ECEn 671: Mathematics of Signals and Systems

Randal W. Beard

Brigham Young University

September 1, 2023

Section 1

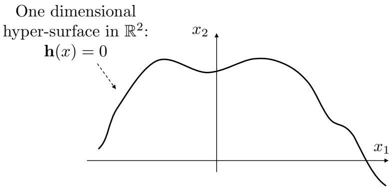
Equality Constraints: Lagrange Multipliers

Equality Constraints: Lagrange Multipliers

Several geometric insights help:

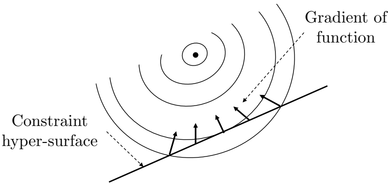
Insight #1

- Geometrically what do the constraints $\mathbf{h}(x) = 0$ look like?
- What if $\mathbf{h}$ is linear, i.e. $\mathbf{h}(x) = Hx = 0$ where $H : \mathbb{R}^n \to \mathbb{R}^m$.
- The constraint implies that x must be in the null space of H, which is a linear space of dimension $n - m$, i.e. an $n - m$ dimensional hyperplane.
- In general, $\mathbf{h}(x) = 0$ is an $n - m$ dimensional hypersurface in $\mathbb{R}^n$.



Equality Constraints: Lagrange Multipliers

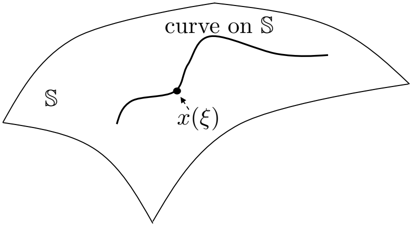
Insight #2



At a constrained minimum the gradient is orthogonal to the hypersurface.

Equality Constraints: Lagrange Multipliers

To formalize, we need some definitions.
Let $\mathbb{S}$ be a hyper-surface of dimension $n - m$. Let x be a curve on $\mathbb{S}$ continuously parameterized by $\xi \in [a, b]$, i.e. $x(\xi) \in \mathbb{S}$.



The derivative of the curve at $x(\xi_0)$ is

$$\dot{x}(\xi_0) = \frac{d}{d\xi} x(\xi_0).$$

Equality Constraints: Lagrange Multipliers

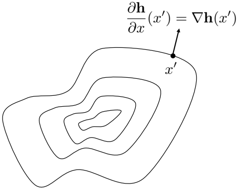
Definition

The <u>tangent plane</u> to a surface $\mathbb{S}$ at $x \in \mathbb{S}$ is the span of the derivatives of all the differentiable curves on $\mathbb{S}$ at x.

The problem with this definition is that it is not constructive, i.e. it doesn't give us a good way to actually construct the tangent plane.

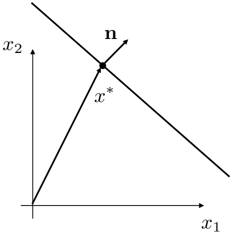
Equality Constraints: Lagrange Multipliers

To construct the tangent plane, recall that the gradient of $\mathbf{h}(x)$ is orthogonal to level curves of $\mathbf{h}(x)$:



Also recall that the formula for the plane is

$$\left\{ y \in \mathbb{R}^n \mid \mathbf{n}^\top (y - x^*) = 0 \right\}.$$



Equality Constraints: Lagrange Multipliers

Therefore the tangent plane of $h(x)$ at x^* is given by

$$P = \{y \in \mathbb{R}^n \mid \nabla \mathbf{h}^\top(x^*)(y - x^*) = 0\}$$

where $\nabla \mathbf{h}^\top(x^*)$ defines an $n - m$ dimensional plane if the rows are linearly independent.

Definition

When the gradient vectors $\nabla h_1, \nabla h_2, \ldots, \nabla h_m$ are linearly independent at x^*, x^* is called a regular point.

We will always assume "regularity" of the constraints.

