# CISC 371 Class 21

## The Lagrange Equation with Linear Equality Constraints

Texts: [1] pp. 285–296; [2] pp. 195–203; [3] p. 215; [4] pp. 374–391; [5] pp. 448–459

*Main Concepts:*

- *Lagrange multipliers add variables to a problem*
- *Lagrange function includes equality constraints*
- *Lagrange equation imposes gradient constraints*
- *Linear equality constraints produce a linear equation*

**Sample Problem, Machine Inference:** What is the force balance in a two-spring mechanical system?

The gradient of an objective function at a minimizer, and the gradient of a property function at that minimizer, are related by Theorem 20.14. In this equation, the value $\mu$ is easily computed at a point $\vec{w}_0$ that is known to be a local minimizer. This leads us to a simple and powerful question.

Can we use this theorem at a stationary point? That is, can Theorem 20.14 be used to *find* a point $\vec{w}^*$ that is a local minimizer? To do this, we would need to treat $\vec{w}^*$ as an unknown vector, the entries of which are to be computed. The value $\mu$ would also be unknown, and would also need to be computed.

One way of using the Lagrange multiplier is to subtract $-\mu\underline{\nabla}p$ from both sides and set the result to zero. This equation has two arguments : the unknown vector $\vec{w}$, and the unknown multiplier $\mu$. This function is named after its discoverer, Joseph-Louis Lagrange.

**Definition:** Lagrange function

For any convex continuous differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ and $p : \mathbb{R}^n \to \mathbb{R}$, for any local minimizer $\vec{w}_0$ of $f$ for which $p(\vec{w}_0) = 0$, the *Lagrange function* is defined as

$$\mathcal{L}(\vec{w}, \mu) = f(\vec{w}) + \mu p(\vec{w}) \tag{21.1}$$

Recall that a necessary condition for $\vec{w}^*$ to be a local minimizer is that $\vec{w}^*$ is a stationary point. Specifically, for Theorem 20.14 to apply, the augmented vector

$$\begin{bmatrix} \vec{w}^* \\ \mu^* \end{bmatrix}$$

must be a stationary point of Equation 21.1. This implies that the derivative of Equation 21.1 must be a 1-form that has each entry equal to zero. This would be two linear equations

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \vec{w}} &= \underline{0} \\ \frac{\partial \mathcal{L}}{\partial \mu} &= 0 \end{aligned} \tag{21.2}$$

We prefer to solve for vector values rather than for 1-form values, so we will take the transposes of Equation 21.2. We also prefer to solve a single linear equation whenever possible. Doing this gives us an equation that is also named after its discoverer.

**<u>Definition:</u>** Lagrange equation

For any convex continuous differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ and $p : \mathbb{R}^n \to \mathbb{R}$, for any local minimizer $\vec{w}_0$ of $f$ subject to the condition that $p(\vec{w}_0) = 0$, the *Lagrange equation* is defined from the Lagrange function $\mathcal{L}(\vec{w}, \vec{\mu})$ as

$$\begin{bmatrix} \left[\dfrac{\partial \mathcal{L}}{\partial \vec{w}}\right]^T \\ \\ \left[\dfrac{\partial \mathcal{L}}{\partial \mu}\right]^T \end{bmatrix} (\vec{w}^*, \mu^*) = \vec{0} \tag{21.3}$$

We can use Equation 21.3 to solve the first problem from the previous class. We will re-state the functions of the problem for clarity.

**<u>Problem:</u>** squared length, linear equality

$$\begin{aligned} f_1(\vec{w}) &= \vec{w}^T I \vec{w} \\ p_1(\vec{w}) &= \begin{bmatrix} -1 & 1 \end{bmatrix} \vec{w} + 3 \end{aligned} \tag{21.4}$$

The Lagrange function for Problem 21.4 is

$$\begin{aligned} \mathcal{L}_1(\vec{w}, \mu) &= f_1(\vec{w}) - \mu p_1(\vec{w}) \\ &= \vec{w}^T I \vec{w} + \mu(\begin{bmatrix} -1 & 1 \end{bmatrix} \vec{w} + 3) \end{aligned} \tag{21.5}$$

