Fundamentals of Linear Algebra and Optimization
Lagrange Multipliers

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

March 19, 2025

Constrained Optimization

In many practical situations, we need to look for local extrema of a function J *under additional constraints*. This situation can be formalized conveniently as follows. We have a function $J\colon \Omega \to \mathbb{R}$ defined on some open subset Ω of a normed vector space, but we also have some subset U of Ω, and we are looking for the local extrema of J *with respect to the set U*.

Constrained Optimization

In many practical situations, we need to look for local extrema of a function J *under additional constraints*. This situation can be formalized conveniently as follows. We have a function $J\colon \Omega \to \mathbb{R}$ defined on some open subset Ω of a normed vector space, but we also have some subset U of Ω, and we are looking for the local extrema of J *with respect to the set U*.

The elements $u \in U$ are often called *feasible solutions* of the optimization problem consisting in finding the local extrema of some objective function J with respect to some subset U of Ω defined by a set of constraints. Note that in most cases, U is *not* open. In fact, U is usually closed.

Constrained Local Extrema

Definition. If $J\colon \Omega \to \mathbb{R}$ is a real-valued function defined on some open subset Ω of a normed vector space E and if U is some subset of Ω, we say that J has a *local minimum* (or *relative minimum*) at the point $u \in U$ *with respect to U* if there is some open subset $W \subseteq \Omega$ containing u such that

$$J(u) \leq J(w) \quad \text{for all } w \in U \cap W.$$

Constrained Local Extrema

Definition. If $J\colon \Omega \to \mathbb{R}$ is a real-valued function defined on some open subset Ω of a normed vector space E and if U is some subset of Ω, we say that J has a *local minimum* (or *relative minimum*) at the point $u \in U$ *with respect to U* if there is some open subset $W \subseteq \Omega$ containing u such that

$$J(u) \leq J(w) \quad \text{for all } w \in U \cap W.$$

Similarly, we say that J has a *local maximum* (or *relative maximum*) at the point $u \in U$ *with respect to U* if there is some open subset $W \subseteq \Omega$ containing u such that

$$J(u) \geq J(w) \quad \text{for all } w \in U \cap W.$$

In either case, we say that J has a *local extremum* at u *with respect to U*.

Equality Constraints

In order to find necessary conditions for a function $J\colon \Omega \to \mathbb{R}$ to have a local extremum with respect to a subset U of Ω (where Ω is open), we need to *incorporate* the definition of U into these conditions. This can be done when the set U is defined by a set of equations,

$$U = \{x \in \Omega \mid \varphi_i(x) = 0,\ \ 1 \leq i \leq m\},$$

where the functions $\varphi_i\colon \Omega \to \mathbb{R}$ are continuous (and usually differentiable).

Equality Constraints

In order to find necessary conditions for a function $J\colon \Omega \to \mathbb{R}$ to have a local extremum with respect to a subset U of Ω (where Ω is open), we need to *incorporate* the definition of U into these conditions. This can be done when the set U is defined by a set of equations,

$$U = \{x \in \Omega \mid \varphi_i(x) = 0,\ \ 1 \leq i \leq m\},$$

where the functions $\varphi_i\colon \Omega \to \mathbb{R}$ are continuous (and usually differentiable).

The equations $\varphi_i(x) = 0$ are called *equality constraints*.

Necessary Condition for Constrained Extrema

In the case of equality constraints, a *necessary condition* for a local extremum with respect to U can be given in terms of *Lagrange multipliers*.

Necessary Condition for Constrained Extrema

Theorem (*Necessary condition for a constrained extremum in terms of Lagrange multipliers*). Let Ω be an open subset of $\mathbb{R}^n$, consider m C^1-functions $\varphi_i \colon \Omega \to \mathbb{R}$ (with $1 \leq m < n$), let

$$U = \{v \in \Omega \mid \varphi_i(v) = 0,\ 1 \leq i \leq m\},$$

and let $u \in U$ be a point such that the derivatives $d\varphi_i(u) \in \mathcal{L}(\mathbb{R}^n; \mathbb{R})$ are *linearly independent*; equivalently, assume that the $m \times n$ matrix $\big((\partial \varphi_i / \partial x_j)(u)\big)$ has rank m.

