C477: Computational Optimisation Constrained Optimisation – Optimality Conditions

Panos Parpas
Department of Computing
Imperial College London

www.doc.ic.ac.uk/∼pp500
p.parpas@imperial.ac.uk

Outline

1. **Equality Constrained Optimisation**
    - Regular Points
    - First Order Necessary Conditions (Lagrange's Theorem)
    - Simple Proof of Lagrange's Theorem
    - Lagrange Multipliers – Interpretation
    - Second Order Necessary Conditions
    - Second Order Sufficient Conditions
    - Examples: Portfolio Optimisation, Optimal Control
2. **Problems with Inequality Constraints**
    - Definitions
    - Karush-Kuhn-Tucker Conditions
    - Second Order Conditions

Additional material:

- Chapter 19, & 20 in *An Introduction to Optimization*, Chong & Zhak, Third Edition.
- Chapter 11, in *Linear and Nonlinear Programming*, Luenberger & Ye, Third Edition.

Outline

1. Equality Constrained Optimisation

- Regular Points
- First Order Necessary Conditions (Lagrange's Theorem)
- Simple Proof of Lagrange's Theorem
- Lagrange Multipliers – Interpretation
- Second Order Necessary Conditions
- Second Order Sufficient Conditions
- Examples: Portfolio Optimisation, Optimal Control

2. Problems with Inequality Constraints

- Definitions
- Karush-Kuhn-Tucker Conditions
- Second Order Conditions

Equality Constrained Optimisation

Problem Formulation

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{s.t } \boldsymbol{h}(\boldsymbol{x}) = 0,$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $h = [h_1, \ldots, h_m]^T$, and $m \leq n$.

Definition (Regular points)

A point $\boldsymbol{x}^*$ satisfying the constraints i.e. $\boldsymbol{h}(\boldsymbol{x}^*) = 0$ is called **regular** if the **gradient vectors** $\{\nabla h_1(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)\}$ are **linearly independent**.

Equality Constrained Optimisation

Problem Formulation

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{s.t } \boldsymbol{h}(\boldsymbol{x}) = 0,$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $h = [h_1, \ldots, h_m]^T$, and $m \leq n$.

Definition (Regular points)

A point $\boldsymbol{x}^*$ satisfying the constraints i.e. $\boldsymbol{h}(\boldsymbol{x}^*) = 0$ is called **regular** if the **gradient vectors** $\{\nabla h_1(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)\}$ are **linearly independent**.

Lagrange's Theorem

Theorem (First Order Necessary Conditions)

Let $\boldsymbol{x}^*$ be a local minimiser (or maximiser) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $\boldsymbol{h}(\boldsymbol{x}) = 0$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \le n$. Assume that $\boldsymbol{x}^*$ is a regular point. Then there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that,

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

Terminology

- The vector λ is called the **Lagrange multiplier vector.**
- The **Lagrange condition** consists of the two equations,

$$\begin{aligned} \boldsymbol{h}(\boldsymbol{x}^*) &= 0 \\ \nabla f(\boldsymbol{x}^*) + \nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* &= 0. \end{aligned}$$

- The **Lagrangian** function associated with the optimisation problem is,

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

- Note that $\nabla L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$ is the Lagrange condition.

Lagrange's Theorem

Theorem (First Order Necessary Conditions)

Let $\boldsymbol{x}^*$ be a local minimiser (or maximiser) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $\boldsymbol{h}(\boldsymbol{x}) = 0$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$. Assume that $\boldsymbol{x}^*$ is a regular point. Then there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that,

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

Terminology

- The vector $\boldsymbol{\lambda}$ is called the **Lagrange multiplier vector.**
- The **Lagrange condition** consists of the two equations,

$$\begin{aligned} \boldsymbol{h}(\boldsymbol{x}^*) &= 0 \\ \nabla f(x^*) + \nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* &= 0. \end{aligned}$$

- The **Lagrangian** function associated with the optimisation problem is,

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

- Note that $\nabla L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$ is the Lagrange condition.

Example 1

$$\min_x \; x_1^2 + x_2^2$$
$$\text{s.t. } x_1^2 + 2x_2^2 - 1 = 0.$$

Find all the points that satisfy the First Order Necessary conditions and identify points that could potentially be local minimisers or maximisers.