The derivatives of Equation 21.5 are

$$\frac{\partial \mathcal{L}_1}{\partial \vec{w}} = 2\vec{w}^T I + \begin{bmatrix} -1 & 1 \end{bmatrix} \mu \tag{21.6}$$

$$\frac{\partial \mathcal{L}_1}{\partial \mu} = \begin{bmatrix} -1 & 1 \end{bmatrix} \vec{w}^T + 3 \tag{21.7}$$

Collecting the terms $\vec{w}$ and $\mu$ into an augmented vector of unknowns, and setting the transposes of the derivatives to zero, gives

$$\begin{bmatrix} 2I & \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ \begin{bmatrix} -1 & 1 \end{bmatrix} & 0 \end{bmatrix} \begin{bmatrix} \vec{w}^* \\ \mu^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix} \tag{21.8}$$

A numerical solution to Equation 21.8 gives the minimizer $\vec{w}^*$ and the Lagrange multiplier as

$$\begin{aligned} \vec{w}^* &= \begin{bmatrix} 1.5 \\ -1.5 \end{bmatrix} \\ \mu^* &= 3 \end{aligned} \tag{21.9}$$

The values of Equation 21.9 are the same as those in Problem 20.4.

**Problem:** Hooke springs, external force

Suppose that two springs are modeled using Hooke's Law, and that the springs are connected to ground via a single mass. Further suppose that the springs are compressed by independent forces, and that there is also a known external force. We study the case where the springs are in equilibrium, which implies that the mass is not moving. This is illustrated in Figure 21.1.

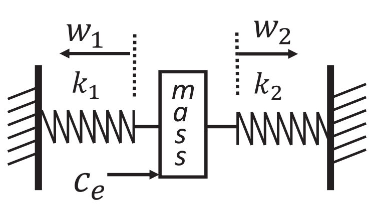


**Figure 21.1:** A two-spring mechanical system. Spring 1 has stiffness coefficient $k_1$ and Spring 2 has coefficient $k_2$. The springs are displaced by $w_1$ and $w_2$ from their reference points. An external load $c_e$ is applied to the system, which is assumed to be in equilibrium.

The Hooke model of a spring gives each force $c_i$, and each potential energy $z_i$, as functions of the spring stiffness $k_i$ and the displacement $x_i$. These forces and energies are

$$\begin{aligned}
c_i &= k_i w_i \qquad (21.10)\\
z_i &= \frac{1}{2} k_i w_i^2\\
\Rightarrow \quad z_1 + z_2 &= \frac{1}{2} \vec{w}^T K \vec{w}\\
\text{where} \quad K &= \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}
\end{aligned}$$

The springs are in equilibrium so, taking into account the direction of the application of force, the forces are in balance when

$$\begin{aligned}
c_2 - c_1 &= c_e\\
\equiv \quad -k_1 x_1 + k_2 x_2 &= c_e \qquad (21.11)\\
\equiv \quad \begin{bmatrix} -k_1 & k_2 \end{bmatrix} \vec{w} &= c_e
\end{aligned}$$

We can formulate this as an optimization problem that has a quadratic objective function:

$$\begin{aligned}
f_2(\vec{w}) &= \frac{1}{2} \vec{w}^T K \vec{w}\\
\text{where} \quad p_2(\vec{w}) &= \underline{m}\vec{w} - c_e \qquad (21.12)\\
K &= \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}\\
\underline{m} &= \begin{bmatrix} -k_1 & k_2 \end{bmatrix}
\end{aligned}$$

The Lagrange function for Problem 21.12 is

$$\begin{aligned}
\mathcal{L}_2(\vec{w}, \mu) &= f_2(\vec{w}) - \mu p_2(\vec{w}) \qquad (21.13)\\
&= \frac{1}{2} \vec{w}^T K \vec{w} + \mu(\underline{m}\vec{w} - c_e)
\end{aligned}$$

The derivatives of Equation 21.5 are

$$\frac{\partial \mathcal{L}_2}{\partial \vec{w}} = \vec{w}^T K + \underline{m}\mu \qquad (21.14)$$

$$\frac{\partial \mathcal{L}_2}{\partial \mu} = \underline{m}\vec{w} - c_e \qquad (21.15)$$

Collecting the terms $\vec{w}$ and $\mu$ into an augmented vector of unknowns, and setting the transposes of the derivatives to zero, gives

$$\begin{bmatrix} K & \underline{m}^T \\ \underline{m} & 0 \end{bmatrix} \begin{bmatrix} \vec{w}^* \\ \mu^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ c_e \end{bmatrix} \tag{21.16}$$

Observation: by construction, if $K$ is symmetric and positive definite, and $\vec{m} \neq \vec{0}$, then the matrix in Equation 21.16 is symmetric and *indefinite*. Two eigenvalues are positive and one eigenvalue is negative. A derivation of this observation is in the extra notes for this class.

