A proof of the method of Lagrange Multipliers.

The technique of Lagrange multipliers allows you to maximize / minimize a function, subject to an implicit constraint. Let $f\colon \mathbb{R}^d \to \mathbb{R}^n$ be a C^1 function, $C \in \mathbb{R}^n$ and $M = \{f = C\} \subseteq \mathbb{R}^d$. (We will always assume that for all $x \in M$, $\operatorname{rank}(Df_x) = n$, and so M is a $d - n$ dimensional manifold.) Now suppose you are given a function $h\colon \mathbb{R}^d \to \mathbb{R}$, and want to find the local extrema of h on M. That is, you want to minimize or maximize h subject to the constraint $f = C$.

Suppose h attains a constrained local extremum at a, subject to the constraint $f = C$. Sketching level sets of h one guesses that if f attains a constrained extremum at $a \in M$, then the manifolds M and $\{h = h(a)\}$ are both tangent at a. More precisely, if $\Gamma = \{h = h(a)\}$ is a manifold, then the geometric intuition suggests $TM_a \subseteq T\Gamma_a$. We know that $T\Gamma_a = \ker Df_a$ and $TM_a = \ker Dh_a$, and one can easily check that

$$TM_a \subseteq T\Gamma_a \iff \nabla h(a) \in \operatorname{span}\{\nabla f_1(a), \dots, \nabla f_n(a)\}\,.$$

That is $TM_a \subseteq T\Gamma_a$ if and only if there exists $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ such that

$$\nabla h(a) = \sum_{i=1}^{n} \lambda_i \nabla f_i(a)\,. \tag{1}$$

This is the method of *Lagrange multipliers*, and we will prove that it works shortly.

Theorem 1 (Lagrange multipliers)**.** *Let $h\colon \mathbb{R}^d \to \mathbb{R}$, $f\colon \mathbb{R}^d \to \mathbb{R}^n$ be C^1 functions, $C \in \mathbb{R}^n$ and $M = \{f = C\} \subseteq \mathbb{R}^d$. Assume*[1] *that for all $x \in M$, $\operatorname{rank}(Df_x) = n$. If h attains a constrained local extremum at a, subject to the constraint $f = C$, then there exists $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ such that (1) holds.*

Before proving the theorem, we make a few remarks. First, in order to find a constrained extremum, the above theorem says (1) holds for some $\lambda_1, \dots \lambda_n \in \mathbb{R}$. Second, since $a \in M$, we know $f(a) = 0$. Thus in order to practically find constrained maxima / minima, we simultaneously solve the equations

$$\nabla h(a) = \sum_{1}^{n} \lambda_i \nabla f_i(a) \qquad \text{and} \qquad f(a) = C\,, \tag{2}$$

where both $a \in \mathbb{R}^d$ and $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ are unknown.

This is a system of equations with $d + n$ unknowns (d coordinates of a, and each of the λ_i's). Equating each coordinate of both sides of the first equation in (2), we get d equations. Equating each coordinate of both sides in the second equation in (2) we get another n equations. Thus the system (2) is a system of $d + n$ equations with $d + n$ unknowns. Typically you will only have finitely many solutions to this system, and thus only finitely many candidates at which constrained local extrema can occur.

To determine whether each of these points is a local maximum or minimum there is a test involving the *bordered Hessian*. This is, however, quite complicated and is usually more trouble than it is worth, so one usually uses some ad-hoc method to decide whether each of the solutions above is a local maximum or not.

Example 2. Find necessary conditions for $h(x, y) = y$ to attain a local maxima/minima of subject to the constraint $y = f(x)$.

Of course, from one variable calculus, we know that the local maxima / minima must occur at points where $f' = 0$. Let's revisit it using the constrained optimization technique above.

Solution. Note our constraint is of the form $y - f(x) = 0$. So at a local maximum we must have

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \nabla h = \lambda \nabla (y - f(x)) = \lambda \begin{pmatrix} -f'(x) \\ 1 \end{pmatrix} \quad \text{and} \quad y = f(x)\,.$$

This forces $\lambda = 1$ and hence $g'(x) = 0$, as expected. □

Example 3. Maximise xy subject to the constraint $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

[1] From the proof it is clear that one only needs to assume $\operatorname{rank}(Df_a) = n$ at the constrained extremum, and not at all points in M.

