x_2 + \mu(x_1^2 + x_2^2 - 1)$. Note that as we don't have inequality constraints in the problem, conditions (2) & (3) of KKT are not relevant. KKT conditions are: (1) $x_1^2 + x_2^2 = 1$

(2) $\nabla_x L(x,\mu) = \begin{bmatrix} 1 + 2\mu x_1 \\ 1 + 2\mu x_2 \end{bmatrix} = 0 \in \mathbb{R}^2$.

Note that if $\mu = 0$, (2) would yield $\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0$, which is wrong.

Hence, $\mu \neq 0$ has to hold. Then, (2) imply: $x_1 = \frac{-1}{2\mu}$, $x_2 = \frac{-1}{2\mu}$.

In particular, we have $x_1 = x_2$. Together with (1): $2x_1^2 = 1 \Rightarrow x_1^2 = \frac{1}{2}$

$\Rightarrow x_1 = \pm \frac{1}{\sqrt{2}}$.

There are 2 KKT points: $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and $-\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

$f_0(x) = x_1 + x_2 \Rightarrow f_0\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \sqrt{2}$, $f_0\left(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right) = -\sqrt{2}$.

The minimum is attained at $\hat{x} = \left(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$, then it's the optimal solution.

Ex: let's consider the same problem, posed as follows:

$$\begin{aligned} &\text{minimize} \quad x_1 + x_2 \\ &\text{subject to} \quad (x_1^2 + x_2^2 - 1)^2 = 0. \end{aligned}$$

$h(x) = (x_1^2 + x_2^2 - 1)^2 \Rightarrow \nabla h(x) = \begin{bmatrix} 4x_1(x_1^2 + x_2^2 - 1) \\ 4x_2(x_1^2 + x_2^2 - 1) \end{bmatrix}$ is the gradient of the single active constraint.

Note that for any feasible $x \in \mathbb{R}^2$, we have $x_1^2 + x_2^2 = 1$. Then, for all feasible x, $\nabla h(x) = 0 \in \mathbb{R}^2$, which is linearly dependent. This means that any feasible solution is irregular! (The problem is NOT well-posed.)

On the other hand, if we write KKT points:

1) $(x_1^2 + x_2^2 - 1)^2 = 0 \Rightarrow x_1^2 + x_2^2 = 1$

2) $L(x,\mu) = x_1 + x_2 + \mu (x_1^2 + x_2^2 - 1)^2$

$$\nabla_x L(x,\mu) = \begin{bmatrix} 1 + 4\mu x_1 (x_1^2 + x_2^2 - 1) \\ 1 + 4\mu x_2 (x_1^2 + x_2^2 - 1) \end{bmatrix} = 0$$

There's no KKT point as the two conditions cannot be satisfied simultaneously.

Example: minimize $x_1^2 + 3x_2 - x_3$

subject to $x_1^2 + x_2^2 + x_3^2 = 6$

$x_1^2 + 2x_2 + 2x_3 = 8$

Note that the feasible region is closed and bounded. Indeed,

$$S = \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 = 6, \underbrace{x_1^2 + 2x_2 + 2x_3 = 8}_{\text{also closed.}}\} \subseteq \underbrace{\{x \in \mathbb{R}^3 \mid \|x\|_2^2 = 6\}}_{\text{closed \& bounded.}}.$$

Then, by Weierstrass theorem, there exists a solution to this problem.

Let's check irregular points first:

$h_1(x) = x_1^2 + x_2^2 + x_3^2 - 6 \Rightarrow \nabla h_1(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{bmatrix}$

$h_2(x) = x_1^2 + 2x_2 + 2x_3 - 8 \Rightarrow \nabla h_2(x) = \begin{bmatrix} 2x_1 \\ 2 \\ 2 \end{bmatrix}$

They are linearly dependent if $\nabla h_1(x) = k \cdot \nabla h_2(x)$ for some $k \in \mathbb{R}$.

Note that if $x_1 \neq 0$, then $\nabla h_1(x) = k \cdot \nabla h_2(x)$ may occur only for $k = 1$.

Then, $x_2 = x_3 = 1$ yield linearly dependent vectors.