Example 1

The Lagrangian is given by $L(x, \lambda) = x_1^2 + x_2^2 + \lambda(x_1^2 + 2x_2^2 - 1)$
The optimality condition $\nabla L(x, \lambda) = 0$ leads to the following set of equations,

$$\begin{aligned} 2x_1 + 2\lambda x_1 &= 0 \\ 2x_2 + 4\lambda x_2 &= 0 \\ x_1^2 + 2x_2^2 &= 1 \end{aligned}$$

Clearly all feasible points are regular (why). From the first equation above either $x_1 = 0$ or $\lambda = -1$. If $x = 0$, the second and third equations imply that $\lambda = -\frac{1}{2}$ and $x_2 = \pm 1/\sqrt{2}$. If $\lambda = -1$ then the second and third equations imply that $x_1 = \pm 1$ and $x_2 = 0$. Thus the following points satisfy the Lagrange conditions,

$$x^{(1)} = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad x^{(2)} = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \quad x^{(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad x^{(4)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Since $f(x^{(1)}) = f(x^{(2)}) = 1/2$, then $x^{(1)}$ and $x^{(2)}$ <u>could be</u> minimisers.
Since $f(x^{(3)}) = f(x^{(4)}) = 1$, then $x^{(3)}$ and $x^{(4)}$ <u>could be</u> maximisers.

Example 2: Computing the maximum eigenvalue

$$\max_{\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$$
$$\text{s.t. } \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} = 1.$$

Where $\boldsymbol{Q}$ and $\boldsymbol{P}$ are symmetric positive definite matrices in $\mathbb{R}^{n \times n}$. Find all the points that satisfy the First Order Necessary Conditions and identify points that could potentially be local minimisers or maximisers.

Example 2: Computing the maximum eigenvalue

The Lagrangian for this problem is,

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \lambda(1 - \boldsymbol{x}^T P \boldsymbol{x}).$$

The Lagrange condition is,

$$\begin{aligned} 2\boldsymbol{Q}\boldsymbol{x} - 2\boldsymbol{\lambda}\boldsymbol{P}\boldsymbol{x} &= 0 \\ 1 - \boldsymbol{x}^T \boldsymbol{P}\boldsymbol{x} &= 0 \end{aligned}$$

Since $\boldsymbol{P} \succ 0$, $\boldsymbol{P}^{-1}$ exists and the first equation implies that $\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{x} = \lambda \boldsymbol{x}$. The last relationship combined with the constraint gives,

$$\lambda = \boldsymbol{x}^{*T} \boldsymbol{Q} x^*.$$

Therefore λ^* is the maximum eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$ and the associated eigenvector is $\boldsymbol{x}^*$.