Equality Constraints: Lagrange Multipliers

Lemma (Moon Lemma 18.1)

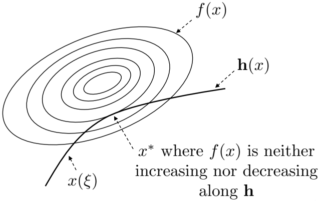
Let $x(\xi)$ be a curve on $\mathbf{h}(x) = 0$ such that

$$x(\xi)|_{\xi=0} = x^*,$$

is a constrained local minimum of f. Then

$$\left.\frac{d}{d\xi} f(x(\xi))\right|_{\xi=0} = 0.$$

Geometry: at point p, f is neither increasing nor decreasing along $x(\xi)$.



Equality Constraints: Lagrange Multipliers: Proof

Proof.

Expanding $f(x(\xi))$ in a Taylor series:

$$f(x(\xi)) = f(x(0)) + \xi \frac{d}{d\xi} f(x(\xi))\bigg|_{\xi=0} + O(|\xi|)$$

If $f(x(0))$ is a local minimum then for $|\xi|$-small

$$\xi \frac{d}{d\xi} f(x(\xi))\bigg|_{\xi=0} \geq 0$$

for all ξ both positive and negative. Therefore

$$\frac{d}{df} f(x(\xi))\bigg|_{\xi=0} = 0$$

where $\frac{d}{d\xi} f(x(\xi)) = \nabla^\top f(x(\xi)) \dot{x}(\xi)$ and $\dot{x}(\xi)$ is an element of the tangent plane. □

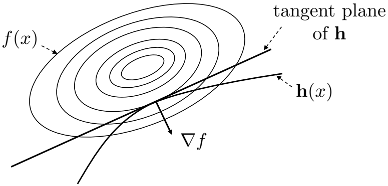
Equality Constraints: Lagrange Multipliers

Lemma (Moon Lemma 18.2)

If x^* is a regular point of $\mathbf{h}(x) = 0$ and a local constrained minimum, then

$$\nabla \mathbf{h}^\top(x^*)y = 0 \Rightarrow \nabla f^\top(x^*)y = 0$$

i.e. if y is in the tangent plane of $\mathbf{h}$ at x^*, then y is orthogonal to the gradient of f.



Proof of Lemma 18.2

Proof.

Translate the coordinate system such that $x^* = 0$. Regularity implies that the tangent plane is given by

$$P(x^*) \triangleq \{z \in \mathbb{R}^n \mid \nabla \mathbf{h}^\top (x^*) z = 0\}$$

Let $y \in P(x^*)$ then

$$\nabla \mathbf{h}^\top (x^*) y = 0.$$

Now choose a smooth curve $x(\xi)$ on $\mathbf{h}(x) = 0$ such that $x(0) = x^*$ and $\dot{x}(0) = y$.

From Lemma 18.1, $\nabla f^\top (x^*) y = 0$. $\square$

Key Insight

At a constrained local minimum $\nabla f(x^*)$ and the columns of $\nabla \mathbf{h}(x^*)$ are parallel, i.e., there is some scalar μ_i such that

$$\begin{aligned}
& \nabla f(x^*) = \mu_i \nabla h_i(x^*) \qquad i = 1, \ldots, m \\
\implies & m \nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla h_i(x^*) \\
\implies & \nabla f(x^*) - \sum_{i=1}^{m} \frac{\mu_i}{m} \nabla h_i(x^*) = 0 \\
\implies & \boxed{\nabla f(x^*) + \nabla \mathbf{h}(x^*) \lambda = 0}
\end{aligned}$$

where

$$\lambda = \begin{pmatrix} -\frac{\mu_1}{m} & \ldots & -\frac{\mu_m}{m} \end{pmatrix}^\top .$$

The vector $\lambda \in \mathbb{R}^m$ is called the Lagrange Multiplier.

