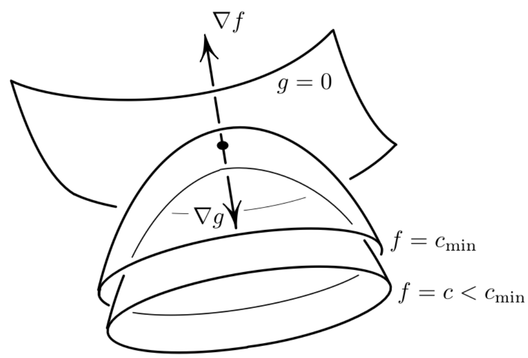
MAT 307, Multivariable Calculus with Linear Algebra – Fall 2024

1 Lagrange Multipliers

We give the geometric picture of Lagrange Multipliers, in the simplest setting.



We seek to minimize $f : \mathbb{R}^n \to \mathbb{R}$ subject to the constraint $g : \mathbb{R}^n \to \mathbb{R}$, where $x \in \mathbb{R}^n$. From the figure, we can see that if x_* is a minimum of f subject to $g = 0$, then the level sets of f and g passing through x_* are tangent at x_*. Thus ∇f and ∇g are parallel at x_*, e.g. there is a $\lambda \in \mathbb{R}^n$ so that $\nabla(f - \lambda g) = 0$ at x_*.

Let use argue this more careful. Consider *any* curve $\vec{r}(s)$ lying on the surface $g = 0$, with $\vec{r}(0) = x_*$. Differentiating $g(\vec{r}(s)) = 0$ in s, we find

$$0 = \frac{d}{ds} g(\vec{r}(s))\Big|_{s=0} = \dot{\vec{r}}(0) \cdot \nabla g(x_*)$$

Here, $\dot{\vec{r}}(0)$ can be *any* vector which is tangent to the surface $g = 0$ at the point x_*. As such, it shows that $\nabla g(x_*)$ is normal to the surface $g = 0$ at x_*. On the other hand, since x_* is a critical point, we have also that

$$0 = \frac{d}{ds} f(\vec{r}(s))\Big|_{s=0} = \dot{\vec{r}}(0) \cdot \nabla f(x_*).$$

This shows that $\nabla f(x_*)$ is likewise normal to the surface $g = 0$ at x_*. Since there is a unique normal direction to a surface of codimension 1[1], this shows that $\nabla f(x_*) = \lambda \nabla g(x_*)$.

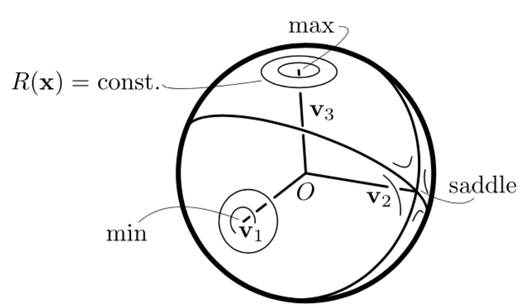
Physical Interpretation: Lagrange's theorem that $\nabla f(x_*) = \lambda \nabla g(x_*)$ is the statement that a particle in a force field constrained frictionlessly to a surface

[1]We require the simple result: let $\vec{a}, \vec{b}, \vec{v}$ be vectors in $\mathbb{R}^n$ with $\vec{b} \neq 0$ and are such that if $\vec{b} \cdot \vec{v} = 0$ then $\vec{a} \cdot \vec{v} = 0$. Then $a = \lambda b$ for some $\lambda \in \mathbb{R}$. Proof: consider a decomposition of $\vec{a}$ as $\vec{a} = \lambda b + r$ with $\lambda \in \mathbb{R}$ and $\vec{r} \cdot \vec{b} = 0$. Since $\vec{b} \cdot \vec{r} = 0$, we have $\vec{a} \cdot \vec{r} = 0$ by hypothesis. Taking the dot product with $\vec{r}$ shows $\vec{a} \cdot \vec{r} = \|\vec{r}\|^2$. Thus $\vec{r} = 0$.

is in equilibrium if and only if the force field is perpendicular to the surface. Indeed, interpret f as the potential energy of a particle in a force field $\vec{F} = -\nabla f$ in $\mathbb{R}^n$. The particle is constrained to the surface $g = 0$. If the potential energy f restricted to $g = 0$ is minimal at x_*, then x_* is an equilibrium, i.e., the field force $\vec{F}$ is balanced by reaction force $\vec{R}$ of the constraint. And since $\vec{R}$ is normal to the surface, we have $\vec{R} = \lambda \nabla g$ for some $\lambda \in \mathbb{R}$.