Let's check feasibility:

$$\left.\begin{array}{l} h_1(x_1,1,1) = x_1^2+1+1=6 \Rightarrow x_1^2=4 \\ h_2(x_1,1,1) = x_1^2+2+2=8 \Rightarrow x_1^2=4 \end{array}\right\}$$

Thus, $x_1=\pm 2$, $x_2=x_3=1$ is feasible.

Two irregular feasible points from this case: $(2,1,1)$ & $(-2,1,1)$

If $x_1=0$, then $\nabla h_1(x) = \begin{bmatrix} 0 \\ 2x_2 \\ 2x_3 \end{bmatrix}$, $\nabla h_2(x) = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$ are linearly dependent

as long as $x_2=x_3=k \in \mathbb{R}$.

Let's check feasibility:

$$\left.\begin{array}{l} h_1(0,k,k) = 2k^2 = 6 \\ h_2(0,k,k) = 2k+2k=8 \end{array}\right\} \left.\begin{array}{l} k^2=3 \\ k=2 \end{array}\right\}$$

has no solution.

$(0,k,k)$ is not feasible for any k.

Now, let's check KKT points:

$$L(x,\mu) = x_1^2+3x_2-x_3+\mu_1(x_1^2+x_2^2+x_3^2-6)+\mu_2(x_1^2+2x_2+2x_3-8)$$

$$\nabla_x L(x,\mu) = \begin{bmatrix} 2x_1(\mu_1+\mu_2+1) \\ 3+2\mu_1x_2+2\mu_2 \\ -1+2\mu_1x_3+2\mu_2 \end{bmatrix} = 0 \in \mathbb{R}^3.$$

Note that $\mu_1 \neq 0$ since we have $\begin{pmatrix} 2\mu_2 = -3-2\mu_1x_2 \\ 2\mu_2 = 1-2\mu_1x_3 \end{pmatrix}$ and it would lead to a contradiction.

Since $\mu_1 \neq 0$, we can write $x_2 = \dfrac{-3-2\mu_2}{2\mu_1}$, $x_3 = \dfrac{1-2\mu_2}{2\mu_1}$ ⭑

For the first equation, there are two cases: $x_1 = 0$ OR $\mu_1 + \mu_2 = -1$.

- If $x_1 = 0$, then $x_2^2 + x_3^2 = 6$

$$2(x_2 + x_3) = 8 \Rightarrow x_2 + x_3 \overset{(*)}{=} \frac{-3-2\mu_2}{2\mu_1} + \frac{1-2\mu_2}{2\mu_1} = 4$$

$$\Rightarrow 2\mu_2 = -1 - 4\mu_1 \quad (1)$$

$$x_2^2 + x_3^2 = 6 \overset{(*)}{\Rightarrow} \left(\frac{3+2\mu_2}{2\mu_1}\right)^2 + \left(\frac{1-2\mu_2}{2\mu_1}\right)^2 = 6$$

$$\Rightarrow (3+2\mu_2)^2 + (1-2\mu_2)^2 = 24\mu_1^2$$

$$\overset{(1)}{\Rightarrow} (2-4\mu_1)^2 + (2+4\mu_1)^2 = 24\mu_1^2$$

$$\Rightarrow 8 + 32\mu_1^2 = 24\mu_1^2 \Rightarrow \mu_1^2 = -1 \Rightarrow \text{No real solution!}$$

- If $\mu_1 + \mu_2 = -1$, from $(*)$, $x_2 = \dfrac{-3+2+2\mu_1}{2\mu_1} = 1 - \dfrac{1}{2\mu_1}$
  $(\mu_1 = -1 - \mu_2)$

$$x_3 = \frac{1+2+2\mu_1}{2\mu_1} = 1 + \frac{3}{2\mu_1}$$

$$x_1^2 + 2x_2 + 2x_3 = 8 \Rightarrow x_1^2 + 2 - \frac{1}{\mu_1} + 2 + \frac{3}{\mu_1} = 8 \Rightarrow x_1^2 = 4 - \frac{2}{\mu_1}$$

$$x_1^2 + x_2^2 + x_3^2 = 6 \Rightarrow 4 - \frac{2}{\mu_1} + \left(1 - \frac{1}{2\mu_1}\right)^2 + \left(1 + \frac{3}{2\mu_1}\right)^2 = 6$$

→ Check that this is not possible!