Necessary Conditions

Theorem (Moon Theorem 18.3 (Necessary conditions for equality constraints))

Let x^* be a local extremum of f subject to the constraints $h(x) = 0$, and let x^* be a regular point. Then there is a $\lambda \in \mathbb{R}^n$ such that

$$\nabla f(x^*) + \nabla \mathbf{h}(x^*)\lambda = 0.$$

Corollary

Let

$$L(x, \lambda) = f(x) + \mathbf{h}^\top(x)\lambda.$$

Then if x^* is a regular local extremum, then

$$\nabla_x L(x^*, \lambda^*) = \frac{\partial L}{\partial x}(x^*, \lambda^*) = 0$$

and

$$\nabla_\lambda L(x^*, \lambda^*) = \frac{\partial L}{\partial \lambda}(x^*, \lambda^*) = 0.$$

Proof of Theorem 18.3

Proof.

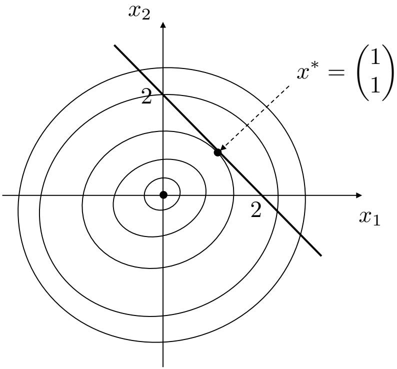
$$\frac{\partial L}{\partial x} = \nabla f(x^*) + \nabla h(x^*)\lambda^* = 0 \tag{1}$$

$$\frac{\partial L}{\partial \lambda} = h(x^*) = 0 \tag{2}$$

Equation (2) implies that the constraints are satisfied.
Equation (1) comes from Theorem 18.3. □

Lagrange Multipliers: Example 1

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 + x_2 = 2 \end{aligned}$$



The Lagrangian is

$$L = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 2).$$

Lagrange Multipliers: Example 1, cont.

The necessary conditions for a minimum are

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= 2x_1 + \lambda = 0 \quad &\implies x_1 = -\frac{\lambda}{2} \\
\frac{\partial L}{\partial x_2} &= 2x_2 + \lambda = 0 \quad &\implies x_2 = -\frac{\lambda}{2} \\
\frac{\partial L}{\partial \lambda} &= x_1 + x_2 - 2 = 0 \quad &\implies -\lambda - 2 = 0.
\end{aligned}$$

Therefore $\lambda = -2$, $x_1 = 1$, $x_2 = 1$, which implies that

$$x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

as expected.

Lagrange Multipliers: Example 18.4.7 (Maximum Entropy)

Let $\mathbb{X}$ be a random variable with probability mass function
$p(\mathbb{X} = x_i) = p_i \qquad i = 1, \ldots, m,$
Then, the entropy of $\mathbb{X}$ is defined as

$$H = -\sum_{i=1}^{m} p_i \log p_i$$

Question: Which pmf has maximum entropy?

To answer, lets solve the optimization problem:

$$\begin{aligned} \max \quad & H \\ \text{s.t.} \quad & \sum p_i = 1, \\ & p_i \geq 0 \end{aligned}$$

Lagrange Multipliers: Example 18.4.7 (Maximum Entropy)

Lets ignore the inequality constraint for now and go back later.
The Lagrangian is

$$L = -\sum p_i \log p_i + \lambda\left(\sum p_i - 1\right)$$

The necessary conditions are

$$\frac{\partial L}{\partial p_i} = 0 \qquad i = 1, \ldots, m$$
$$\frac{\partial L}{\partial \lambda} = 0$$

where

$$\frac{\partial L}{\partial p_i} = -\log p_i - 1 + \lambda = 0 \qquad i = 1, \ldots, m$$
$$\implies \lambda = 1 + \log p_i$$
$$\implies \log p_i = \lambda - 1.$$

Lagrange Multipliers: Example 18.4.7 (Maximum Entropy)

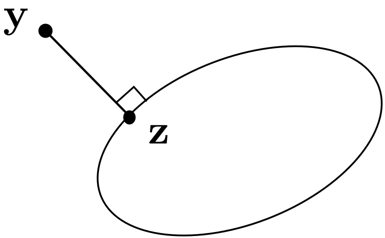
So p_i must be constant for all i. The constant $\sum p_i = 1 \Rightarrow p_i = \frac{1}{n} \geq 0$, so p_i satisfies the inequality constraint.

