# ECEn 671: Mathematics of Signals and Systems
# Moon: Chapter 18.

Randal W. Beard

Brigham Young University

December 8, 2020

# Table of Contents

# Section 1

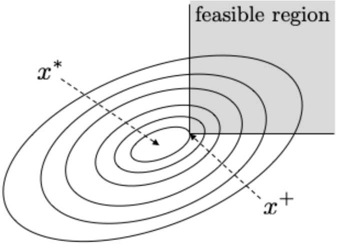
## Constrained Optimization

# Constrained Optimization

In Chapter 14 we studied unconstrained minimization of continuously differentiable functions.

In Chapter 18 we focus on constrained optimization problems.

For example, given the level curves,




Note that the constrained optimum $x^+$ does not equal the unconstrained optimum $x^*$.

The unconstrained optimum is $x^*$; the constained optimum is $x^+$.

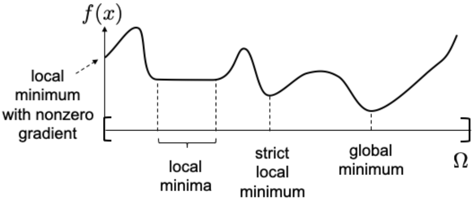
# Constrained Optimization

## Definition

Let $\Omega \subseteq \mathbb{R}^n$ be the feasible region. Then $x^* \in \Omega$ is a <u>local minimum</u> of $f : \mathbb{R}^n \to \mathbb{R}$ over $\Omega$ if $\exists \epsilon > 0$ such that

$$x \in \Omega \cap \{y \in \mathbb{R}^n : |u - x^*| < \epsilon\} \quad \Longrightarrow \quad f(x) \geq f(x^*).$$

If $f(x) > f(x^*)$ then $x^*$ is a <u>strict local minimum</u>. If true for all $\epsilon > 0$ then $x^*$ is a global minimum.



# Constrained Optimization

## Definition

Let $x \in \Omega$ and $d \in \mathbb{R}^n$, then

$$y = x + \alpha d$$

is a feasible point if $y \in \Omega$.

## Definition

The vector $d$ is a feasible direction at $x$, if $\exists \epsilon_0 > 0$ such that

$$x + \epsilon d \in \Omega$$

for every $0 \leq \epsilon \leq \epsilon_0$.

# Constrained Optimization

Recall, if $f : \mathbb{R}^n \to \mathbb{R}$ then the gradient vector is

$$\frac{\partial f}{\partial x} = \nabla_x f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

and the Hessian matrix is

$$\frac{\partial^2 f}{\partial x^2} = \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

If $\Omega = \mathbb{R}^n$ then a necessary condition for $x^*$ to be a local minima is that $\nabla_x f(x^*) = 0$. What about constrained optimization problems?

# Constrained Optimization

## Theorem (Moon Theorem 18.1)

*Let $\Omega \subseteq \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^1$ (continuously differentiable) on $\Omega$.*

1. *If $x^*$ is a local minimum of $f$ over $\Omega$, then for any feasible direction $d \in \mathbb{R}^n$ at $x^*$*

$$[\nabla_x f(x^*)]^\top d \geq 0$$

2. *If $x^*$ is an interior point of $\Omega$, then*

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

3. *If in addition, $f \in \mathcal{C}^2$ and $\nabla_x f(x^*)^\top d = 0$, then*

$$d^\top \nabla^2 f(x^*) d \geq 0$$

*Note that this is a weaker condition than psd Hessian.*

# Proof of Theorem 18.1

1. By Taylor series expansion,

$$f(x^* + \epsilon d) = f(x^*) + \epsilon \nabla_x f(x^*)^\top d + O(\epsilon)$$
$$\Longrightarrow f(x^* + \epsilon d) - f(x^*) = \epsilon \nabla_x f(x^*)^\top d + O(\epsilon)$$

Since $x^*$ is a local minimum, for $\epsilon$ sufficiently small we must have that

$$\Longrightarrow f(x^* + \epsilon d) - f(x^*) \geq 0$$
$$\Longrightarrow \nabla_x f(x^*)^\top d \geq 0.$$