Necessary Condition for Constrained Extrema

If $J\colon \Omega \to \mathbb{R}$ is a function which is differentiable at $u \in U$ and if J has a local constrained extremum at u, then there exist m numbers $\lambda_i(u) \in \mathbb{R}$, uniquely defined, such that

$$dJ(u) + \lambda_1(u) d\varphi_1(u) + \cdots + \lambda_m(u) d\varphi_m(u) = 0;$$

Necessary Condition for Constrained Extrema

If $J\colon \Omega \to \mathbb{R}$ is a function which is differentiable at $u \in U$ and if J has a local constrained extremum at u, then there exist m numbers $\lambda_i(u) \in \mathbb{R}$, uniquely defined, such that

$$dJ(u) + \lambda_1(u)d\varphi_1(u) + \cdots + \lambda_m(u)d\varphi_m(u) = 0;$$

or equivalently,

$$\nabla J(u) + \lambda_1(u)\nabla\varphi_1(u) + \cdots + \lambda_m(u)\nabla\varphi_m(u) = 0.$$

Lagrange Multipliers

Definition. The numbers $\lambda_i(u)$ involved in the preceding theorem are called the *Lagrange multipliers* associated with the constrained extremum u.

Lagrange Multipliers

Definition. The numbers $\lambda_i(u)$ involved in the preceding theorem are called the *Lagrange multipliers* associated with the constrained extremum u.

The linear independence of the linear forms $d\varphi_i(u)$ is equivalent to the fact that the Jacobian matrix $\big((\partial\varphi_i/\partial x_j)(u)\big)$ of $\varphi = (\varphi_1, \ldots, \varphi_m)$ at u has rank m. If $m = 1$, the linear independence of the $d\varphi_i(u)$ reduces to the condition $\nabla\varphi_1(u) \neq 0$.

The Lagrangian

A fruitful way to reformulate the use of Lagrange multipliers is to introduce the notion of the Lagrangian associated with our constrained extremum problem.

The Lagrangian

A fruitful way to reformulate the use of Lagrange multipliers is to introduce the notion of the Lagrangian associated with our constrained extremum problem.

Definition. The *Lagrangian* associated with our constrained extremum problem is the function $L\colon \Omega \times \mathbb{R}^m \to \mathbb{R}$ given by

$$L(v, \lambda) = J(v) + \lambda_1 \varphi_1(v) + \cdots + \lambda_m \varphi_m(v),$$

with $\lambda = (\lambda_1, \ldots, \lambda_m)$.

Critical Point of the Lagrangian

Proposition. There exists some $\mu = (\mu_1, \dots, \mu_m)$ and some $u \in U$ such that

$$dJ(u) + \mu_1 d\varphi_1(u) + \cdots + \mu_m d\varphi_m(u) = 0$$

if and only if

$$dL(u, \mu) = 0,$$

or equivalently

$$\nabla L(u, \mu) = 0;$$

that is, iff (u, μ) is a *critical point* of the Lagrangian L.

Lagrangian System

If we write out explicitly the condition

$$dJ(u) + \lambda_1 d\varphi_1(u) + \cdots + \lambda_m d\varphi_m(u) = 0,$$

we get the $n \times m$ system

$$\begin{aligned} \frac{\partial J}{\partial x_1}(u) + \lambda_1 \frac{\partial \varphi_1}{\partial x_1}(u) + \cdots + \lambda_m \frac{\partial \varphi_m}{\partial x_1}(u) &= 0 \\ &\vdots \\ \frac{\partial J}{\partial x_n}(u) + \lambda_1 \frac{\partial \varphi_1}{\partial x_n}(u) + \cdots + \lambda_m \frac{\partial \varphi_m}{\partial x_n}(u) &= 0, \end{aligned}$$

Lagrangian System

and it is important to note that the matrix of this system is the *transpose* of the Jacobian matrix of φ at u. If we write $\mathrm{Jac}(\varphi)(u) = \big((\partial \varphi_i/\partial x_j)(u)\big)$ for the Jacobian matrix of φ (at u), then the above system is written in matrix form as

$$\nabla J(u) + (\mathrm{Jac}(\varphi)(u))^{\top}\lambda = 0,$$

where λ is viewed as a column vector, and the Lagrangian is equal to

$$L(u, \lambda) = J(u) + (\varphi_1(u), \ldots, \varphi_m(u))\lambda.$$

The Lagrangian Technique

The beauty of the Lagrangian is that the constraints $\{\varphi_i(v) = 0\}$ have been incorporated into the function $L(v, \lambda)$, and that the necessary condition for the existence of a constrained local extremum of J is reduced to the necessary condition for the existence of a local extremum of the *unconstrained* L.

Lagrangian Technique Counterexample

One should be careful to check that the assumptions of the preceding theorem are satisfied (in particular, the linear independence of the linear forms $d\varphi_i$).

Lagrangian Technique Counterexample

One should be careful to check that the assumptions of the preceding theorem are satisfied (in particular, the linear independence of the linear forms $d\varphi_i$).