In other words: the uniform pmf maximizes entropy.

In other words: all possibilities are equally likely.

Lagrange Multipliers: Example: Constrained Least Squares

Given an ellipsoid and a point y outside the ellipsoid, find the point z in the ellipsoid nearest to y.



The equation for an ellipsoid is given by

$$E = \{z \in \mathbb{R}^n : z^\top L^\top L z \leq 1.$$

Lagrange Multipliers: Example: Constrained Least Squares

So we need to solve the following constrained optimization problem:

$$\begin{array}{cc} \min\limits_{z \in \mathbb{R}^n} & \|z - y\| \\ \text{s.t.} & z^\top L^\top L z = 1 \end{array}$$

The Lagrangian is

$$L = (y - z)^\top (y - z) + \lambda (z^\top L^\top L z - 1)$$

Lagrange Multipliers: Example: Constrained Least Squares

The necessary conditions are

$$\frac{\partial L}{\partial z} = -2(y - z) + 2\lambda L^\top L z = 0$$
$$\frac{\partial L}{\partial \lambda} = z^\top L^\top L z - 1 = 0$$

The first equations gives

$$(I + \lambda L^\top L) z = y$$
$$\implies z = (I + \lambda L^\top L)^{-1} y.$$

Therefore, λ must satisfy

$$g(\lambda) = y^\top (I + \lambda L^\top L)^{-1} L^\top L (I + \lambda L^\top L)^{-1} y = 1$$

which must be solved numerically using a root finding technique using e.g. Newton's method.

Lagrange Multipliers: Example

Consider the following optimization problem:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & \frac{1}{2} x^\top A x \\ \text{s.t.} & Bx = c \end{array}$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^m$.
The Lagrangian is

$$L = \frac{1}{2} x^\top A x + \lambda^\top (Bx - c)$$

Lagrange Multipliers: Example

The necessary conditions are

$$\frac{\partial L}{\partial x} = Ax + B^\top \lambda = 0$$
$$\frac{\partial L}{\partial \lambda} = Bx - c = 0$$

If A is invertible, then

$$x = -A^{-1}B^\top \lambda$$
$$\implies -BA^{-1}B^\top \lambda = c.$$

If $BA^{-1}B^\top$ is invertible then

$$\lambda = -(BA^{-1}B^\top)^{-1}c$$
$$\implies x = A^{-1}B^\top(BA^{-1}B^\top)^{-1}c$$

which is the weighted norm pseudo-inverse.

Section 2

Sufficient Conditions

Lagrange Multipliers: Sufficient Conditions

The necessary conditions tell us where a local extremum might exist, but not whether it is a local min, max, or saddle point.

Are there sufficient conditions for constrained optimization problems?

Lagrange Multipliers: Sufficient Conditions

For the unconstrained problem, we look at the Hessian of f for sufficient conditions. For unconstrained problems we look at the second derivative of L with respect to x.

$$\underbrace{L}_{1\times 1} = \underbrace{f}_{1\times 1} + \underbrace{\lambda^\top}_{1\times m} \underbrace{h}_{m\times 1} = f + \sum_{i=1}^{m} \lambda_i h_i$$

so

$$\underbrace{\nabla_x L}_{n\times 1} = \underbrace{\nabla_x f}_{n\times 1} + \underbrace{\nabla h}_{n\times m} \underbrace{\lambda}_{m\times 1} = \nabla_x f + \sum_{i=1}^{m} \lambda_i \nabla_x h_i$$

$$\nabla^2_{xx} L = \nabla^2_{xx} f + \sum_{i=1}^{m} \lambda_i \nabla^2_{xx} h_i$$

We will drop the xx notation (unless not obvious) to get

$$\boxed{\nabla^2 L = \nabla^2 f + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i}$$

Lagrange Multipliers: Sufficient Conditions

Let $P(x^*) = \{y \in \mathbb{R}^n \mid \nabla h(x^*) y = 0\}$ be the tangent plane at x^*.

