Lagrange Multipliers.

Lagrange multipliers, also sometimes called undetermined multipliers, are used to find the stationary points of a function of ~~several~~ several ~~variables~~ variables subject to one or more constraints.

Let us say that we are finding the maximum of a function $f(x_1, x_2)$. subject a constraint relating x_1 and x_2, $g(x_1, x_2) = 0$

One solution can be

express x_2 as x_1 using the constraint

$$x_2 = h(x_1).$$

This can be substituted into $f(x_1, x_2)$ to give a function of x_1 alone in the form $f(x_1, h(x_1))$

The max w.r.t to x_1 can be found by differentiating the usual way.

$$x_2^* = h(x_1^*).$$

One problem with this approach is that it may be difficult to find an analytic solution of the constraint eq^n which allows x_2 to be expressed as an explicit function of x_1.

* — This approach treats x_1 and x_2 differently and spoils the natural symmetry b/w them.

—— x ——

A more elegant, and often simpler approach is based on 'λ' called Lagrangian Multiplier.

Considering things geometrically.

Let's assume a D-dimensional variable x with components $x_1, \ldots, x_D$.

The constraint eq^n $g(x) = 0$
it represents a (D−1) dimensional surface in x-space.

We first note that any point on the constraint surface the gradient $\nabla g(x)$ of the constraint function will be orthogonal to the surface.

For visualization,

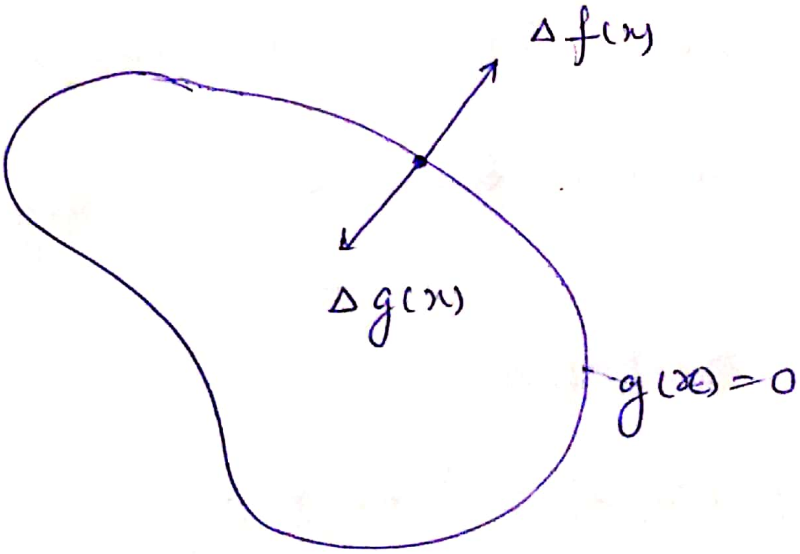
consider a point x that lies on the constraint surface, and consider a nearby point $x+\varepsilon$ that also lies on ~~that~~ the surface.

$$g(x+\varepsilon) \simeq g(x) + \varepsilon^T \nabla(g(x))$$

because both x and $x+\varepsilon$ lie on the constraint surface, we have $g(x) = g(x+\varepsilon)$ and $\varepsilon^T \nabla g(x) \simeq 0$.

In the limit $\|\varepsilon\| \to 0$ we have $\varepsilon^T \nabla g(x) = 0$, and because ε is then parallel to the constraint surface $g(x)=0$.

we see that ∇g is normal to the surface.



Next we seek a point x^* on the constraint surface such that $f(x)$ is maximized.

Such a point must have property that the vector $\nabla f(x)$ is also orthogonal to the constraint surface because otherwise it could increase the value of $f(x)$ by moving a short distance along the constraint surface.

Thus ∇f and ∇g are parallel (or anti parallel) vectors.

Hence $\nabla f + \lambda \nabla g = 0$

where $\lambda \neq 0$ is known as <u>Lagrange multiplier</u>.

Note - λ can have either sign +ve or -ve.

Now, we can write

$$L(x, \lambda) = f(x) + \lambda g(x)$$

The constrained stationarity condition is obtained by setting $\nabla_x L = 0$.

Furthermore, the condition $\partial L / \partial \lambda = 0$ leads to the constraint eq^n $g(x) = 0$.

Thus to find max^m of a function $f(x)$ subject to constraint $g(x) = 0$, we define lagrangian function and then we find the stationarity point of $L(x, \lambda)$ w.r.t x, λ. If we are only interested in x^*, then we can eliminate λ from the stationary equations without needing to find its value. (hence, the term 'undetermined multiplier').

Example -.

$$f(x_1, x_2) = 1 - x_1^2 - x_2^2 \quad \text{subject to}$$

$$g(x_1, x_2) = x_1 + x_2 - 1 = 0.$$

$$\therefore L(x, \lambda) = f(x) + \lambda g(x)$$

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

The conditions for this Lagrangian to be stationary w.r.t x_1, x_2 & λ

$$-2x_1 + \lambda = 0$$

$$-2x_2 + \lambda = 0$$

$$x_1 + x_2 - 1 = 0$$

Solutions to these eqn give stationary pt $(x_1^*, x_2^*) = (\frac{1}{2}, \frac{1}{2})$ and corresponding

$\lambda = 1$

Now, we will consider an inequality constraint $g(x) \geq 0$.

There are now two kinds of solution,

① constrained stationary point

$$g(x) > 0$$

in this case constraint is inactive

② lies on boundary

$$g(x) = 0$$

constraint is said to be active.

In the former case, the function $g(x)$ plays no role and so the stationary point of ~~Lagrange~~ or ~~function~~ condition is simply $\nabla f(x) = 0$.

. Here $\lambda = 0$.

In the later case.

$$\nabla f(x) = -\lambda \nabla g(x) \quad \text{for some value of } \lambda > 0.$$

because $f(x)$ will be max^{m} only when $g(x) > 0$

i.e oriented away from the region $g(x) >$.

For either of these two cases, the product $\lambda g(x) = 0$. Thus the sol^{n} to the problem of max^{m} $f(x)$ with subject to $g(x) \geq 0$

$$g(x) \geq 0$$
$$\lambda \geq 0$$
$$\lambda g(x) = 0$$

These are known as KKT conditions

Karush - Kuhn - Tucker.

Scanned with CamScanner

Note that if we wish to minimize (rather than maximize) $f(x)$ subjected to $g(x) \geqslant 0$, then we minimize

$$L(x, \lambda) = f(x) - \lambda (g(x))$$

with respect to x, again $\lambda \geqslant 0$.

—— x ——.

Generalization -

$$L(x, \{\lambda_j\}, \{\mu_k\}) = f(x) + \sum_{j=1}^{J} \lambda_j g_j(x) + \sum_{k=1}^{K} \mu_k h_k(x).$$