Lagrange Multipliers

The method known as "Lagrange Multipliers" is an approach to the general problem of finding the maximum or minimum value of a function $\boldsymbol{g}: \mathbb{R}^{\mathbf{n}} \to \mathbb{R}$ when the variable is not allowed to range over all of $\mathbb{R}^{\mathbf{n}}$ but is constrained to lie in some subset. The method applies to subsets defined by the vanishing of a differentiable function $\boldsymbol{f}: \mathbb{R}^{\mathbf{n}} \to \mathbb{R}$. That is, the subset is of the form $\left\{\vec{x} \in \mathbb{R}^n : f(\vec{x}) = 0\right\}$.

The idea of the method is nicely illustrated by the case that $\boldsymbol{f}(x, y) = x^2 + y^2 - 1$, so the subset is the compact set $\mathcal{C} = \left\{(x, y) : x^2 + y^2 = 1\right\}$, and the function g has the form $\boldsymbol{g}(x, y) = Ax^2 + 2Bxy + Cy^2$. The function $\boldsymbol{g}$ of course attains its maximum and minimum on the circle $\mathcal{C}$ since $\mathcal{C}$ is compact and g is continuous. To locate the maximum and minimum points, we could note that $\mathcal{C}$ can be traced out as $\gamma(t) = (\cos t, \sin t)$, $t \in [0, 2\pi]$.

Then $\frac{d}{dt} g(\gamma(t)) = \operatorname{grad} g\big|_{(\cos t, \sin t)} \cdot \underbrace{(-\sin t, \cos t)}_{\gamma'(t)}$ by the Chain Rule. We want this to be 0, since the derivative is 0 at maxima and minima.

Now $(-\sin t, \cos t)$ is perpendicular to $(\cos t, \sin t) = \dfrac{\text{grad } f}{\|\text{grad } f\|} \neq \vec{0}$. [This is an aspect of the general idea that a "level curve" of a differentiable function $\mathbb{R}^2 \to \mathbb{R}$, is perpendicular to the gradient of the function.]

So $\text{grad}\, g\big|_{\gamma(t)}$ being perpendicular to $(-\sin t, \cos t)$ is the same as $\text{grad}\, g\big|_{\gamma(t)}$ being a multiple of $\text{grad } f\big|_{\gamma(t)}$. Thus, we should look for (x, y) with $x^2+y^2=1$ and with $\lambda(x, y) = (2Ax + 2By, 2Bx + 2Cy)$ for some λ. This is the same as saying that $\begin{pmatrix} A & B \\ B & C \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \dfrac{\lambda}{2}\begin{pmatrix} x \\ y \end{pmatrix}$

or (x, y) is an eigenvector of $\begin{pmatrix} A & B \\ B & C \end{pmatrix}$.

You can interpret this two ways: Since you know how to find eigenvectors for symmetric matrices, this shows how to solve the max/min problem--or at least to find candidates for the solutions. Looked at the other way around, the fact that g on $\{(x, y) : x^2 + y^2 = 1\}$ has a max and a min implies that eigenvectors exist! This example is very suggestive. The following theorem is what it suggests:

Theorem (Lagrange): If ***f***, ***g***: U: $\mathbb{R}^{\mathbf{n}} \to \mathbb{R}$, where U is open, are continuously differentiable functions , if ***f***(p)=0 but $\text{grad } f\big|_p \neq \vec{0}$, and if $g(p) = \sup\left\{g(\vec{x}) : \vec{x} \in U \;\&\; f(\vec{x}) = 0\right\}$ then ∃ λ ∈ ℝ ∍ $\text{grad}\, g\big|_p = \lambda \cdot \text{grad } f\big|_p$.

[The same holds if $g(p) = \inf\left\{g(\vec{x}) : \vec{x} \in U \;\&\; f(\vec{x}) = 0\right\}$ by applying the theorem to – f!][NB: Continuity of grad ***g*** is actually not needed!]