We can use specific values and solve a sample form of this problem. Suppose that the stiffnesses and external load are

$$\begin{aligned} k_1 &= 1 \\ k_2 &= 2 \\ c_e &= 6 \end{aligned} \tag{21.17}$$

When we substitute the values of Equation 21.17 into Equation 21.14, we have

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 2 \\ -1 & 2 & 0 \end{bmatrix} \begin{bmatrix} \vec{w}^* \\ \mu^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} \tag{21.18}$$

When we numerically solve Equation 21.18, using 2 decimal places of numerical precision for display, we have

$$\begin{bmatrix} \vec{w}^* \\ \mu^* \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ -2 \end{bmatrix} \tag{21.19}$$

Equation 21.19 provides the same values as we get when we use the `MATLAB` Optimization Toolbox to perform constrained optimization of Problem 21.12 with the values of Equation 21.18. The signs of the eigenvalues of the matrix of Equation 21.18 are, as expected, a mixture of 2 positive eigenvalues and 1 negative eigenvalue.

## 21.1 General Equality Constraints

For an objective function $f(\cdot)$ that is quadratic in $\vec{w}$, having a symmetric positive definitive $K$ and a linear constraint matrix $M_{l\times n}$ that is full rank, optimization of a quadratic objective function with linear equality constraints is

$$\begin{aligned}
\vec{w}^* &= \underset{\vec{w}\in\mathbb{R}^n}{\operatorname{argmin}} f(\vec{w}) \\
\text{and} \quad p(\vec{w}^*) &= 0 \\
\text{where:} \\
f(\vec{w}) &= \frac{1}{2}\vec{w}^T K\vec{w} + \vec{q}^T\vec{w} \\
p(\vec{w}^*) &= M\vec{w}^* - \vec{c}
\end{aligned} \tag{21.20}$$

There are $l$ Lagrange multipliers $\mu_i$ in the Lagrange function of Problem 21.20. The objective function maps to a scalar value and the Lagrange multipliers must also map to a scalar value for the Lagrange function to be well formed.

The extra notes for this class have a derivation using linear algebra. The result is simple to state.

**<u>Theorem:</u>** KKT matrix gives the minimizer of a quadratic objective with linear equality constraints

*A solution to Problem 21.20 is given by the linear equation*

$$\begin{bmatrix} K & M^T \\ M & 0 \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{\mu} \end{bmatrix} = \begin{bmatrix} -\vec{q} \\ \vec{c} \end{bmatrix} \tag{21.21}$$

**Proof:** See the extra notes for this class.

<u>Observation</u>: Equation 21.16 is a special case of Theorem 21.21, where $M = \underline{m}$ and $\vec{q} = \vec{0}$. The matrix in Theorem 21.21 is often named the KKT matrix because the solution is a KKT point, which we will define and use in a subsequent class in this course.

## 21.2 Extra Notes on Linearly Constrained Quadratic Objectives

**Derivation:** solution to a quadratic objective with linear equality constraints

In solving Problem 21.20, there are two useful ways to represent the sum of the products of each Lagrange multiplier with the $i^{\text{th}}$ constraint, which produces a scalar result. This is because the sum of the products is effectively a dot product, which can be written in two ways as the product of a transposed vector and a vector.

$$\begin{aligned} \mu_i(\vec{m}_i^T\vec{w} - c_i) &= (\vec{w}^T\vec{m} - c_i)\mu_i \\ \Rightarrow \quad g(\vec{w},\vec{\mu}) &\stackrel{\text{def}}{=} \sum_{i=1}^{l} \mu_i(\vec{m}_i^T\vec{w} - c_i) \\ &= \vec{\mu}^T[M\vec{w} - \vec{c}] \\ &= [\vec{w}^T M^T - \vec{c}^T]\vec{\mu} \end{aligned} \tag{21.22}$$

