lagrange multipliers

- if we have a problem of the form $min/max_x\ f(x, y)$ such that $g_i(x, y) = 0$, where the g_i's just define a bunch of constraints on x, y, we can do the following.
- imagine the contours of f (the curves along which f is the same). so $f = 3$ is a contour.
- theorem: at the constrained min/max point of f, f and g are tangents to each other. that means their gradients at that point are proportional to each other.

  proof idea: f is “nice” and continuous with its derivative well behaved. lets say at the constrained min/max, they f wasn't tangent to g. that means we could move a bit along g and either increase/decrease f. which is a contradiction since we assumed we're at a point on g where f is max/min. QED.
- skipping over some basic math, this basically means that the lagrangian, defined as $\mathcal{L}(x, y, \lambda_i ... \lambda_j) = f(x, y) - \sum_i \lambda_i g_i(x, y)$ has gradient 0. we find all the solutions to $\nabla\mathcal{L} = \vec{0}$. plug in each solution into f and see which gives the min/max f. this is the constrained min/max.
- To see why the tangency theorem boils down to $\nabla\mathcal{L} = 0$, expand out $\nabla\mathcal{L}$ and also write out the conditions for parallel normals of f with each of g_i. You will see that they are both the same thing. if you're too lazy to do this math, this video does the manipulation to boil it down.
- If we have an inequality constraint of the form $x + 1 \leq 0$, we can convert it to an equality constraint by writing it as $x + 1 = s^2$ (since s is squared it has to be non-negative). Now we can use this just like we use any other equality constraint!
- Lagrangian gives way to solve constrained optimization problems with equality constraints. If there are also inequality conditions, we setup something called a generalized Lagrangian and if we solve the generalized Lagrangian problem unconstrained, its equivalent to solving the constrained optimization problem. This proof idea is discussed in my math notes of Deep Learning book but the problem formulation looks like this:

  $\mathcal{L} = f(\mathbf{x}) + \sum_i \lambda_i g_i(\mathbf{x}) + \sum_i \alpha_i h_i(\mathbf{x})$ and we want to find $min_{\mathbf{x}}\ max_{\lambda}\ max_{\alpha}\ \mathcal{L}$