So, we see that there's no KKT point for this problem.

There are two irregular points: $(2,1,1)$ and $(-2,1,1)$.

Check that they yield the same objective function value. Hence both are optimal. The value of the problem is 6.

<u>Question:</u> Is it possible to simplify the problem formulation?

$$\text{minimize} \quad x_1^2 + 3x_2 - x_3$$

$$\text{s.t.} \quad \left.\begin{aligned} x_1^2 + x_2^2 + x_3^2 &= 6 \\ x_1^2 + 2x_2 + 2x_3 &= 8 \end{aligned}\right\}$$

$$x_1^2 = 8 - 2x_2 - 2x_3$$

$$8 - 2x_2 - 2x_3 + x_2^2 + x_3^2 = 6 \quad \text{(only constraint)}$$

$$8 - 2x_2 - 2x_3 + 3x_2 - x_3 \quad \text{(objective function)}$$

$$\text{minimize} \quad x_2 - 3x_3 \quad (+8)$$

$$\text{s.t.} \quad \underbrace{x_2^2 - 2x_2 + 1}_{} + \underbrace{x_3^2 - 2x_3 + 1}_{} = 0$$

$$(x_2 - 1)^2 + (x_3 - 1)^2 = 0$$

The only feasible point is indeed $x_2 = x_3 = 1$. This yields the same optimal objective function value 6.

The Convex Case

Theorem: Consider the convex problem

$$(P)\begin{pmatrix} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0 & \forall i \\ & h_j(x) = 0 & \forall j \end{pmatrix}$$

where $f_0, f_1, \dots, f_m$ are cont.ly differentiable convex and $h_1, \dots, h_p$ are affine functions.

If x^* satisfies KKT conditions, then it's optimal.

Proof: Assume x^* satisfies KKT conditions; $\exists \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p \in \mathbb{R}$ s.t.

1.) Primal feasibility: $f_i(x^*) \leq 0 \quad \forall i, \quad h_j(x) = 0 \quad \forall j$

2.) Dual feasibility: $\lambda_i \geq 0 \quad \forall i$

3.) Complementary slackness: $\lambda_i f_i(x^*) = 0 \quad \forall i$

4.) $\nabla_x L(x^*, \lambda, \mu) = 0 \qquad \left(\nabla f_0(x^*) + \sum_i \lambda_i \nabla f_i(x^*) + \sum \mu_j \nabla h_j(x^*) = 0\right).$

Let's consider the Lagrangian for the fixed λ, μ (only as a function of x):

$$\mathcal{L}(x) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x).$$

Note that as f_i, $i = 0, 1, \dots, m$ are convex, $\lambda_i \geq 0$ and $h_j(x)$ are affine, $\mathcal{L}(x)$ is a convex function of x. Hence, any stationary point is global minimum. Note that by condition 4.), $\nabla \mathcal{L}(x^*) = 0$. Hence, x^* is a global minimum for $\mathcal{L}(x)$.

Next, we will show that $f_0(x^*) \leq f_0(x)$ for any feasible x.

Indeed, for any feasible x, we have $\mathcal{L}(x^*) \leq \mathcal{L}(x)$ from above.

That is, for any feasible x, we have

$$\underbrace{f_0(x^*) + \underbrace{\underbrace{\sum_{i=1}^{m} \lambda_i f_i(x^*)}_{\substack{0 \\ \text{from (3)}}} + \underbrace{\sum_{j=1}^{p} \mu_j h_j(x^*)}_{\substack{0 \\ \text{from (1)}}}}_{=0}}_{f_0(x^*)} \leq f_0(x) + \underbrace{\sum_{i=1}^{m} \underbrace{\lambda_i}_{\geq 0} \underbrace{f_i(x)}_{\substack{\leq 0 \\ \text{as } x \\ \text{is feasible}}}}_{\leq 0} + \sum_{j=1}^{p} \mu_j \underbrace{h_j(x)}_{\substack{=0 \\ \text{as } x \text{ is} \\ \text{feasible}}} \leq f_0(x).$$

$$f_0(x^*) \leq f_0(x)$$

This shows that x^* is optimal for problem (P). ∎

Ex: Consider problem

$$\text{minimize } 4x_1^2 + x_2^2 - x_1 - 2x_2$$
$$\text{subject to } 2x_1 + x_2 \leq 1$$
$$x_1^2 \leq 1.$$