Example. Let $J\colon \mathbb{R}^3 \to \mathbb{R}$ be given by

$$J(x, y, z) = x + y + z^2$$

and $g\colon \mathbb{R}^3 \to \mathbb{R}$ by

$$g(x, y, z) = x^2 + y^2.$$

Since $g(x, y, z) = 0$ iff $x = y = 0$, we have $U = \{(0, 0, z) \mid z \in \mathbb{R}\}$ and the restriction of J to U is given by $J(0, 0, z) = z^2$, which has a minimum for $z = 0$.

Lagrangian Technique Counterexample

However, a “blind” use of Lagrange multipliers would require that there is some λ so that

$$\frac{\partial J}{\partial x}(0,0,z) = \lambda \frac{\partial g}{\partial x}(0,0,z), \;\; \frac{\partial J}{\partial y}(0,0,z) = \lambda \frac{\partial g}{\partial y}(0,0,z), \;\; \frac{\partial J}{\partial z}(0,0,z) = \lambda \frac{\partial g}{\partial z}(0,0,z),$$

Lagrangian Technique Counterexample

However, a "blind" use of Lagrange multipliers would require that there is some λ so that

$$\frac{\partial J}{\partial x}(0,0,z) = \lambda \frac{\partial g}{\partial x}(0,0,z), \ \frac{\partial J}{\partial y}(0,0,z) = \lambda \frac{\partial g}{\partial y}(0,0,z), \ \frac{\partial J}{\partial z}(0,0,z) = \lambda \frac{\partial g}{\partial z}(0,0,z),$$

and since

$$\frac{\partial g}{\partial x}(x,y,z) = 2x, \quad \frac{\partial g}{\partial y}(x,y,z) = 2y, \quad \frac{\partial g}{\partial z}(0,0,z) = 0,$$

Lagrangian Technique Counterexample

the partial derivatives above all vanish for $x = y = 0$, so at a local extremum we should also have

$$\frac{\partial J}{\partial x}(0,0,z) = 0, \quad \frac{\partial J}{\partial y}(0,0,z) = 0, \quad \frac{\partial J}{\partial z}(0,0,z) = 0,$$

but this is absurd since

$$\frac{\partial J}{\partial x}(x,y,z) = 1, \quad \frac{\partial J}{\partial y}(x,y,z) = 1, \quad \frac{\partial J}{\partial z}(x,y,z) = 2z.$$

Lagrangian Technique Counterexample

the partial derivatives above all vanish for $x = y = 0$, so at a local extremum we should also have

$$\frac{\partial J}{\partial x}(0,0,z) = 0, \quad \frac{\partial J}{\partial y}(0,0,z) = 0, \quad \frac{\partial J}{\partial z}(0,0,z) = 0,$$

but this is absurd since

$$\frac{\partial J}{\partial x}(x,y,z) = 1, \quad \frac{\partial J}{\partial y}(x,y,z) = 1, \quad \frac{\partial J}{\partial z}(x,y,z) = 2z.$$

The reader should enjoy finding the reason for the flaw in the argument.

Lagrangian Provides a Necessary Condition

Keep in mind that the preceding theorem gives only a *necessary condition*. The (u, λ) may *not* correspond to local extrema! *Thus it is always necessary to analyze the local behavior of J near a critical point u.*

Lagrangian Technique Example

Example. Let us apply the above method to the following example in which $E_1 = \mathbb{R}$, $E_2 = \mathbb{R}$, $\Omega = \mathbb{R}^2$, and

$$J(x_1, x_2) = -x_2$$
$$\varphi(x_1, x_2) = x_1^2 + x_2^2 - 1.$$

Lagrangian Technique Example

Observe that

$$U = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}$$

is the unit circle, and since

$$\nabla\varphi(x_1, x_2) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix},$$

Lagrangian Technique Example

Observe that

$$U = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}$$

is the unit circle, and since

$$\nabla\varphi(x_1, x_2) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix},$$

it is clear that $\nabla\varphi(x_1, x_2) \neq 0$ for every point $= (x_1, x_2)$ on the unit circle.

Lagrangian Technique Example

If we form the Lagrangian

$$L(x_1, x_2, \lambda) = -x_2 + \lambda(x_1^2 + x_2^2 - 1),$$

a necessary condition for J to have a constrained local extremum is that $\nabla L(x_1, x_2, \lambda) = 0$,

Lagrangian Technique Example

If we form the Lagrangian

$$L(x_1, x_2, \lambda) = -x_2 + \lambda(x_1^2 + x_2^2 - 1),$$

a necessary condition for J to have a constrained local extremum is that $\nabla L(x_1, x_2, \lambda) = 0$, so the following equations must hold:

$$\begin{aligned} 2\lambda x_1 &= 0 \\ -1 + 2\lambda x_2 &= 0 \\ x_1^2 + x_2^2 &= 1. \end{aligned}$$

