Constraints
Let $f(x) = -x_1 - x_2$, $x \in \mathbb{R}^2$. min $f(x)$ yields x_1 and x_2 tending toward infinity. With constraints (subject to)

$$x_1^2 + x_2^2 - 1 = 0 \tag{1}$$

The feasible set is the unit circle.

Taylor series
A function can be approximated by

$$f(x+\delta) \approx f(x) + \nabla f(x)^T \delta \tag{2}$$

Use Taylor for the above problem

$$\begin{aligned} g(x) &= x_1^2 + x_2^2 - 1 \\ g(x^\star + \delta) &\approx g(x^\star) + \nabla g(x^\star)^T \delta \\ &= \nabla f(x^\star)^T \delta \end{aligned}$$

We have that

$$\nabla f(x^*)^T \delta = 0 \tag{3}$$

Set

$$\nabla f(x^*) = \lambda \nabla g(x^*), \qquad \lambda \in \mathbb{R} \tag{4}$$

Write a Langrangian

$$\begin{aligned} \mathbf{L}(x, \lambda) &= f(x) + \lambda g(x) \\ \mathbf{L}(x^*, \lambda^*) &= \nabla f(x^*) - \lambda \nabla g(x^*) = 0 \end{aligned}$$

Multiple equality constraints
$g_i(x) = 0, i = 1, \ldots, K$
The Jacobian is given by $D\bar{g}$, $\bar{g}$ being a vector.

$$D\bar{g}(x) = \begin{bmatrix} (\nabla g_1(x))^T \\ \vdots \\ (\nabla g_K(x))^T \end{bmatrix} \qquad \in \mathbb{R}^{K \times N} \tag{5}$$

Assume $D\bar{g}$ has full row rank

$$\bar{G}(x+\delta) \approx \bar{g} + D\bar{g}(x)\delta = 0$$

This is because of g being 0 for all x.

Let $G = \{\nabla g_8 x), \dots, \nabla g_K(x)\}$.
Let the orthogonal projection of $\nabla f(x^*)$ onto G be

$$\sum_{i=1}^{K} \lambda_i \nabla g_i(x^*) \tag{6}$$

This gives

$$\nabla f(x^*) = \sum_{i=1}^{K} \lambda_i \nabla g_i(x^*) + r \tag{7}$$

with r being the residual orthogonal to the projection. Choose $s = -r$. Is this a feasible step?

$$\begin{aligned} s^T \nabla f(x^*) &= s^T \left(\sum_{i=1}^{K} \lambda_i \nabla g(_i(x^*) - s \right) = -s^T s \\ &= -\|S\|_2^2 \end{aligned}$$

Start at an optimal point and take step s

$$f(x^* + s) \approx f(x^*) + \nabla f(x^*)^T s \tag{8}$$

The right term can not be negative, as the optimal point is not a minimum then. The norm is always positive, and s can therefore not be a feasible step.

Necessary condition for optimization

$$\nabla f(x^*) - \sum_{i=1}^{K} \lambda_i \nabla g_i(X^*) = 0 \tag{9}$$

Example
Write the lagrangian

$$\mathcal{L}(x, \lambda) = -x_1 - x_2 + \lambda(x_1^2 + x_2^2 - 1) \tag{10}$$

Minimize $\mathcal{L}(x, \lambda)$. Take the derivative $\nabla \mathcal{L}(x, \lambda)$.

$$\begin{aligned} \frac{\partial}{\partial x_1} \mathcal{L}(x, \lambda) &= -1 + 2\lambda x_1 = 0 \\ \frac{\partial}{\partial x_2} \mathcal{L}(x, \lambda) &= -1 + 2\lambda x_2 = 0 \end{aligned}$$

This means that $\lambda = \frac{1}{2x_2}$. From this we know, that $x_1 = x_2$. We know that $g(x) = 0$.

Inequality constraints
If the constraints are inequalities the following holds

$$\nabla f(x^*) = \sum_{i=1}^{K} \mu_i \nabla h_i(x^*) \tag{11}$$

where $h(x) \geq 0$ and $\mu \geq 0$

$$\begin{aligned} f(x^* + s) &\approx f(x^*) + (\nabla f(x^*))^T s \\ &= f(x^*) + ((Dh(x^*)^T \mu)^T s \\ &= f(x^*) + \mu^T \mathbf{e} \\ &= f(x^*) + \mu_1 \end{aligned}$$

KKT

$$\begin{aligned} D\mathcal{L}(x, \mu) &= 0 \\ \nabla f(x^*) &= \sum_{i=1}^{P} \mu - i \nabla h_i(x^*) \\ \mu_i &\geq 0, \qquad i = 1, \ldots, P \\ h_i(x^*) &\geq 0, \qquad \forall i \\ \mu_i h_i(x^*) &= 0, \qquad \forall i \end{aligned}$$

Linear Program
Minimize $f(x)$
s.t. $g(x) \geq 0$
Constraints and objectives have to be linear.

Minimize $c^T x + q$ subject to $Ax = b$.
Linearizing a Problem
Minimize $\|x\|_1$ subject to $Ax = b$. This is equal to minimizing $\sum_{i=1}^{K} t_i$ with

1. $x_i \leq t_i$
2. $-x_i \leq t_i$
3. $Ax = b$