Theorem (Moon Theorem 18.4)

Let f and h be C^2

1. (Necessity) Suppose that x^* is a local constrained min of f and that x^* is regular. Then $\exists \lambda$ such that

$$\nabla f(x^*) + \nabla h(x^*) \lambda = 0.$$

2. (Sufficiency) If

1. $h(x^*) = 0$
2. $\exists \lambda$ *such that* $\nabla f(x^*) + \nabla h(x^*) \lambda = 0$
3. $y^\top \nabla^2 L(x^*) y \geq 0 \qquad \forall y \in P(x^*)$

then x^* is a local constrained min of f.

Lagrange Multipliers: Sufficient Conditions: Example 18.5.1

$$\begin{aligned} \max \quad & x_1 x_2 + x_2 x_3 + x_1 x_3 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 3 \end{aligned}$$

The Lagrangian is

$$L = x_1 x_2 + x_2 x_3 + x_1 x_3 + \lambda(x_1 + x_2 + x_3 - 3)$$

The necessary conditions are therefore

$$\nabla_x L = \begin{pmatrix} x_2 + x_3 + \lambda \\ x_1 + x_3 + \lambda \\ x_2 + x_1 + \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\nabla_\lambda L = x_1 + x_2 + x_3 - 3 = 0$$

Lagrange Multipliers: Sufficient Conditions: Example 18.5.1

Therefore, we must solve

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 3 \end{pmatrix}.$$

The solution is:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix}^* = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -2 \end{pmatrix}$$

Lagrange Multipliers: Sufficient Conditions: Example 18.5.1

Is the soluution a local max?

$$\nabla^2 L = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} + \lambda \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Note that

$$\text{eig} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = -1, -1, 2$$

and so $\nabla^2 L$ is indefinite.

However, the sufficient condition requires that we restrict attention to $P(x^*)$.

Lagrange Multipliers: Sufficient Conditions: Example 18.5.1

Note that $\nabla h = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \forall x$ and so

$$P(x^*) = \{x \in \mathbb{R}^n \mid x_1 + x_2 + x_3 = 0\}$$

Therefore

$$x \in P \implies x = \begin{pmatrix} x_1 \\ x_2 \\ -(x_1 + x_2) \end{pmatrix}.$$

Restricting attention to P gives

$$x^\top \nabla^2 L x = -x_1^2 - x_3^2 - (x_1 + x_2)^2 \leq 0.$$

Therefore x^* is local maximum.

Lagrange Multipliers: Sufficient Conditions: Example 18.5.1

What did we do in this example to check the negative definite condition? We first projected the x on to the null space of $\nabla h(x^*)$ In general we can check condition (3)

$$y^\top \nabla^2 L(x^*) y \geq 0 \qquad \forall y \in P(x^*)$$

as follows:

Let E be an orthonormal basis for $\mathcal{N}(\nabla h(x^*))$, then

$$y^\top \nabla^2 L(x^*) y \geq 0 \qquad \forall y \in P(x^*) \iff E^\top \nabla^2 L(x^*) E \geq 0.$$

Lagrange Multipliers: Sufficient Conditions: Example 18.5.1

Using Matlab:

```
>> E = null([1,1,1])
```

$$\texttt{>> E =} \begin{pmatrix} -0.5774 & -0.5774 \\ 0.7887 & -0.2113 \\ -0.2113 & 0.7887 \end{pmatrix}$$

Therefore

$$E^\top \nabla^2 L(x^*) E = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \leq 0,$$

verifying the sufficient condition.

Lagrange Multipliers

Question: Is there a physical interpretation of Lagrange Multipliers?

In the book (Section 18.6) it is shown that for the optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x) = c \end{aligned}$$

where $c \neq 0$, and the solution is given by $x^*(c)$. If we let x^* be a function of c and $x^* = x^*(0)$, then

$$\frac{\partial f}{\partial c}(x^*(c))\bigg|_{c=0} = -\lambda$$

In other words, λ indicates how f changes near the optimum as the constraint values are changed.

Another way of looking at it is that the Lagrange multipliers indicate the sensitivity of x^* to changes in $h(x)$, or the steepness of f along h.