Note that $f_0(x) = 4x_1^2 + x_2^2 - x_1 - 2x_2$, $f_1(x) = 2x_1 + x_2 - 1$, $f_2(x) = x_1^2 - 1$ are all convex functions. Hence, this is a convex problem. KKT conditions are sufficient for optimality:

1- Primal feasibility: $2x_1 + x_2 \leq 1$, $x_1^2 \leq 1$

2. Dual feasibility: $\lambda_1 \geq 0$, $\lambda_2 \geq 0$

3. Complementary slackness: $\lambda_1(2x_1 + x_2 - 1) = 0$, $\lambda_2(x_1^2 - 1) = 0$

4. Lagrangian: $L(x, \lambda) = 4x_1^2 + x_2^2 - x_1 - 2x_2 + \lambda_1(2x_1 + x_2 - 1) + \lambda_2(x_1^2 - 1)$

$$\nabla_x L(x, \lambda) = \begin{bmatrix} 8x_1 - 1 + 2\lambda_1 + 2\lambda_2 x_1 \\ 2x_2 - 2 + \lambda_1 \end{bmatrix} = 0 \in \mathbb{R}^2.$$

By considering condition 3, we consider different cases:

Case 1: $\lambda_1 = 0, \lambda_2 = 0$.

Then, from condition 4, we have $x_1 = \frac{1}{8}, x_2 = 1$.

Note that $2x_1 + x_2 = \frac{1}{4} + 1 \nleq 1$. Hence, this is not feasible. No KKT points from this case.

Case 2: $\lambda_1 = 0, x_1^2 = 1$. Hence, $x_1 = 1$ or $x_1 = -1$.

From condition 4, we obtain $x_2 = 1$. Two canditate points: $(1,1), (-1,1)$

Note that for $(1,1)$, we have

$2x_1 + x_2 = 3 \nleq 1$. Hence, it's not feasible.

Consider $(-1,1)$: $2x_1 + x_2 = -2+1 = -1 \leq 1$ ✓ ; $x_1^2 = 1 \leq 1$ ✓ — Primal feasibility is satisfied.

We still need to check conditions 2 & 4.

From 4, we have $8x_1 - 1 + 2\lambda_1 + 2\lambda_2 x_2 = -8 - 1 + 2\lambda_2(-1)$

$$= -9 - 2\lambda_2 = 0$$

Then, $\lambda_2 = \frac{-9}{2} \ngeq 0$. Not dual feasible. No KKT points from this case!

Case 3: $\lambda_2 = 0, 2x_1 + x_2 = 1$.

Condition 4: $8x_1 + 2\lambda_1 - 1 = 0 \Rightarrow x_1 = \frac{1-2\lambda_1}{8}$

$2x_2 - 2 + \lambda_1 = 0 \Rightarrow x_2 = \frac{2-\lambda_1}{2}$

$\frac{1-2\lambda_1}{4} + \frac{2-\lambda_1}{2} = 1$

$\Rightarrow 1 - 2\lambda_1 + 4 - 2\lambda_1 = 4$

$\Rightarrow \lambda_1 = \frac{1}{4} \geq 0$ ✓

(Condition 2 is satisfied)

Let's check condition 1: $x_1 = \frac{1-\frac{1}{2}}{8} = \frac{1}{16}$ $(x_1^2 \leq 1$ ✓$)$

$x_2 = 1 - 2x_1 = \frac{7}{8}$. Hence, $\hat{x} = \left(\frac{1}{16}, \frac{7}{8}\right)$ is a KKT point.