We begin with an application to linear algebra [2]. We have the

Theorem 1. *Let M be a real symmetric $n \times n$ matrix. Critical points of the quadratic form $f(\vec{x}) = (A\vec{x}, \vec{x})$ restricted to the unit sphere $|\vec{x}|^2 = 1$ are eigenvectors of M. Moreover, the critical values are the eigenvalues.*



Proof. We use Lagrange multipliers, where $f(\vec{x}) = (A\vec{x}, \vec{x})$ and $g(\vec{x}) = \|\vec{x}\|^2$. Note that $\nabla f(\vec{x}) = A\vec{x}$ and $\nabla g(\vec{x}) = \vec{x}$. We conclude that $\vec{v}$ is a critical point of $f(\vec{x})$ subject to the constraint $g = 1$ if and only if, for some $\lambda \in \mathbb{R}$ we have

$$A\vec{v} = \lambda \vec{v}.$$

Moreover, the critical value is $(A\vec{v}, \vec{v}) = (\lambda \vec{v}, \vec{v}) = \lambda$. ☐

As such, we have the formulae

$$\lambda_{\min} = \min_{\vec{y} \neq 0} \frac{(A\vec{y}, \vec{y})}{(\vec{y}, \vec{y})}, \qquad \lambda_{\max} = \max_{\vec{y} \neq 0} \frac{(A\vec{y}, \vec{y})}{(\vec{y}, \vec{y})}.$$

This goes by the name Rayleigh Ritz theorem, and $\frac{(A\vec{y}, \vec{y})}{(\vec{y}, \vec{y})}$ the Rayleigh quotient. The next largest eigenvalue can be obtained by similar formulae, if the maximum is taken among all vectors orthogonal to the eigenvector corresponding to the maximum eigenvalue. This continues, replacing the single eigenvector with the span of the eigenvectors corresponding to of the bigger eigenvalues. There is an

[2] This is technically outside the scope of the class. I include it because it is a nice application of Lagrange multipliers to an important theorem. This proof (and the pictures) is due to Mark Levi and can be found in his book *Classical Mechanics with Calculus of Variations and Optimal Control*.

improvement of this procedure that does not require knowledge of the higher eigenspaces called the Courant-Fischer Minimax Theorem.

When we have constraint equations and constraint inequalities such as

- Find min of $f(x,y) = x^2 - 2xy$ subject to the constraint inequalities $x \geq 1, y \geq 1$: this is standard min/max problem with boundaries and infinity to check.
- Find min of $f(x,y,z) = x^2 + y^2 + z^2$ subject to the constraint equation $3x+5y+z = 9$: can eliminate z to find min of $F(x,y) = x^2+y^2+(9-3x-5y)^2$ with no constraint equation.
- Find min of $f(x,y) = x^2 + y^2$ subject to the constraint equation $e^{x+y} = xy + 2$: can't eliminate a variable, so must use Lagrange multipliers.

Example 1. Find the min and max of $f(x,y) = (x-1)^2 + y^2$ on the curve $g(x,y) = x^3 - y^2 = 0$ where $x \geq 0$. Set $\nabla f = \lambda \nabla g$, we have

$$2x - 2 = 3\lambda x^2, \qquad 2y = -2\lambda y.$$

Either $y = 0$, so $x^3 = y^2 = 0$, but $2x - 2 = -2 \neq 0$. So $\lambda = -1$, and local (global min) of f at $x = \frac{-1+\sqrt{7}}{3}$, $y = \pm\sqrt{\left(\frac{-1+\sqrt{7}}{3}\right)^3}$. Note that $\nabla g(0,0) = \vec{0}$, and f attains a local max at $(0,0)$ from geometric picture. But the Lagrange multiplier method does not find this local extrema. The global max of f doesn't exist.

Caveat: In Lagrange multiplier method, we need that $\nabla g \neq \vec{0}$. Technically, $\nabla g(0,0)$ is parallel to $\nabla f(0,0)$ but no λ exists such that $\nabla f = \lambda \nabla g$ at $(0,0)$. When the two gradients are parallel, we can also have $\lambda \nabla f = \nabla g$, with $\lambda = 0$. So when the constraint g is not a smooth curve, one should check the point where $\nabla g = \vec{0}$.

