Constrained Optimization

Dr Ramkrishna Pasumarthy

Professor, IIT Madras

DSA 2026

DSA July 2026

Overview [1]

[1] This content is largely based on the book "Nocedal, J., Wright, S. J. (2006). Numerical optimization (2nd ed.). Springer Science+Business Media.

Formulation

- There are now equality and inequality constraints in the values that the argument can take

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

- $\mathcal{E}$: index for **equality constraints**
- $\mathcal{I}$: index for **inequality constraints**
- The theory for constrained optimization is built by looking at three examples

Example 1: Single Equality Constraint

$$min_{x_1,x_2}\ (x_1 + x_2) \text{ subject to } x_1^2 + x_2^2 - 2 = 0$$

- $f(x) = x_1 + x_2$
- $\mathcal{I} = \phi$
- $\mathcal{E} = \{1\}$
- $c_1(x) = x_1^2 + x_2^2 - 2$
- Intuition tells that the constraint set is a circle and the global minimum is $(-1, -1)$ as that is the point possible in the circle where $x_1 + x_2$ is the most negative and hence lowest
- From any other point, it is feasible to move in the circle to decrease f - simply move anti-clockwise or clockwise in the direction towards $(-1, -1)$

Geometry of the Problem

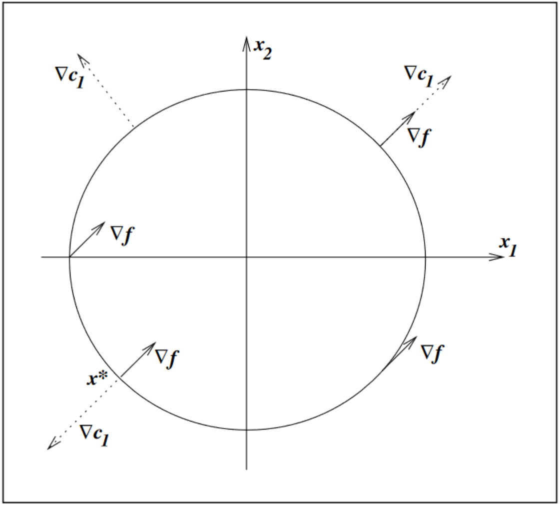

Figure: ∇f parallel to ∇c_1 at the minimum $(-1, -1)$

Observations

- One observes that at the minimum $x^* = (-1, -1)$, the constraint normal $\nabla c_1(x^*)$ is parallel to $\nabla f(x^*)$
- Or in other words, there exists a scalar λ_1 such that

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

- It can be verified that $\lambda_1 = \frac{1}{2}$

Motivation for the Lagrangian

- One can derive the condition $\nabla f(x^*) = \lambda_1 \nabla c_1(x^*)$ from an unconstrained optimization by considering the problem

$$min_{x_1,x_2,\lambda_1} L = f(x) - \lambda_1 c_1(x)$$

- We have that the unconstrained first order conditions for this problem are

$$\begin{aligned} \nabla_x L = (\nabla_x f - \lambda_1 \nabla_x c_1) = 0 \\ \nabla_{\lambda_1} L = -c_1(x) = 0 \end{aligned}$$

- Both constraint and the gradient conditions come out of the function L called the **Lagrangian** and λ_1 , the **Lagrange multiplier** for c_1
- **Warning**: This being first order is only a necessary condition. Eg. $(1, 1)$ also satisfies the condition but is a maxium and not a minimum

Proof of the First Order Condition

- First Order Taylor Expansions:

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

- For c to be feasible (remain in the circle) upto first order, $c_1(x+s) - c_1(x) \approx 0$ and hence $\nabla c_1(x)^T s = 0$. We require that any feasible direction (satisfying $\nabla c_1(x)^T s = 0$) should lead to only increase of the function f.
- So, the component of the gradient along the direction where $\nabla c_1(x)^T s = 0$ must vanish
- The feasible direction can be characterised as $c_{feas} = \{\nabla c_1(x)\}^\perp$
- Hence we must have

$$\nabla f(x) - \frac{\nabla c_1(x) \nabla c_1(x)^T}{||\nabla c_1(x)||^2} \nabla f(x) = 0$$