Example 3: The regularity condition is needed

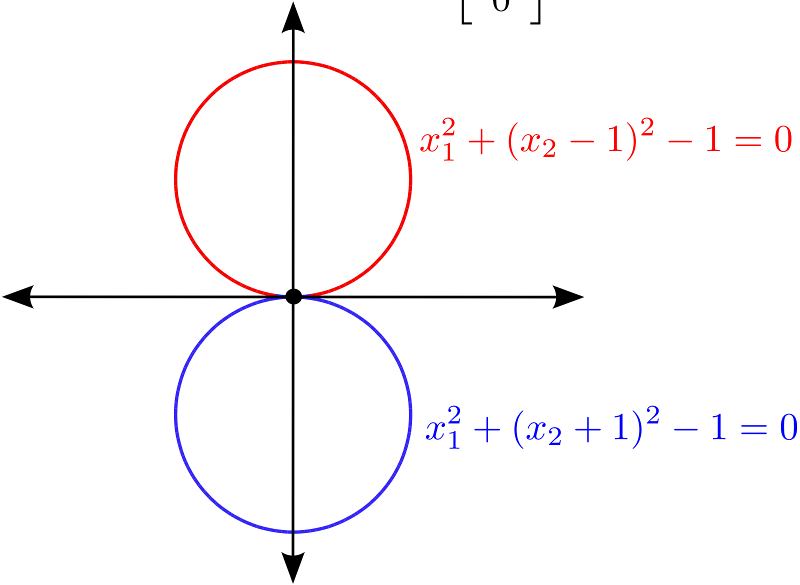
$$\min_{x_1,x_2} x_1$$

$$\text{s.t.}\quad h_1(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0$$

$$h_2(x) = x_1^2 + (x_2 + 1)^2 - 1 = 0$$

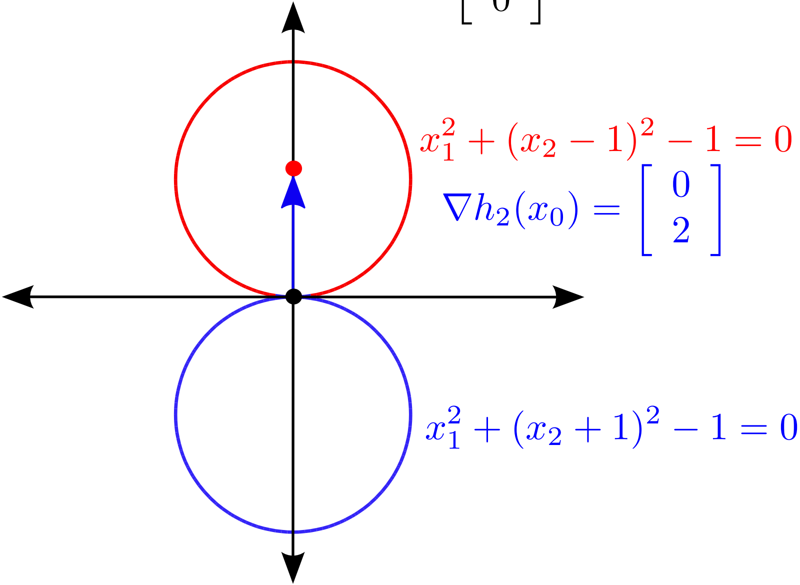
Example 3: The regularity condition is needed

Only feasible point: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$



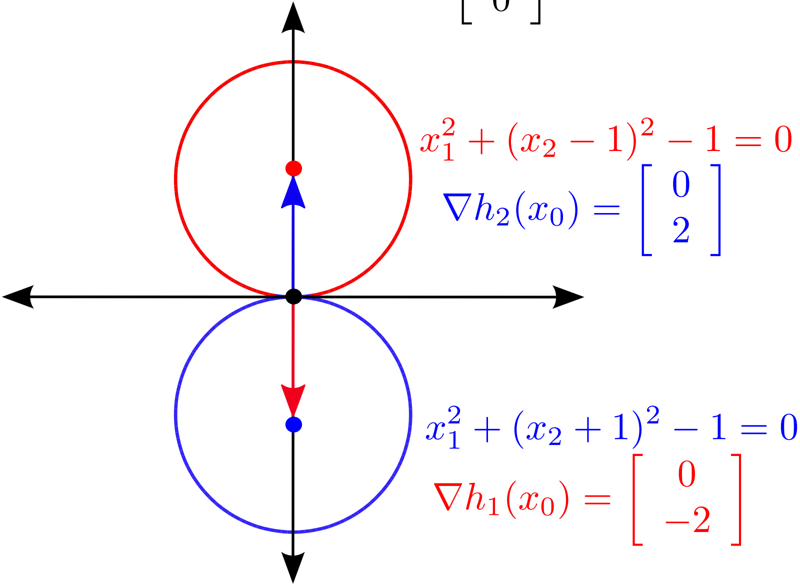
Example 3: The regularity condition is needed

Only feasible point: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$



Example 3: The regularity condition is needed

Only feasible point: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$



Example 3: The regularity condition is needed

Clearly $x_0 = (0,0)^T$ is the global optimum but,

$$\nabla f(x_0) + \lambda_1 \nabla h_1(x_0) + \lambda_2 \nabla h_2(x_0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 0 \\ -2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 2 \end{bmatrix} = 0$$

Has no solution.