Solution. At a local maximum,

$$\begin{pmatrix} y \\ x \end{pmatrix} = \nabla(xy) = \lambda \nabla\Big(\frac{x^2}{a^2} + \frac{y^2}{b^2}\Big) = \lambda \begin{pmatrix} 2x/a^2 \\ 2y/b^2. \end{pmatrix}$$

which forces $y^2 = x^2 b^2/a^2$. Substituting this in the constraint gives $x = \pm a/\sqrt{2}$ and $y = \pm b/\sqrt{2}$. This gives four possibilities for xy to attain a maximum. Directly checking shows that the points $(a/\sqrt{2}, b/\sqrt{2})$ and $(-a/\sqrt{2}, -b/\sqrt{2})$ both correspond to a local maximum, and the maximum value is $ab/2$. □

Before proving Theorem 1 we give an incorrect proof, that is found in many places.

Wrong proof of Theorem 1. Let $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ and define the function $H\colon \mathbb{R}^{d+n} \to \mathbb{R}$ by

$$H(x, \lambda) = h(x) - \lambda \cdot (f(x) - C)\,.$$

Now extremize H. We know that at local extrema of H we must have $DH = 0$. Equating $\partial_{x_i} H = 0$ for all i gives the first equation in (2). Equating $\partial_{\lambda_i} H = 0$ gives the second equation in 2. This shows that any local extremum of H must be a local constrained extremum of h given the constraint $f = C$. □

The error in the above proof is that it does not rule out the possibility of the existence of constrained local extrema of h (given the constraint $f = C$), that *are not* local extrema of H. A correct proof requires the implicit function theorem.

Correct proof of Theorem 1. Suppose h attains a constrained local extremum at a. Reordering the coordinates if necessary, we assume without loss of generality that the last n columns of Df_a are linearly independent. Let $m = d - n$ and denote points in $\mathbb{R}^d$ by (x, y) with $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. By the implicit function theorem we know that there exists open sets $U \ni a$, $V \subseteq \mathbb{R}^m$ and a C^1 function $g\colon V \to \mathbb{R}^d$ such that

$$M \cap U = \{(x, g(x)) \mid x \in V\}\,.$$

Now write $a = (b, c)$ with $b \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ and observe that if (b, c) is a constrained local extremum of h given $f = C$, then b must be an *unconstrained* local extremum of the function

$$H(x) \stackrel{\text{def}}{=} h(x, g(x))\,,$$

and hence $DH_b = 0$.

Now write $Dh = (\mathcal{D}_x h\ \mathcal{D}_y)$, where $\mathcal{D}_x h$ and $\mathcal{D}_y h$ are the sub-matrices of Dh obtained by taking the first m and n columns respectively. Since $DH_a = 0$, the chain rule implies

$$0 = DH_a = \begin{pmatrix} \mathcal{D}_x h_a & \mathcal{D}_y h_a \end{pmatrix} \begin{pmatrix} I \\ Dg_b \end{pmatrix} = \mathcal{D}_x h_a + \mathcal{D}_y h_a Dg_b\,. \tag{3}$$

We can compute Dg_a by differentiating $f(x, g(x)) = C$ implicitly. This gives

$$0 = \begin{pmatrix} \mathcal{D}_x f_a & \mathcal{D}_y f_a \end{pmatrix} \begin{pmatrix} I \\ Dg_b \end{pmatrix} = \mathcal{D}_x f_a + \mathcal{D}_y f_a Dg_b \implies Dg_b = -(\mathcal{D}_y f_a)^{-1} \mathcal{D}_x f_a\,.$$

Substituting this in (3) gives

$$\mathcal{D}_x h_a = \mathcal{D}_y h_a \Big((\mathcal{D}_y f_a)^{-1} \mathcal{D}_x f_a \Big) = \Big(\mathcal{D}_y h_a (\mathcal{D}_y f_a)^{-1} \Big) \mathcal{D}_x f_a = \Lambda \mathcal{D}_x f_a\,, \tag{4}$$

where Λ is the $1 \times n$ matrix defined by

$$\Lambda = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_n \end{pmatrix} \stackrel{\text{def}}{=} \mathcal{D}_y h_a (\mathcal{D}_y f_a)^{-1}$$

Also note

$$\mathcal{D}_y h_a = \mathcal{D}_y h_a \Big((\mathcal{D}_y f_a)^{-1} \mathcal{D}_y f_a \Big) = \Big(\mathcal{D}_y h_a (\mathcal{D}_y f_a)^{-1} \Big) \mathcal{D}_y f_a = \Lambda \mathcal{D}_y f_a\,.$$

Combined with (4) this shows

$$Dh_a = \begin{pmatrix} \mathcal{D}_x h_a & \mathcal{D}_y h_a \end{pmatrix} = \Lambda \begin{pmatrix} \mathcal{D}_x f_a & \mathcal{D}_y f_a \end{pmatrix} = \Lambda Df_a\,.$$

Transposing this gives (1), finishing the proof. □