Since the problem is convex, we don't need to check the last case, i.e. $(2x_1 + x_2 = 1,\ x_1^2 = 1)$.

$\hat{x} = \left(\frac{1}{16}, \frac{7}{8}\right)$ is an optimal solution.

Considering problem (P)

$$\begin{aligned} \text{minimize } & f_0(x) \\ \text{subject to } & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p. \end{aligned}$$

Until now, we have seen that under some assumptions on the problem, KKT conditions may be sufficient and/or necessary.

- If f_i, $i = 1, \ldots, m$ and h_j, $j = 1, \ldots, p$ are all linear, then KKT is necessary. (f_0 can be linear/convex/nonconvex as long as it's continuously diff.)
- If $f_0, f_1, \ldots, f_m$ are convex, $h_1, \ldots, h_p$ are linear, then KKT is sufficient
- For nonconvex case, KKT is necessary for regular points only.

Next, we will see another necessary condition for which the constraints are as in convex programming but the objective function is free to be non-convex.

Defn: Consider problem (P) for which $f_1, f_2, \ldots, f_m$ are convex and $h_1, h_2, \ldots, h_p$ are linear (affine). If there exists a feasible point x^* such that $f_i(x^*) < 0$ for all $i = 1, \ldots, m$, then x^* is called a Slater point. In this case, we say that the Slater's condition is satisfied.

Result: If for problem (P), the Slater's condition is satisfied, then KKT conditions are necessary for local optimality.

Example: Consider problem:

$$(P) \quad \text{maximize} \quad (x_1+1)^2 + (x_2+1)^2 \quad \text{subject to} \quad x_1^2 + x_2^2 \leq 8$$

First, let's write the problem in minimization form:

$$\boxed{\begin{array}{ll} \text{minimize} & -(x_1+1)^2-(x_2+1)^2 \\ \text{s.t.} & x_1^2+x_2^2 \leq 8 \end{array}}$$

This is not a convex program as $f_0(x) = -(x_1+1)^2-(x_2+1)^2$ is not convex. However, the constraints are convex. The Slater's condition is also satisfied, consider e.g. $x=(0,0)$. Then, KKT conditions are necessary. Feasible region is closed and bounded, hence there's a solution, which has to be among KKT points:

1.) $x_1^2+x_2^2 \leq 8$

2.) $\lambda \geq 0$

3.) $\lambda(x_1^2+x_2^2-8) = 0$

4.) $L(x,\lambda) = -(x_1+1)^2-(x_2+1)^2+\lambda(x_1^2+x_2^2-8)$

$$\nabla_x L(x,\lambda) = \begin{bmatrix} -2(x_1+1)+2\lambda x_1 \\ -2(x_2+1)+2\lambda x_2 \end{bmatrix} = 0$$

If $\lambda=0$, then $x_1 = x_2 = -1$ from condition 4. This is both primal & dual feasible. Hence a KKT point.

If $x_1^2+x_2^2=8$, $\lambda \neq 0$ $\Rightarrow$ from 4, we obtain: $\left.\begin{array}{l}(\lambda-1)x_1 = 1 \\ (\lambda-1)x_2 = 1\end{array}\right\} \; x_1=x_2=\dfrac{1}{\lambda-1}$

$2x_1^2 = 8 \Rightarrow x_1^2=4 \Rightarrow x_1 = \pm 2, \; x_2 = \pm 2 \; (x_1=x_2)$.

If $x_1=2 \Rightarrow 2 = \dfrac{1}{\lambda-1} \Rightarrow \lambda = 1+\dfrac{1}{2} \geq 0$ ✓ $\Rightarrow (2,2)$ is a KKT point

$x_1=-2 \Rightarrow -2 = \dfrac{1}{\lambda-1} \Rightarrow \lambda = 1-\dfrac{1}{2} \geq 0$ ✓ $\Rightarrow (-2,-2)$ is a KKT point.

There are 3 KKT points: $(-1,-1)$, $(-2,-2)$, $(2,2)$.

$(x_1+1)^2+(x_2+1)^2$ is maximum at $(2,2)$, hence it's the optimal solution for this problem.