Lagrangian Technique Example

The second equation implies that $\lambda \neq 0$, and then the first yields $x_1 = 0$, so the third yields $x_2 = \pm 1$, and we get two solutions:

$$\lambda = \frac{1}{2}, \qquad (x_1, x_2) = (0, 1)$$

$$\lambda = -\frac{1}{2}, \qquad (x_1', x_2') = (0, -1).$$

Lagrangian Technique Example

The second equation implies that $\lambda \neq 0$, and then the first yields $x_1 = 0$, so the third yields $x_2 = \pm 1$, and we get two solutions:

$$\lambda = \frac{1}{2}, \qquad (x_1, x_2) = (0, 1)$$
$$\lambda = -\frac{1}{2}, \qquad (x_1', x_2') = (0, -1).$$

We can check immediately that the first solution is a minimum and the second is a maximum.

Lagrangian Technique Example

The second equation implies that $\lambda \neq 0$, and then the first yields $x_1 = 0$, so the third yields $x_2 = \pm 1$, and we get two solutions:

$$\lambda = \frac{1}{2}, \qquad (x_1, x_2) = (0, 1)$$
$$\lambda = -\frac{1}{2}, \qquad (x_1', x_2') = (0, -1).$$

We can check immediately that the first solution is a minimum and the second is a maximum.

The reader should look for a geometric interpretation of this problem.

Lagrangian for Quadratic Optimization

Example. Let us now consider the case in which J is a quadratic function of the form

$$J(v) = \frac{1}{2} v^\top A v - v^\top b,$$

where A is an $n \times n$ *symmetric matrix*, $b \in \mathbb{R}^n$, and the constraints are given by a linear system of the form

$$Cv = d,$$

where C is an $m \times n$ matrix with $m < n$ and $d \in \mathbb{R}^m$. We also assume that C has rank m.

Lagrangian for Quadratic Optimization

Example. Let us now consider the case in which J is a quadratic function of the form

$$J(v) = \frac{1}{2}v^\top Av - v^\top b,$$

where A is an $n \times n$ *symmetric matrix*, $b \in \mathbb{R}^n$, and the constraints are given by a linear system of the form

$$Cv = d,$$

where C is an $m \times n$ matrix with $m < n$ and $d \in \mathbb{R}^m$. We also assume that C has rank m. In this case the function φ is given by

$$\varphi(v) = (Cv - d)^\top,$$

Lagrangian for Quadratic Optimization

and since we showed earlier that

$$d\varphi(v)(w) = (Cw)^\top,$$

the condition that the Jacobian matrix of φ at u has rank m is satisfied.

Lagrangian for Quadratic Optimization

and since we showed earlier that

$$d\varphi(v)(w) = (Cw)^\top,$$

the condition that the Jacobian matrix of φ at u has rank m is satisfied. The Lagrangian of this problem is

$$L(v, \lambda) = \frac{1}{2} v^\top A v - v^\top b + (Cv - d)^\top \lambda = \frac{1}{2} v^\top A v - v^\top b + v^\top C^\top \lambda - d^\top \lambda,$$

where λ is viewed as a column vector.

Lagrangian for Quadratic Optimization

and since we showed earlier that

$$d\varphi(v)(w) = (Cw)^\top,$$

the condition that the Jacobian matrix of φ at u has rank m is satisfied. The Lagrangian of this problem is

$$L(v, \lambda) = \frac{1}{2} v^\top A v - v^\top b + (Cv - d)^\top \lambda = \frac{1}{2} v^\top A v - v^\top b + v^\top C^\top \lambda - d^\top \lambda,$$

where λ is viewed as a column vector. Now because A is a symmetric matrix, we showed earlier that

$$\nabla L(v, \lambda) = \begin{pmatrix} Av - b + C^\top \lambda \\ Cv - d \end{pmatrix}.$$

Lagrangian for Quadratic Optimization

Therefore, the necessary condition for constrained local extrema is

$$\begin{aligned} Av + C^\top \lambda &= b \\ Cv &= d, \end{aligned}$$

Lagrangian for Quadratic Optimization

Therefore, the necessary condition for constrained local extrema is

$$\begin{aligned} Av + C^\top \lambda &= b \\ Cv &= d, \end{aligned}$$

which can be expressed in matrix form as

$$\begin{pmatrix} A & C^\top \\ C & 0 \end{pmatrix} \begin{pmatrix} v \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix},$$

where the matrix of the system is a symmetric matrix.

Lagrangian and Quadratic Optimization

This example will be further discussed in the next module. As we will show, the function J has a minimum iff A is positive definite, so in general, if A is only a symmetric matrix, the critical points of the Lagrangian do *not* correspond to extrema of J.