# Proof of Theorem 18.1, cont.

2. If $x^*$ is an interior point then every $d \in \mathbb{R}^n$ is feasible at $x^*$, i.e.

$$\langle \nabla_x f(x^*), d \rangle_{\mathbb{R}^n} = 0, \qquad \forall d \in \mathbb{R}^n.$$

Therefore,

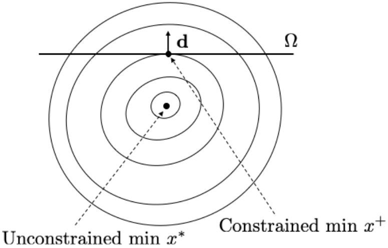
$$\begin{aligned} & \nabla_x f(x^*)^\top d \geq 0 \quad \text{and} \quad \nabla_x f(x^*)^\top (-d) \geq 0 \\ \Longrightarrow\ & \nabla_x f(x^*)^\top d = 0, \quad \forall d \in \mathbb{R}^n \\ \Longrightarrow\ & \nabla_x f(x^*) = 0 \end{aligned}$$

since $R^n$ is a finite dimensional vector space .

# Proof of Theorem 18.1, cont.

3. If $\nabla_x f(x^*)^\top d = 0$ then the Taylor series for $f$ is

$$
\begin{aligned}
& f(x^* + \epsilon d) = f(x^*) + \epsilon^2 d^\top \nabla^2 f(x^*) d + O(\epsilon^2) \\
\implies & 0 \le f(x^* + \epsilon d) - f(x^*) = \epsilon^2 d^\top \nabla^2 f(x^*) d + O(\epsilon^2) \\
\implies & d^\top \nabla^2 f(x^*) d \ge 0.
\end{aligned}
$$




Note: Any feasible $d$ points uphill.

Note: The function is concave in feasible region.

# Constrained Optimization: Sufficient Conditions

Are there sufficient conditions?

First, suppose that the constraints are not active, i.e. $x^*$ is an interior point of $\Omega$. (We will consider the active constraint case later.)

Theorem (Moon Theorem 18.2)

*Let $f \in \mathcal{C}^2$ on $\Omega$ and let $x^*$ be an interior point of $\Omega$. If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite then $x^*$ is a strict local minimum of $f$.*

# Constrained Optimization: Sufficient Conditions: Proof

Proof.

Let $d$ be any unit vector in $\mathbb{R}^n$ then

$$
\begin{aligned}
& f(x^* + \epsilon d) = f(x^*) + \epsilon \nabla f(x^*)^\top d + \epsilon^2 d^\top \nabla^2 f(x^*) d + O(\epsilon^2) \\
\implies & f(x^* + \epsilon d) - f(x^*) = \epsilon^2 d^\top \nabla^2 f(x^*) d + O(\epsilon^2)
\end{aligned}
$$

Since $\nabla^2 f(x^*)$ is positive definite, it follows that for $\epsilon$ sufficiently small

$$f(x^* + \epsilon d) - f(x^*) > 0,$$

which implies that $x^*$ is a strict local minimum. $\square$

**Note:** we cannot generalize this theorem to the case when $\nabla^2 f(x^*)$ is p.s.d.. Why?

Section 2

# General Constrained Optimization

# Constrained Optimization

In general we have two types of constraints:

1. Equality constraints of the form

$$h_i(x) = 0$$

For example:

$$h_1(x) \triangleq x_1^2 + x_1 x_2 x_3 + \tan(x_3)\cos(x_2) = 0$$

2. Inequality constraints of the form

$$g_i(x) \leq 0$$

For example

$$\begin{aligned} x_1 \geq 0, & \qquad x_2 \geq 0 \\ \Longrightarrow g_1(x) \triangleq -x_1 \leq 0, & \qquad g_2(x) \triangleq -x_2 \leq 0 \end{aligned}$$

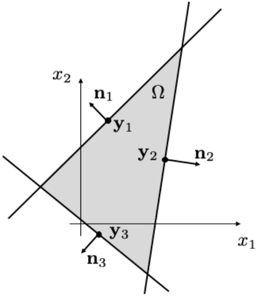
# Constrained Optimization

In fact a region $\Omega \subset \mathbb{R}^n$ can always be described by inequality constraints.

