Constrained Optimization

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E}, \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}$$

equality const.
inequality const.

Feasible set: $\Omega = \{x \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I}\}$

so, $\min_{x \in \Omega} f(x)$

Definition: The active set $\mathcal{A}(x)$ at any feasible x consists of the equality const. indices from $\mathcal{E}$ together with the indices of the inequality cons. i for which $c_i(x) = 0$; that is

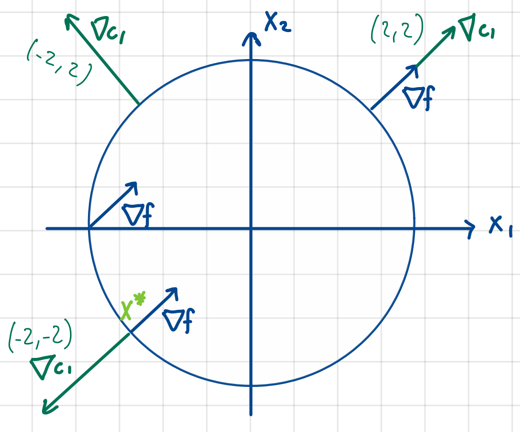
$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(x) = 0\}$$

At a feasible point x, the inequality cost. $i \in \mathcal{I}$ is said to be **active** if $c_i(x) = 0$ and **inactive** if the strict inequality $c_i(x) > 0$ is satisfied.

○ A Single Equality Constraint

$$\min\ x_1 + x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0$$

Here $f(x) = x_1 + x_2$ $\quad \mathcal{I} = \emptyset \quad \mathcal{E} = \{1\} \quad c_1 = x_1^2 + x_2^2 - 2$



$$\nabla f = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \nabla c_1 = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

Feasible set for this problem is circle of radius $\sqrt{2}$ centered at origin. (Just boundary not interior)

The solution x^* is obviously $(-1,-1)^T$. See that at x^*, the constraint normal $\nabla c_1(x^*)$ is parallel to $\nabla f(x^*)$. There is a scalar λ_1^* (here $\lambda_1^* = -1/2$)

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$$

By introducing the **Lagrangian function**:

$$\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x)$$

Lagrange Multiplier

$$\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x)$$

At solution x^*, there is a scalar λ_1^* such that

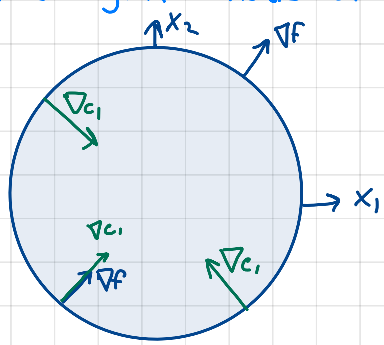
★ $\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0$ ← This is **necessary** but **not sufficient.**

$x\,(1,1)$ with $\lambda_1 = 1/2$ satisfies ★ but it is not optimal solution!

○ A Single Inequality Constraint

$$\min \quad x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0$$

Now, feasible region consists of the circle and its interior!



$$\nabla f = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \nabla c_1 = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}$$

Constraint normal ∇c_1 points toward the interior of the feasible region at each point on the boundary of the circle.

A given feasible point $\mathbf{x}$ is not optimal if we can find a small step $\mathbf{s}$ that both retains feasibility and decreases the objective function f to first order.

C1: ○ The step $\mathbf{s}$ improves the objective function, to first order, if $\nabla f(x)^T s < 0$.

$$f(x+s) - f(x) < 0$$

$$\cancel{f(x)} + \nabla f(x)^T s - \cancel{f(x)} < 0$$

$$\nabla f(x)^T s < 0$$

C2: ○ $\mathbf{s}$ retains feasibility if $\quad 0 \leq c_1(x+s) \approx c_1(x) + \nabla c_1(x)^T s$

$$⊛ \quad c_1(x) + \nabla c_1(x)^T s \geq 0.$$

Case I: x lies strictly inside the circle, $c_1(x) > 0$.

Any step vector s satisfies the condition ⊛, provided only that its length is sufficiently small.

Whenever $\nabla f(x) \neq 0$, we can obtain a step s that satisfies both C1, C2

$$s = -\alpha \nabla f(x) \quad \text{for any positive scalar } \alpha \text{ sufficiently small.}$$

However, no step s is given when $\nabla f(x) = 0$.

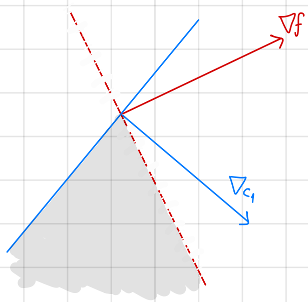
Case 2: x lies on the boundary of the circle. $c_1(x)=0$. Now, the conditions become

Open half-space → $\nabla f(x)^T s < 0$

Closed half-space → $\nabla c_1(x)^T s \geq 0$

$$c_1(x+s) \geq 0$$
$$\approx$$
$$c_1(x) + \nabla c_1(x)^T s \geq 0$$

lies on the boundary, so it is zero.



