Lagrange Multipliers: Geometric Interpretation

The Core Idea

At an optimal point (stationary point) of a constrained optimization problem, the statement says:

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 + \cdots + \lambda_m \nabla g_m$$

Where:

- ∇f = gradient of the objective function (loss)
- ∇g_i = gradient of the i-th constraint
- λ_i = Lagrange multipliers (the coefficients)

What Does This Mean?

1. Geometric Interpretation

- The gradient of the loss function is **parallel to** (or lies in the span of) the constraint gradients
- You cannot improve the objective by moving along the constraint surface
- The gradient ∇f is "blocked" by the constraints

2. From Your Lecture Notes

The notes explain this beautifully:

> "At the minimum, the level curves are tangent to each other, so the normal vectors ∇f and ∇g are parallel."

And more generally:

> "Why the method works: at constrained min/max, moving in any direction along the constraint surface $g = c$ should give $df/ds = 0$. So, for any $\hat{u}$ tangent to $\{g = c\}$, $\frac{df}{ds}\big|_{\hat{u}} = \nabla f \cdot \hat{u} = 0$, i.e. $\hat{u} \perp \nabla f$. Therefore ∇f is normal to tangent plane to $g = c$, and so is ∇g, hence the gradient vectors are parallel."

Example from Your Notes

For the problem: minimize $f(x, y) = x^2 + y^2$ subject to $xy = 3$

At the optimum:

- $\nabla f = (2x, 2y)$
- $\nabla g = (y, x)$
- The condition becomes: $(2x, 2y) = \lambda(y, x)$

This gives:

$$2x = \lambda y$$
$$2y = \lambda x$$

The **Lagrange multiplier λ is the coefficient** that makes ∇f equal to a scalar multiple of ∇g.

Why "Linear Combination"?

When you have **multiple constraints** $g_1 = c_1, g_2 = c_2, \ldots, g_m = c_m$:

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 + \cdots + \lambda_m \nabla g_m$$

The gradient of the loss is a **weighted sum** (linear combination) of all constraint gradients, where the weights are the Lagrange multipliers.

Physical Intuition

Think of it this way:

- You want to minimize f, so naturally you'd move in the direction of $-\nabla f$
- But the constraints "push back" with forces proportional to ∇g_i
- At equilibrium (the optimal point), these forces balance: ∇f equals the combined effect of all constraint forces

The Lagrange multipliers λ_i tell you **how strongly** each constraint is "pushing" against your objective at the optimal point.