## Example

Feasible Region $\Omega$:

$$(x - \mathbf{y}_1)^\top \mathbf{n}_1 \leq 0$$
$$(x - \mathbf{y}_2)^\top \mathbf{n}_2 \leq 0$$
$$(x - \mathbf{y}_3)^\top \mathbf{n}_3 \leq 0$$




Where $\mathbf{n}_i$ is a vector normal to the linear constraint.

# Constrained Optimization

A general constrained optimization problem can be written as

$$
\begin{aligned}
\min_{x \in \Omega} \quad & f(x) \\
\text{s.t.} \quad & h_1(x) = 0, \\
& \vdots, \\
& h_m(x) = 0, \\
& g_1(x) \leq 0, \\
& \vdots, \\
& g_p(x) \leq 0
\end{aligned}
$$

# Constrained Optimization

Letting

$$\mathbf{h} = (h_1 \dots h_m)^\top$$
$$\mathbf{g} = (g_1 \dots g_p)^\top,$$

we have

$$\begin{aligned} \min_{x \in \Omega} \quad & f(x) \\ \text{s.t.} \quad & \mathbf{h}(x) = 0, \\ & \mathbf{g}(x) \leq 0 \end{aligned}$$

Equality constraints are easier to deal with than inequality constraints.

We will first treat equality constraints, then inequality constraints.

Section 3

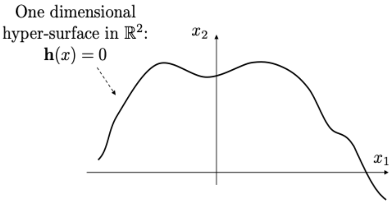
# Equality Constraints: Lagrange Multipliers

# Equality Constraints: Lagrange Multipliers

Several geometric insights help:

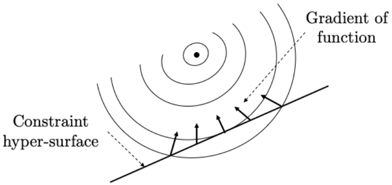
Insight #1

- 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

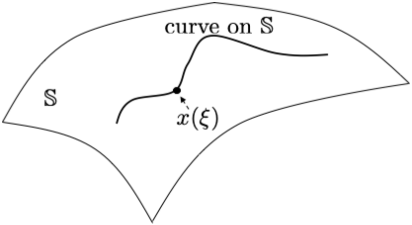



At a constrained minimum the gradient is orthogonal to the hypersurface.

# 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

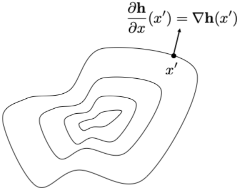
### Definition

The tangent plane 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.

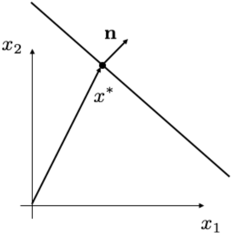
# Equality Constraints: Lagrange Multipliers

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)

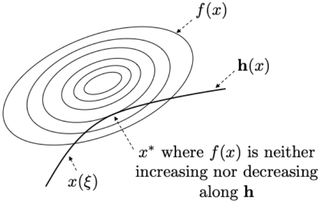
*Let* $x(\xi)$ *be a curve on* $\mathbf{h}(x) = 0$ *such that*

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

*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:

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

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

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

for all $\xi$ both positive and negative. Therefore

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

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. $\square$

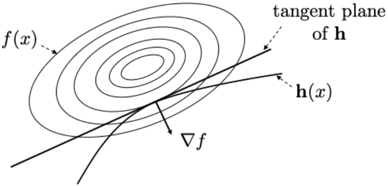
# Equality Constraints: Lagrange Multipliers

## Lemma (Moon Lemma 18.2)

*If $x^*$ is a regular point of $\mathbf{h}(x) = 0$ <u>and</u> 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 $\mu_i$ 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.

