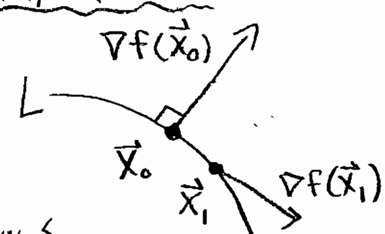
Notes on Lagrange Multipliers, a guide to Stewart 14.8 [Andrew Critch, Math 53, 09Su]

★ Suppose $f: \mathbb{R}^{n=2\text{ or }3} \to \mathbb{R}^1$, and L is a curve or surface in $\mathbb{R}^n$. Say we want to maximize (or minimize) f but only allowing points of L as inputs. I.e., we are wondering which points $\vec{x} \in L$ give the largest value of $f(\vec{x})$.

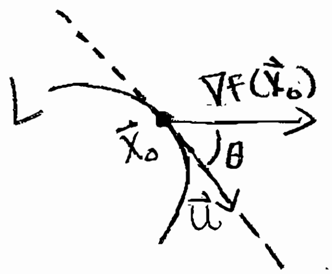
Suppose $\vec{x}_0 \in L$ is such a maximizing input. Then:

★ Theorem: $\boxed{\nabla f(\vec{x}_0) \perp L \text{ at } \vec{x}_0}$



(Note: This is not the theorem that says $\nabla f(x_0)$ is perpendicular to the level set of f at $\vec{x}_0$, because L is not a level set of f here! At most points $\vec{x}_1 \in L$, $\nabla f(\vec{x}_1)$ will not be $\perp$ to L.)

Why is this true? If L makes an angle of $\theta < \pi/2$ with $\nabla f(\vec{x}_0)$ at $\vec{x}_0$, then moving the input along L



instantaneously in the direction of the tangent vector $\vec{u}$ (see picture) will change f at a rate of

$$D_{\vec{u}} f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \vec{u}$$

$$= |\nabla f(\vec{x}_0)| \cdot |\vec{u}| \cdot \cos\theta$$

positive since $\theta < \pi/2$!

$$> 0.$$

This means $f(\vec{x}_0)$ can be increased slightly by moving $\vec{x}_0$ (inside L) in the (approximate) direction of $\vec{u}$, so it is not yet maximized.

Now, if L is defined by implicit equations, often called "constraints", then we can find $\vec{x}_0$ by a method, which depends on the number of constraints: for us, one or two.

(1) If the set L is defined by one implicit equation, say $\{g=c\}$ (so L is a surface in $\mathbb{R}^3$ or a curve in $\mathbb{R}^2$), then we already know that $\nabla g(\vec{x}_0)$ gives the normal direction to L at $\vec{x}_0$ (here we assume $\nabla g(\vec{x}_0) \neq 0$), so $\nabla f(\vec{x}_0) \parallel \nabla g(\vec{x}_0)$, i.e.

$$\boxed{\nabla f(\vec{x}_0) = \lambda \nabla g(x_0)}$$

writing out the components of this equation, and remembering also that $g(\vec{x}_0)=C$, we have $n+1$ equations in $n+1$ variables:

$$\left.\begin{array}{l} g=C \\ f_x=\lambda g_x \\ f_y=\lambda g_y \\ (f_z=\lambda g_z) \\ \vdots \end{array}\right\} \text{at } \vec{x}_0$$

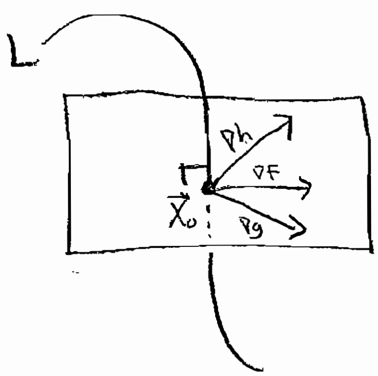
if in $\mathbb{R}^3$ ↘ $(f_z=\lambda g_z)$

<-- Please read the text for examples of how to solve these equations... they can be quite tricky!

The scalar λ is called a "Lagrange multiplier," and solving this system for $x_0, y_0 (, z_0)$ is called "the method of Lagrange multipliers."
The inputs $\vec{x}_0$ you find ("Lagrange" inputs?) are like critical inputs: if you check them all, you will find the "true" $\vec{x}_0 \in L$ which maximizes $f(\vec{x}_0)$.

(2) If L is defined by two implicit equations, $\left\{ \begin{matrix} g = c_1 \\ h = c_2 \end{matrix} \right\}$, for example a curve in $\mathbb{R}^3$, then we know two normal vectors to L at $\vec{x}_0$: $\nabla g(\vec{x}_0)$ and $\nabla h(\vec{x}_0)$.

(here we assume $\nabla g(\vec{x}_0)$ and $\nabla h(\vec{x}_0)$ are not 0 and not parallel to each other). Since ∇f, ∇g, and ∇h (at $\vec{x}_0$) all lie in the plane of normal vectors to L at $\vec{x}_0$, we can express ∇f as a "linear combination" of ∇g and ∇h there:



See last page for an optional explanation of this fact

$$\boxed{\underbrace{\nabla f = \lambda \cdot \nabla g + \mu \cdot \nabla h}_{\text{at } \vec{x}_0}}, \therefore \quad \left. \begin{matrix} g = c_1 \\ h = c_2 \\ f_x = \lambda g_x + \mu h_x \\ f_y = \lambda g_y + \mu g_y \\ f_z = \lambda g_z + \mu g_z \end{matrix} \right\} \text{at } \vec{x}_0$$

Now we have $n+2$ variables in $n+2$ unknowns we can solve to find "Lagrange" inputs, $\vec{x}_0$, and check them to find the maximum!

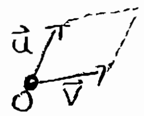
Of course, if we can parametrize the curve L nicely, we can solve a one-variable calculus problem instead of a system of 5 equations, which is probably much easier! But parametrization is often extremely difficult or impossible in practice, so then Lagrange multipliers come to the rescue.

//

How two vectors "span" a plane (optional)

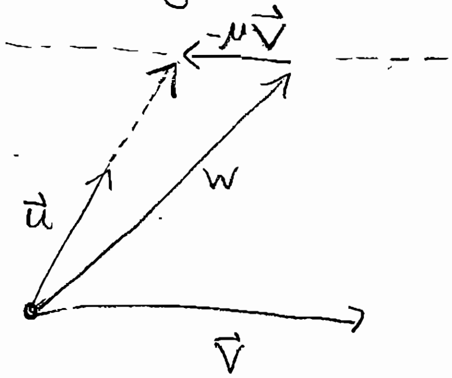
Say $\vec{u}$ and $\vec{v}$ are non-zero, non-parallel vectors in $\mathbb{R}^3$ (or $\mathbb{R}^n$)

Slide them into position at the origin, so they determine a plane:



Now imagine that this sheet of paper is that plane,

and imagine a third vector $\vec{w}$ in the same plane:



If we subtract ($\pm$) multiples $\mu\vec{v}$ of the vector $\vec{v}$ (μ being a scalar), the tip of the resulting vector $\vec{w}-\mu\vec{v}$ moves along the dotted line as μ changes. For some μ, $\vec{w}-\mu\vec{v}$ will line up with $\vec{u}$, say $\vec{w}-\mu\vec{v}=\lambda\vec{u}$ for some scalar λ. Rearranging gives:

$$\boxed{\vec{w}=\lambda\vec{u}+\mu\vec{v}}$$