- Hence , $\lambda_1 = \frac{\nabla c_1(x)^T \nabla f(x)}{||\nabla c_1(x)||^2}$

Example 2: Single Inequality Constraint

$$min_{x_1,x_2}\ (x_1 + x_2) \text{ subject to } 2 - x_1^2 - x_2^2 \geq 0$$

- Same as before: Equality replaced by inequality
- $f(x) = x_1 + x_2$
- $\mathcal{E} = \phi$
- $\mathcal{I} = \{1\}$
- $c_1(x) = 2 - x_1^2 - x_2^2$
- Intuition tells that the constraint set is a disc and the global minimum is still $(-1, -1)$ as that is the point possible in the disc where $x_1 + x_2$ is the most negative and hence lowest
- One notes that $-\lambda_1 = \frac{1}{2} \geq 0$

Geometry of the Problem

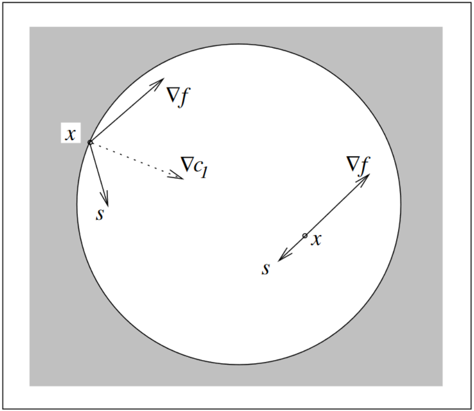

Figure:

- ∇c_1 is towards the interior of the feasible region
- ∇c_1 is parallel to ∇f at $(-1, -1)$
- s should also point in the interior of the feasible region

First Order Condition: In Interior

- Consider the case when x is inside the interior of the circle - it is possible to move in all directions from x. i.e. $x + s \in$ disc , for all s with $||s||$ small enough
- In this case, $c_1(x) < 0$
- Then by choosing $s = -\nabla f(x)$, one can decrease the value of the function as for unconstrained first order condition
- So, we must have

$$\nabla f(x) = 0 \text{ when } c_1(x) < 0$$

First Order Condition: In the Boundary

- At the boundary, $c_1(x) = 0$
- What is the restriction of s at the boundary?

$$\color{red}{0 \leq} c_1(x+s) \approx \underbrace{c_1(x)}_{=0 \text{ at boundary}} + \nabla c_1(x)^T s$$

- So, s should satisfy $\nabla c_1(x)^T s \geq 0$
- f should be decreasing in some direction when x is not a minimum - i.e. or

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

Geometry of the search direction constraints

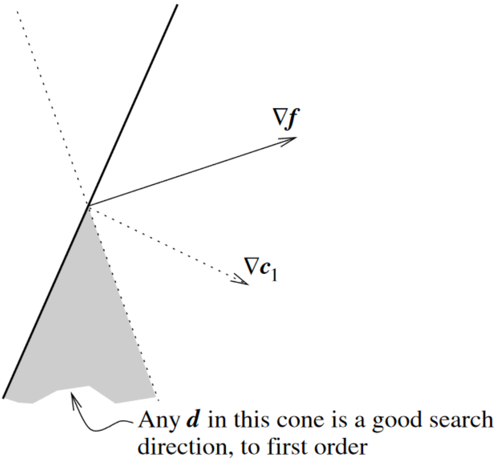

Figure: Intersection of the spaces $\nabla c_1(x)^T s \geq 0$ and $\nabla f(x)^T s < 0$

First Order Condition at Boundary

- The gray area where f can decrease if and only if $\nabla f(x)$ and $\nabla c_1(x)$ are parallel - point in the same direction. i.e.

$$\nabla f(x) = \lambda_1 c_1(x) \quad \text{where} \quad \lambda_1 \geq 0$$

So, we have a first order condition at the boundary $c_1(x) = 0$ that

$$\begin{aligned} c_1(x) &= 0 \\ \nabla f(x) &= \lambda_1 c_1(x) \quad \text{where} \quad \lambda_1 \geq 0 \end{aligned}$$