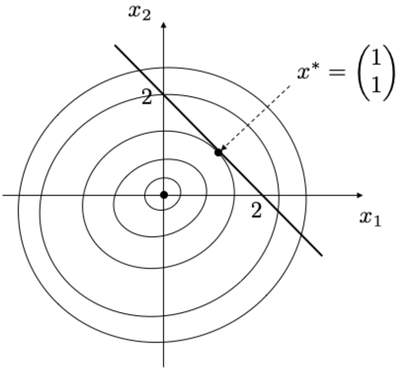
$$\frac{\partial L}{\partial x} = \nabla f(x^*) + \nabla h(x^*)\lambda^* = 0 \tag{1}$$

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

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

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

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

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

The necessary conditions are

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

where

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

# Lagrange Multipliers: Example 18.4.7 (Maximum Entropy)

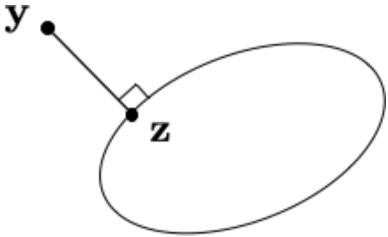
So $p_i$ must be constant for all $i$. The constant $\sum p_i = 1 \Rightarrow p_i = \frac{1}{n} \geq 0$, so $p_i$ satisfies the inequality constraint.

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:

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

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

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

Therefore, $\lambda$ 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:

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

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

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

# 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 4

## 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, $\lambda$ 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 $h$.

Section 5

# Inequality Constraints: Kuhn-Tucker Conditions

# Inequality Constraints

Lets first consider the problem with just inequality constraints, i.e.

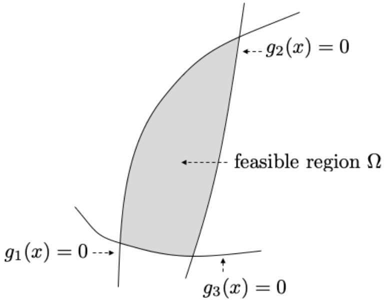
$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & \mathbf{g}(x) \leq 0 \end{array}$$

where $\mathbf{g}(x) \leq 0$ means that

$$\begin{pmatrix} g_1(x) \\ \vdots \\ g_q(x) \end{pmatrix} \leq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

i.e., element-wise.

For example, let $x \in \mathbb{R}^2$ and let $q = 3$.



# Inequality Constraints

Case I. If the local min is in the interior of $\Omega$, then clearly

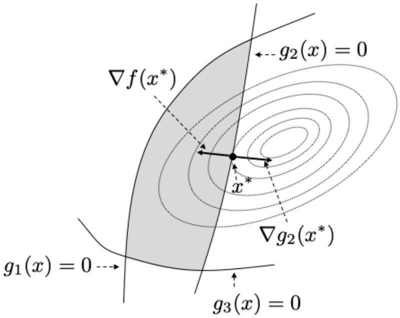
$$\nabla f(x^*) = 0$$

or

$$\nabla f(x^*) + 0 \cdot \nabla g_1(x^*) + 0 \cdot \nabla g_2(x^*) + 0 \cdot g_3(x^*) = 0.$$

# Inequality Constraints

Case II. The local minimum is on the boundary but not at a corner

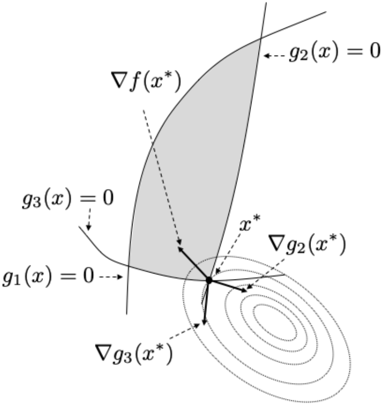



Since in this case $g_1$ is an equality constraint, we must have that $\nabla f(x^*) \parallel \nabla g_1(x^*)$. In fact, in this case the two vectors point in opposite directions! Therefore

$$\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + 0 \cdot \nabla g_2(x^*) + 0 \cdot g_3(x^*) = 0.$$

# Inequality Constraints

## Case III.




In this case, $\nabla f(x^*)$ is in the linear span of $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$ where the coefficients are negative. Therefore

$$\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) + 0 \cdot g_3(x^*) = 0$$

where $\mu_1 > 0$ and $\mu_2 > 0$.

# Inequality Constraints

In general, for inequality constraints at a local minimum $x^*$ we have that

1. $\nabla f(x^*) + \nabla \mathbf{g}(x^*)\mu = 0$
2. $\mathbf{g}(x^*)^\top \mu = 0$
3. $\mu \geq 0$

Conditions (1) and (3) together mean that $\nabla f(x^*)$ is contained in the (negative) linear span of $\{\nabla g_1(x^*), \ldots, \nabla g_q(x^*)\}$.