The value the Lagrange multiplier λ attains at an optimum solution of f subject to the constraint $g = c$ is the rate of change of f with respect to the constraint variable c:

$$\frac{df}{dc} = \lambda$$

Since at the optimal solution $\vec{x}(c)$, $g(\vec{x}(c)) = c$,

$$\frac{dg}{dc} = \nabla g(\vec{x}(c)) \cdot \vec{x}'(c) = 1, \qquad \frac{df}{dc} = \nabla f(\vec{x}(c)) \cdot \vec{x}'(c) = \lambda \nabla g(\vec{x}(c)) \cdot \vec{x}'(c) = \lambda.$$

In economics, λ is known as the shadow price for the constraint - it represents the value of the resource (constraint).

To optimize f subject to multiple constraints $g_1 = c_1, \ldots, g_k = c_k$, let S be the $(n-k)$-dimensional intersection of the k hypersurfaces S_i, level surface of $g_i = c_i$

for $i = 1, \cdots, k$. Suppose that ∇g_i's are linearly independent everywhere on S. Any vector tangent to S must be tangent to each S_i, and hence perpendicular to each ∇g_i. Each ∇g_i is perpendicular to the intersection S. If f attains an extremum at $P \in S$, ∇f must be perpendicular to all vectors tangent to S at P. So ∇f is in the k-dimensional space spanned by the normal vectors to S_i's at P. It follows from linear algebra that

$$\nabla f = \lambda_1 \nabla g_1 + \cdots + \lambda_k \nabla g_k$$

for some constants $\lambda_, \ldots, \lambda_k$.

Remark. Vector $\vec{v}_1, \cdots, \vec{v}_k$ are linearly independent if $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_k\vec{v}_k = \vec{0}$ if and only if $a_1 = a_2 = \cdots = a_k = 0$.

Example 2. Find the closest point to the origin which belongs to the cone

$$x^2 + y^2 = z^2$$

and to the plane

$$x + y + z = 2.$$

We need to minimize $f(x, y, z) = x^2+y^2+z^2$ subject to $g_1(x, y, z) = x^2+y^2-z^2 = 0$ and $g_2(x, y, z) = x + y + z - 2 = 0$.

We introduce λ_1 and λ_2, where

$$\nabla f(x, y, z) = \lambda_1 \nabla g_1(x, y, z) + \lambda_2 \nabla g_2(x, y, z).$$

$$2x = 2\lambda_1 x + \lambda_2, 2y = 2\lambda_1 y + \lambda_2, 2z = -2\lambda_1 z + \lambda_2, x^2 + y^2 = z^2, x + y + z = 2.$$

From the first two coordinates, we get $(x - y) = \lambda_1(x - y)$.
If $x \neq y$ then $\lambda_1 = 1$ and $\lambda_2 = 0$; from which we get $z = 0$ and $x = y = 0$, a contradiction.
If $x = y$, we get $z = \pm\sqrt{2}x$.

$$(2 \pm \sqrt{2})x = 2.$$

The critical points are

$$P_1 = (2 - \sqrt{2}, 2 - \sqrt{2}, \sqrt{2}(2 - \sqrt{2})), \qquad P_2 = (2 + \sqrt{2}, 2 + \sqrt{2}, -\sqrt{2}(2 + \sqrt{2})).$$

The geometry of the surfaces show that P_1 and P_2 are both local min. P_1 is closer to the origin than P_2.

To show that P_1 is closest to the origin on the whole set

$$D = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 = z^2, x + y + z = 2\}.$$

Let

$$K = \{(x, y, z) \in F \mid x^2 + y^2 + z^2 \leq 25\},$$

K is compact, so f attains min somewhere in K, and so it must attain the min at the point P_1. Since $f \geq 25$ on outside of K on D, P_1 is the global min for f on F, $f(P_1) = 4(2-\sqrt{2})^2$.

D is unbounded, so f has no global max. Also note that $z = 2 - x - y$, $x^2 + y^2 = (2 - x - y)^2$,

$$2(x+y) = 2 + xy, \qquad y = \frac{2(x-1)}{x-2}, \qquad z = 2 - x - y$$

As $x \to 2$, $y \to \pm\infty$, $z \to \mp\infty$.

Alternatively, note that on the cone $z = \pm r$, and on the plane $r(\cos\theta + \sin\theta \pm 1) = 2$. To optimize $x^2 + y^2 + z^2 = r^2 + z^2 = 2r^2$, we just need to optimize $r = \frac{2}{\cos\theta + \sin\theta \pm 1}$. Taking $\frac{dr}{d\theta}$ we have $\tan\theta = 1$, so $\theta = \pi/4, 5\pi/4$, where $x = y$.