Proof: By translation and rotation of $\mathbb{R}^{\mathbf{n}}$ coordinates and by replacing ***f*** by u***f*** for a suitable u∈ℝ, u≠0, we can assume WOLOG that $\vec{p} = \vec{0}$ and $\text{grad } f\big|_p = (0,..,0,1)$. By the Implicit Function Theorem , there is a differentiable function ***F***: $\mathbb{R}^{\mathbf{n-1}} \to \mathbb{R}$ defined on a neighborhood of (0,0,...,0) $\in \mathbb{R}^{\mathbf{n-1}}$ ∍ $\boldsymbol{f}(x_1,x_2,\ldots,x_{n-1},\boldsymbol{F}(x_1,x_2,\ldots,x_{n-1}))=0$. Let $\gamma_k(t)=(0,\ldots,0,\boldsymbol{t},0,\ldots,0,\boldsymbol{F}(0,\ldots,0,\boldsymbol{t},0,\ldots,0)$, ***t*** in the k[th] slots. Then $\frac{d}{dt} f(\gamma_k(t)\big|_{t=0} = 0$ so $0 = \text{grad } f\big|_{\vec{0}} \cdot \gamma_k{}'(t)$. But $\text{grad } f\big|_{\vec{0}} \cdot \gamma_k{}'(t) =$ the n[th] component of $\gamma_k{}'(\boldsymbol{t}) = \left(0,\ldots,1,0,\ldots,0,\left.\frac{\partial F}{\partial x_k}\right|_{(0,\ldots,0)}\right)$.

Thus $\left.\frac{\partial F}{\partial x_k}\right|_{(0,\ldots,0)} = 0$, ∀ k=1,…,n-1 .

Next we compute $\left.\frac{d}{dt}g(\gamma_k(t))\right|_{t=0}$, k=1,…,n-1 . This must be 0 since $g(\vec{0})$ is a maximum for all points in $\{\vec{x}:\vec{x}\in U, f(\vec{x})=0\}$ and $f(\gamma_k(t))=0$, $\forall$ ***t***.

But $\frac{d}{dt}g(\gamma_k(t))\Big| = \gamma_k{}'(0)\cdot \text{grad}g\big|_{\vec{0}}$ = the k[th] component of $\text{grad}g\big|_{\vec{0}}$ by the facts that $\left.\frac{\partial F}{\partial x_k}\right|_{(0,\ldots,0)} = 0$ and that

$\gamma_k{}'(t)\big|_{t=0} = (0,\ldots,1,0,\ldots,\left.\frac{\partial F}{\partial x_k}\right|_{(0,\ldots,0)}) = (0,\ldots,1,0,\ldots0,0)$. Thus only the n[th] component of $\text{grad}\,\text{g}\big|_{\vec{0}}$ can possibly be nonzero so that $\text{grad}\,\text{g}\big|_{\vec{0}}$ is a multiple of $\text{grad}\, f\big|_{\vec{0}} = (0,\ldots,0,1)$ as required. □

This argument shows, as part of the proof, that the "level set" $\{\vec{x}: f(\vec{x})=0\}$ is, in a neighborhood of $\vec{0}$, actually a "graph" over the (x_1,…,x_{n-1},0) coordinate hyperplane. In $\mathbb{R}^3$, this would exhibit the level set as locally a "smooth surface" in the intuitive (and precise, too) sense. When n>3, the level set is a "smooth hypersurface". This all depends on $\text{grad}\, f\big|_p \neq \vec{0}$.

The intuition here is that if $\text{grad}\, g|_{\vec{0}}$ were not a multiple of $\text{grad}\, f|_{\vec{0}}$ then one could move in $\{\vec{x} : f(\vec{x}) = 0\}$ along the "projection" of $\text{grad}\, g|_{\vec{0}}$ on the $(x_1,\ldots,x_{n-1})$ subspace and in that direction, ***g*** would have nonzero derivative, contradicting that ***g*** has a maximum on $\{\vec{x} : f(\vec{x}) = 0\}$ at $\vec{0}$. But the Implicit Function Theorem argument is crucial: otherwise one does not know that movement is possible in that direction while still remaining on the level "surface" $\{\vec{x} = (x_1,\ldots,x_n) : f(\vec{x}) = 0\}$. If n=2 and $f(x, y) = x^2+y^2$, for example, then $\{\vec{x} : f(\vec{x}) = 0\}$ contains only one point! No movement is possible.

Exercise: Given a symmetric matrix $A=(a_{ij})$ apply Lagrange multipliers to $f(\vec{x}) = \|\vec{x}\|^2 - 1$ and $g(\vec{x}) = \sum a_{ij} x_i x_j = \sum_i a_{ii} x_i^2 + 2\sum_{i<j} a_{ij} x_i x_j$ to prove the existence of an eigenvector of A (namely, $\vec{x} \ni f(\vec{x}) = 0$ and $g(\overrightarrow{x_0}) = \max \text{of } g \text{ on} \{\vec{x} : f(\vec{x}) = 0\}$ is an eigenvector).