Condition (2): Note that if the constraint is active, i.e. $g_i(x^*) = 0$ then $\mu_i$ can be nonzero, but if $g_i$ is inactive, i.e. $g_i(x^*) < 0$ then $\mu_i$ must be zero to satisfy (2).

# Inequality Constraints

Now lets go back to the general constrained optimization problem:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & \mathbf{h}(x) = 0, \\ & \mathbf{g}(x) \leq 0 \end{aligned}$$

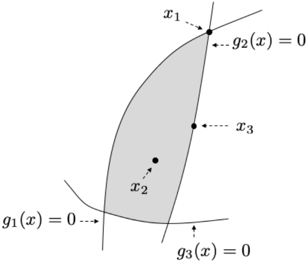
where $f : \mathbb{R}^n \to \mathbb{R}$, $h(x) : \mathbb{R}^n \to \mathbb{R}^p$, $g(x) : \mathbb{R}^n \to \mathbb{R}^q$.

### Definition

$x^*$ is a regular point if $\nabla h_i(x^*), i = 1, \ldots, p$ and $\nabla g_j(x^*)$ are linearly independent for all $j = 1, \ldots, q$ such that $g_j(x^*)$ is active.

# Inequality Constraints

For example, suppose that $\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$, and $\mathbf{g} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix}$.




Then $x^*$ is a regular point at:

- $x_1$ if $\{\nabla h_1(x_1), \nabla h_2(x_1), \nabla g_1(x_1), \nabla g_2(x_1)\}$ are linearly independent.
- $x_2$ if $\{\nabla h_1(x_2), \nabla h_2(x_2)\}$ are linearly independent.
- $x_3$ if $\{\nabla h_1(x_3), \nabla h_2(x_3), \nabla g_1(x_3)\}$ are linearly independent.

# Kuhn Tucker Conditions: Necessary Conditions

**Theorem (Moon Theorem 18.6)**

*Let $x^*$ be a regular local minimum, then $\exists \lambda \in \mathbb{R}^p$ (regular Lagrange multipliers), and $\exists \mu \in \mathbb{R}^q$, such that*

1. $\mu \geq 0$ *(element wise)*
2. $\mathbf{g}^\top(x^*)\mu = 0$
3. $\nabla f(x^*) + \nabla \mathbf{h}^\top(x^*)\lambda + \nabla \mathbf{g}^\top(x^*)\mu = 0.$

# Kuhn Tucker Conditions: Sufficient Conditions

Theorem (Moon 18.7)

*Suppose $f, g, h$ are in $C_2$. If there exist $\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^q$ such that at $x^*$*

1. $\mu \geq 0$
2. $\mathbf{g}^\top(x^*)\mu = 0$
3. $\nabla f(x^*) + \nabla \mathbf{h}^\top(x^*)\lambda + \nabla \mathbf{g}^\top(x^*)\mu = 0$
4. $p^\top(\nabla^2 f(x^*) + \sum_{k=1}^{p} \nabla^2 h_k(x^*)\lambda_k + \sum_{k=1}^{q} \nabla g_k(x^*)\mu_k)p > 0$

*for all $p$ in the tangent plane of the <u>active</u> constraints, then $x^*$ is a local constrained minimum.*

# Kuhn Tucker Conditions: Example 18.9.1

$$\begin{aligned} \min \quad & 3x_1^2 + 4x_2^2 + 6x_1x_2 - 8x_2 - 6x_1 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 9 \leq 0, \\ & 2x_1 - x_2 - 4 \leq 0 \end{aligned}$$

The necessary conditions are:

$$\begin{aligned} 6x_1 + 6x_2 - 6 + \mu_1(2x_1) + \mu_2(2) &= 0 \\ 8x_2 + 6x_1 - 8 + \mu_1(2x_2) + \mu_2(-1) &= 0 \\ \mu_1(x_1^2 + x_2^2 - 9) + \mu_2(2x_1 - x_2 - 4) &= 0 \\ \mu_1 \geq 0, \mu_2 &\geq 0 \end{aligned}$$

# Kuhn Tucker Conditions: Example 18.9.1

Lets try various combinations of active constraints:

Case I (Both inactive) i.e.
$\mu_1 = \mu_2$ 0
Therefore, must solve

$$6x_1 + 6x_2 - 6 = 0$$
$$8x_2 + 6x_1 - 8 = 0$$

i.e.,

$$\begin{pmatrix} 6 & 6 \\ 6 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \end{pmatrix}$$
$$\implies \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Check inequality constraints:

$$g_1(x) = 1 - 9 = -8 \leq 0$$
$$g_2(x) = -1 - 4 \leq 0$$

Therefore

$$x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \mu^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

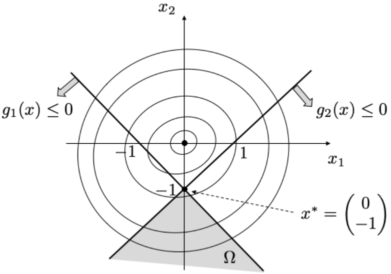
satisfies necessary conditions.
Sufficient condition:

$$\nabla^2 f = \begin{pmatrix} 6 & 6 \\ 6 & 8 \end{pmatrix} > 0$$

implies local minimum.

# Kuhn Tucker Conditions: Example

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 + x_2 + 1 \leq 0, \\ & -x_1 + x_2 + 1 \leq 0 \end{aligned}$$



# Kuhn Tucker Conditions: Example

The necessary conditions are:

$$\begin{aligned} 2x_1 + \mu_1 - \mu_2 &= 0 \\ 2x_2 + \mu_1 + \mu_2 &= 0 \\ \mu_1(x_1 + x_2 + 1) + \mu_2(-x_1 + x_2 + 1) &= 0 \\ \mu_1 \geq 0, \mu_2 &\geq 0 \end{aligned}$$

Try various combinations of active constraints Case 1: (Both inactive)

$$\begin{aligned} 2x_1 &= 0 \\ 2x_2 &= 0 \end{aligned}$$

$$\implies x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

However, both constraints are violated since

$$\begin{aligned} g_1^*(x^*) &= 1 \geq 0 \\ g_2(x^*) &= 1 \geq 0. \end{aligned}$$

# Kuhn Tucker Conditions: Example

## Case 2: $g_1$-active, $g_2$-inactive

$$\begin{aligned}
2x_1 + \mu_1 &= 0 \quad \Longrightarrow \; x_1 = -\frac{1}{2}\mu_1 \\
2x_2 + \mu_1 &= 0 \quad \Longrightarrow \; x_2 = -\frac{1}{2}\mu_1 \\
\mu_1(x_1 + x_2 + 1) &= 0 \\
\mu_1 &> 0
\end{aligned}$$

Last two equations imply that

$$\mu_1(-\frac{1}{2}\mu_1 - \frac{1}{2}\mu_1 + 1) = -\mu_1^2 + \mu_1 = \mu_1(1 - \mu_1) = 0.$$

Solving for $\mu_1$ gives $\mu_1 = 0$ or $\boxed{\mu_1 = 1}$. Therefore

$$x^* = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$$

# Kuhn Tucker Conditions: Example

Checking constraints:

$$g_1(x^*) = -\frac{1}{2} - \frac{1}{2} + 1 = 0 \leq 0 \qquad \text{ok}$$

$$g_2(x^*) = \frac{1}{2} - \frac{1}{2} + 1 = 1 \geq 0 \qquad \text{no}$$

Case 3: $g_1$-inactive, $g_2$-active Similar results to Case 2.

Case 4: Both active

$$\begin{aligned} &\mu_1(\frac{1}{2}\mu_2 - \frac{1}{2}\mu_1 - \frac{1}{2}\mu_2 - \frac{1}{2}\mu_1 + 1) \\ &\quad + \mu_2(-\frac{1}{2}\mu_2 + \frac{1}{2}\mu_1 - \frac{1}{2}\mu_1 - \frac{1}{2}\mu_2 + 1) = 0 \\ \implies &\mu_1(1 - \mu_1) + \mu_2(1 - \mu_2) = 0 \end{aligned}$$

# Kuhn Tucker Conditions: Example

A positive solution is

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} > 0$$

which gives

$$x^* = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

Constraints can be verified to be satisfied.

Sufficient condition:

$$\nabla^2 f + \nabla^2 g \mu = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} 1 + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} 1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} > 0$$

Therefore $x^*$ is a local minimum.