# Lagrange Multipliers



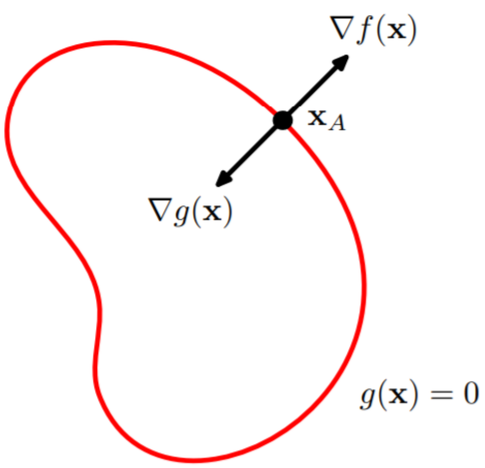
Lagrange multipliers, a.k.a. undetermined multipliers, are used to find **the stationary points** of a function of several variables subject to one or more constraints.

## Equality Constraint

Sometime it may be difficult to find a analytic solution of the constraint between variables $x_i$. We can introduce a parameter $\lambda$ called Lagrange multiplier. From a geometrical perspective, consider a $D$-dimensional variable $\mathrm{x} = (x_1, \ldots, x_D)^T$. The constraint equation $g(\mathrm{x}) = 0$ is a $(D-1)$-dimensional surface in $\mathrm{x}$-space.

> *Lemma-1*. At any point on the surface $g(\mathrm{x}) = 0$ the gradient $\nabla g$ will be orthogonal to the surface.
>
> *Proof*. Consider a point $\mathrm{x}$ on the surface, and consider a nearby point $\mathrm{x} + \epsilon$ that also on the surface. If we make a Taylor expansion around this point, we have:
>
> $$g(\mathrm{x} + \epsilon) \approx g(\mathrm{x}) + \epsilon^T \nabla g(\mathrm{x}) \tag{1}$$
>
> because $g(\mathrm{x} + \epsilon) = g(\mathrm{x})$, we have $\epsilon^T \nabla g(\mathrm{x}) \approx 0$. And if $||\epsilon|| \to 0$, we have $\epsilon^T \nabla g(\mathrm{x}) = 0$. And because at this time $\epsilon$ is parallel to the surface $g(\mathrm{x}) = 0$, so $\nabla g$ is normal to the surface.

A geometrical picture of the technique of Lagrange multipliers in which we seek to maximize a function $f(\mathbf{x})$, subject to the constraint $g(\mathbf{x}) = 0$. If $\mathbf{x}$ is $D$ dimensional, the constraint $g(\mathbf{x}) = 0$ corresponds to a subspace of dimensionality $D - 1$, indicated by the red curve. The problem can be solved by optimizing the Lagrangian function $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$.




> *Lemma-2.* If $\mathrm{x}^*$ on $g(\mathrm{x}) = 0$ and $\forall \mathrm{x}$ on $g(\mathrm{x}) = 0$, we have $f(\mathrm{x}^*) \geq f(\mathrm{x})$, then $\nabla f(\mathrm{x}^*)$ is orthogonal to the surface $g(\mathrm{x}) = 0$.
>
> *Proof*. Otherwise we could increase the value of $f(\mathrm{x})$ by moving a short distance along the surface.

From *Lemma-1* and *Lemma-2* we have:

$$\nabla f + \lambda \nabla g = 0 \tag{2}$$

where $\lambda \neq 0$ is the Lagrange multiplier.

Then we can define the *Lagrangian* function

$$L(\mathrm{x}, \lambda) := f(\mathrm{x}) + \lambda g(\mathrm{x}) \tag{3}$$

And

$$\begin{aligned} \nabla_{\mathrm{x}} L = 0 &\Rightarrow \nabla f + \lambda \nabla g = 0 \\ \frac{\partial L}{\partial \lambda} = 0 &\Rightarrow g(\mathrm{x}) = 0 \end{aligned}$$

Thus to find the maximum of $f$ s.t. $g(\mathrm{x}) = 0$, we solve the equation

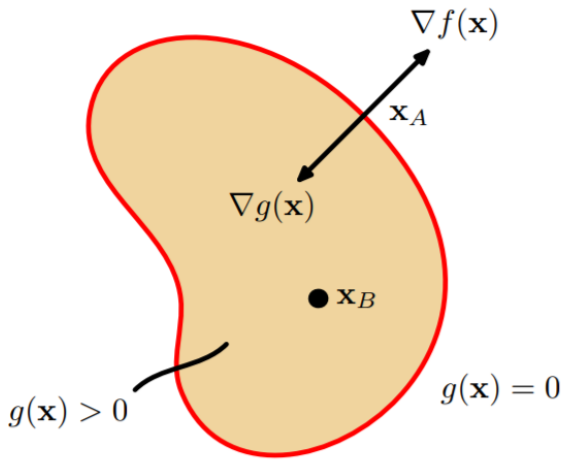
$$\begin{cases} \nabla_{\mathrm{x}} L = 0 & \Rightarrow \quad \nabla f + \lambda \nabla g = 0 \\ \frac{\partial L}{\partial \lambda} = 0 & \Rightarrow \quad g(\mathrm{x}) = 0 \end{cases} \tag{4}$$

and get a $(D+1)$-dimension equations to get stationary points.

## Inequality Constraint

The problem of maximizing $f$ with inequality constraint of the form $g(\mathrm{x}) \geq 0$.

Illustration of the problem of maximizing $f(\mathbf{x})$ subject to the inequality constraint $g(\mathbf{x}) \geqslant 0$.




There are two kinds of solution possible.

- *inactive* constraint: The stationary point lies in the region $g(\mathrm{x}) > 0$. The function $g(\mathrm{x})$ is useless, $\lambda = 0$.
- *active* constraint: The stationary point lies in the surface $g(\mathrm{x}) = 0$. This is analogous to the equality constraint case, $\lambda \neq 0$.

Note that now the sign of $\lambda$ is crucial, because $f$ only be at a maximum if its gradient is oriented away from the region $g(\mathrm{x}) > 0$ (At this time, the stationary point lies in the surface $g(x) = 0$. So the points in the region must has value less then the stationary point lies in the surface, which means the direction of gradient is oriented away from the region). Therefore $\lambda > 0$ (if we want a minimum, we set $\lambda < 0$).

For either cases, we have $\lambda g(\mathrm{x}) = 0$. Then we can define:

> **KKT(Karush-Kuhn-Tucker) conditions**

> $$g(\mathrm{x}) \geq 0$$
> $$\lambda \geq 0$$
> $$\lambda g(\mathrm{x}) = 0$$

And the solution is also straightforward.