Summary of First Order Conditions

- **At boundary:**

$$c_1(x) = 0$$
$$\nabla f(x) = \lambda_1 c_1(x) \quad \text{where} \quad \lambda_1 \geq 0$$

- **At Interior:**

$$c_1(x) > 0$$
$$\nabla f(x) = 0$$

- The interior condition can be rewritten as
  At Interior:

$$c_1(x) > 0$$
$$\nabla f(x) = \lambda_1 \nabla c_1(x)$$
$$\lambda_1 = 0$$

Unified First Order Condition

Both the conditions can be summarized together as

$$
\begin{aligned}
c_1(x) &\geq 0 \\
\lambda_1 &\geq 0 \\
\nabla f(x) &= \lambda_1 \nabla c_1(x) \\
\lambda_1 c_1(x) &= 0
\end{aligned}
$$

Lagrangian

- By defining once again the Lagrange function as $L = f(x) - \lambda_1 c_1(x)$, the first order conditions can be summarized as

$$\begin{aligned} \nabla_{x,\lambda} L &= 0 \\ \text{for } \ \lambda_1 &\geq 0 \\ \lambda_1 c_1(x) &= 0 \end{aligned}$$

- $\lambda_1 c_1(x) = 0$: **complementary slackness condition**
- $c_1(x) = 0 \implies$ constraint is **active** [2]
- $\lambda_1 > 0 \implies$ Lagrange Multiplier is **active**
- Lagrange multiplier can be active only when the constraint is active - complementary
- **NOTE**: An equality constraint is always active !

[2] at the boundary, one is just about to violate the constraint!

Example 3: Two Inequality Constraints

$$\begin{aligned} min_{x_1,x_2}\ (x_1 + x_2) \text{ subject to} \\ x_1^2 + x_2^2 - 2 = 0 \\ x_2 \geq 0 \end{aligned}$$

- $f(x) = x_1 + x_2$
- $\mathcal{I} = \{1, 2\}$
- $\mathcal{E} = \phi$
- $c_1(x) = 2 - x_1^2 - x_2^2$ and $c_2(x) = x_2$
- Intuition tells that the constraint set is the upper semi-circlular disc and the global minimum is $(-\sqrt{2}, 0)$ as that is the point possible in the circle where $x_1 + x_2$ is the most negative and hence lowest
- **NOTE:** At $(-\sqrt{2}, 0)$, both constraints are active - $c_1(x) = c_2(x) = 0$

Geometry of the Problem

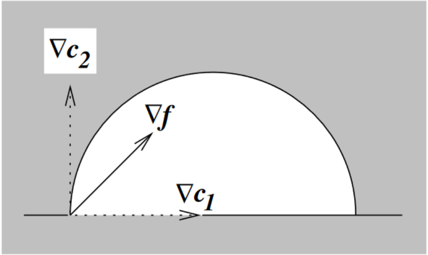

Figure: The gradient $\nabla f(x)$ at $(-\sqrt{2}, 0)$ is a non-negative combination of the gradients of c_1 and c_2 - points interior of the feasible set demanded by both the constraints

First Order Conditions - Generic Case - KKT Conditions

$$L = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

KKT Conditions at a local minimum:

- $\nabla_x L = \nabla f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla c_i(x) = 0$
- $c_i(x) \geq 0$ for all $x \in \mathcal{I}$
- $c_i(x) = 0$ for all $x \in \mathcal{E}$
- $\lambda_i \geq 0$ for all $x \in \mathcal{I}$
- $\lambda_i c_i(x) = 0$ for all $x \in \mathcal{I}$ (complementary slackness)

Second Order Sufficient Condition

- For the unconstrained case, it was noted that $\nabla^2 f(x)$ being positive definite guaranteed a minimum. So, we have

$$w^T(\nabla^2 f)w > 0 \quad \text{for all } w \neq 0$$

- Let C be the set of all possible search directions from x in the constrained case
- For constrained case we have,
Second Order Condition for constrained optimization:

$$w^T(\nabla_{xx} L)w > 0 \quad \text{for all } w \in C$$