If ∇f and ∇c_1 point in the same direction,
$\nabla f = \lambda_1 \nabla c_1$, for some $\lambda_1 \geq 0$,
then the intersection of these two regions is **empty!**

$\nabla f(x) = \lambda_1 \nabla c_1(x)$, if λ_1 is negative here ∇f and ∇c_1 would point opposite direction and make up an entire open half-plane

\# Optimality Conditions for both Case I and 2.

$$\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x)$$

When no first order feasible descent direction exists at some point x^*, we have that

converged!

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0 \quad \text{for some} \quad \lambda_1^* \geq 0.$$

we also require that

$$\lambda_1^* c_1(x^*) = 0.$$ → known as complementarity condition

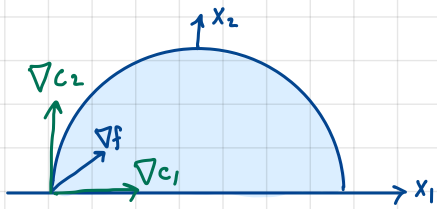
This implies that the Lagrange multiplier λ_1 can be strictly positive only when the corresponding c_1 is active.

For example, in Case 1, $c_1(x^*) > 0$ requires that $\lambda_1^* = 0$.

In Case 2, λ_1^* a nonnegative value.

○ Two Inequality Constraints

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



$$\nabla f = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \nabla c_1 = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix} \quad \nabla c_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\mathcal{L}(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x) \qquad \lambda = (\lambda_1, \lambda_2)^T$$

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \quad \text{for some} \quad \lambda^* \geq 0$$

$$\lambda_1^* c_1(x) = 0 \qquad \lambda_2^* c_2(x) = 0$$

When $x^* = (-\sqrt{2}, 0)^T$, we have

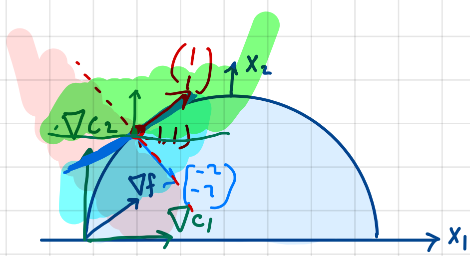
$$\nabla f(x^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \nabla c_1(x^*) = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix} \quad \nabla c_2(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

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

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \lambda_1^* \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix} - \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\underbrace{\lambda_1^* = \frac{1}{2\sqrt{2}}}_{\text{Positive}} \qquad \underbrace{\lambda_2^* = 1}_{\text{Positive}}$$

Let's check the point $x(-1, 1)$



$$\nabla f = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \nabla c_1 = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix} \quad \nabla c_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \nabla c_1 = \begin{pmatrix} 2 \\ -2 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \lambda_1 \begin{bmatrix} 2 \\ -2 \end{bmatrix} - \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

$$1 - 2\lambda_1 = 0 \qquad \lambda_1 = \frac{1}{2}$$

$$1 - \frac{1}{2}(-2) - \lambda_2 = 0$$

$$2 - \lambda_2 = 0$$

$$\lambda_2 = 2$$

λ_1 and λ_2 are positive!

λ_2 can not be positive, since

c_2 is <u>not</u> active at $x(-1, 1)$.

$$\lambda_2 \cdot c_2(x) = 0$$

$$2 \cdot 1 \neq 0 \quad \times \quad \Rightarrow \quad x(-1,1) \text{ is not the optimal solution.}$$

Given the point x and the active set $\mathcal{A}(x)$, 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.

Numerical Optimization
Nocedal & Wright
Page: 321

Theorem 12.1: First-Order Necessary Conditions

It is called first-order, because they are concerned with properties of gradients (first-derivative vectors).

Suppose that x^* is a local solution of $\boxed{\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \begin{cases} c_i(x) = 0, & i \in \mathcal{E}, \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}}$ that the functions f and c_i 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^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$
$$c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E},$$
$$c_i(x^*) \geq 0, \quad \text{for all } i \in \mathcal{I},$$
$$\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I},$$

Complementarity Conditions
$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}.$$

These conditions are often known as the Karush-Kuhn-Tucker (**KKT**) conditions.

If exactly one of λ_i^* and $c_i(x^*)$ is zero for each index $i \in \mathcal{I}$.
e.g., we have that $\lambda_i^* > 0$ for each $i \in \mathcal{I} \cap \mathcal{A}(x)$. → Strict Complementarity

⚠ When LICQ holds, the optimal λ^* is unique!