We know how to differentiate a linear equation of a vector with respect to a vector. We can use the appropriate version of Equation 21.22 to differentiate the vector of interest:

$$\begin{aligned} \frac{\partial g}{\partial \vec{w}} &= \vec{\mu}^T M \\ \frac{\partial g}{\partial \vec{\mu}} &= \vec{w}^T M^T - \vec{c}^T \end{aligned} \tag{21.23}$$

After these preliminaries, we form the Lagrange function of Problem 21.20 as

$$\mathcal{L}(\vec{w},\vec{\mu}) \stackrel{\text{def}}{=} f(\vec{w}) + \vec{\mu}^T[M\vec{w} - \vec{c}] \tag{21.24}$$

Next, we differentiate Equation 21.24, using intermediate results of Equation 21.23, to get

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \vec{w}} &= \vec{w}^T K + \vec{q}^T + \vec{\mu}^T M \\ \frac{\partial \mathcal{L}}{\partial \vec{\mu}} &= \vec{w}^T M^T - \vec{c}^T \end{aligned} \tag{21.25}$$

We can transpose Equations 21.25, which is in terms of 1-forms, to equations that are in terms of vectors. We also impose the necessary condition for stationarity, which is equality to the zero vector, to get

$$\begin{aligned}\left[\vec{w}^T K + \vec{q}^T + \vec{\mu}^T M\right]^T &= \vec{0}_{m\times 1}\\ \left[\vec{w}^T M^T - \vec{c}^T\right]^T &= \vec{0}_{l\times 1}\end{aligned} \tag{21.26}$$

After we simplify by propagating the transposes, and use the symmetry property $K = K^T$, we have

$$\begin{aligned}\left[K\vec{w} + \vec{q} + M^T\vec{\mu}\right]^T &= \vec{0}\\ \left[M\vec{w} - \vec{c}\right] &= \vec{0}\end{aligned} \tag{21.27}$$

We can collect the terms of Equation 21.27 into a single linear equation

$$\begin{bmatrix} K & M^T \\ M & 0 \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{\mu} \end{bmatrix} = \begin{bmatrix} -\vec{q} \\ \vec{c} \end{bmatrix}$$

which is Theorem 21.21.

Temporarily abbreviating the matrix in Theorem 21.21 as

$$W = \begin{bmatrix} K & M^T \\ M & 0 \end{bmatrix} \tag{21.28}$$

we see that $W$ is symmetric and partitioned into blocks.

The matrix $K$ was assumed – in Problem 21.20 – to be symmetric and positive definite so $K^{-1}$ exists. We can define the matrix that reduces the blocks of $W$ in Equation 21.28, using Gaussian elimination, as

$$E_W \stackrel{\text{def}}{=} \begin{bmatrix} I & 0 \\ -MK^{-1} & I \end{bmatrix} \tag{21.29}$$

$E_W$ in Equation 21.29 is a unit lower triangular matrix, so it is non-singular and can be used in a similarity transformation.

Pre-multiplying $W$ by $E_W$, and post-multiplying the result by $E_W^T$, shows that

$$E_W W E_W^T = C = \begin{bmatrix} K & 0 \\ 0 & -MM^T \end{bmatrix} \tag{21.30}$$

The matrix $M$ was assumed – in Problem 21.20 – to be full rank, so $[M^T]$ is symmetric and positive definite. Therefore $[-MM^T]$ is symmetric and negative definite, so the block-diagonal matrix $C$ in Equation 21.30 is indefinite: it has positive eigenvalues arising from the $[K]$ block and negative eigenvalues arising from the $[-MM^T]$ block.

Specifically, every eigenvector of $C$ that has a positive eigenvalue has the form

$$\begin{bmatrix} \vec{w} \\ \vec{0} \end{bmatrix}$$

and every eigenvector of $C$ that has a negative eigenvalue has the form

$$\begin{bmatrix} \vec{0} \\ \vec{\mu} \end{bmatrix}$$

which can be proved by expansion.

End of Extra Notes

## References


[1] Antoniou A, Lu WS: Practical Optimization: Algorithms and Engineering Applications. Springer Science & Business Media, 2007

[2] Beck A: Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. Siam Press, 2014

[3] Boyd S, Vandenberghe L: Convex Optimization. Cambridge University Press, 2004

[4] Chong EKP, Zak SH: An Introduction to Optimization, volume 76. John Wiley & Sons, 2013

[5] Nocedal J, Wright S: Numerical Optimization. Springer Science & Business Media, 2006