The regularity condition may not be needed!

$$
\begin{aligned}
&\min_{x_1, x_2} \; x_2 \\
\text{s.t.} \quad & h_1(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0 \\
& h_2(x) = x_1^2 + (x_2 + 1)^2 - 1 = 0
\end{aligned}
$$

The regularity condition may not be needed!

$$\begin{aligned} &\min_{x_1,x_2} \; x_2 \\ \text{s.t.} \quad & h_1(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0 \\ & h_2(x) = x_1^2 + (x_2 + 1)^2 - 1 = 0 \end{aligned}$$

Again $\boldsymbol{x}_0 = (0,0)^T$ is the only feasible solution optimum & Lagrange theorem is satisfied at this point.

$$\boldsymbol{\nabla f}(\boldsymbol{x}_0) + \lambda_1 \nabla \boldsymbol{h}_1(\boldsymbol{x}_0) + \lambda_2 \boldsymbol{\nabla h}_2(\boldsymbol{x}_0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \lambda_1 \begin{bmatrix} 0 \\ -2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 2 \end{bmatrix} = 0$$

Has many solutions.

Lagrange's Theorem – Simple Proof

Theorem (First Order Necessary Conditions – $n = 2$, $m = 1$)

Let $\boldsymbol{x}^*$ be a local minimiser (or maximiser) of $f : \mathbb{R}^2 \to \mathbb{R}$ subject to $h(\boldsymbol{x}) = 0$, $h : \mathbb{R}^n \to \mathbb{R}$. Then there exists a scalar $\lambda^* \in \mathbb{R}$ such that,

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

Lagrange's Theorem – Simple Proof

Proof.

□

Lagrange's Theorem – Simple Proof

Proof.

Note that the condition,

$$\boldsymbol{\nabla f}(x^*) = -\boldsymbol{\lambda}\boldsymbol{\nabla} h(\boldsymbol{x}^*).$$

Means that $\nabla f(\boldsymbol{x}^*)$ is parallel to $\nabla h(\boldsymbol{x}^*)$. So if we can show that $\nabla f(\boldsymbol{x}^*)$ and $\nabla h(\boldsymbol{x}^*)$ are orthogonal to the same vector then we know that they will be parallel. The basic idea behind the proof is to construct such a vector.
Consider the set,

$$L = \{\boldsymbol{x} \mid h(\boldsymbol{x}) = 0\}.$$

Then we parameterise the set L such that,

$$\boldsymbol{x}(t) \in L,\ \boldsymbol{x}(t^*) = \boldsymbol{x}^*, \text{and } \frac{d\boldsymbol{x}}{dt}(t^*) \neq 0, \text{ where } t, t^* \in (a, b).$$

We will show that $\langle \frac{d\boldsymbol{x}}{dt}(t^*), \boldsymbol{\nabla h}(\boldsymbol{x}^*)\rangle = 0$ and that $\langle \frac{d\boldsymbol{x}}{dt}(t^*), \boldsymbol{\nabla f}(\boldsymbol{x}^*)\rangle = 0$
1. Since $\boldsymbol{x}(t)$ is constant in the curve L (i.e. $h(\boldsymbol{x}(t)) = 0$ if $t \in (a, b)$) then,

$$\frac{dh}{dt}h(\boldsymbol{x}(t^*)) = \left\langle \frac{d\boldsymbol{x}}{dt}(t^*), \boldsymbol{\nabla h}(\boldsymbol{x}^*) \right\rangle = 0$$

2. Consider the composite function $\phi(t) = f(\boldsymbol{x}(t))$. Since $\boldsymbol{x}^*$ is a local minimum of f then t^* is a local minimum of $\phi(t)$. Therefore the first order condition,

$$\frac{d\phi(t^*)}{dt} = 0$$

is satisfied. But,

$$0 = \frac{d\phi(t^*)}{dt} = \left\langle \frac{d\boldsymbol{x}}{dt}(t^*), \boldsymbol{\nabla f}(\boldsymbol{x}^*) \right\rangle.$$

Therefore $\boldsymbol{\nabla f}(\boldsymbol{x}^*)$ is parallel to $\boldsymbol{\nabla} h(\boldsymbol{x}^*)$, and we must have $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = -\lambda\boldsymbol{\nabla} h(\boldsymbol{x}^*)$ for some λ.

Simple Proof – Example

We will illustrate the main idea behind the proof using the problem,

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

Simple Proof – Example

Note that $x^* = [\frac{1}{\sqrt{2}} \; \frac{1}{2}]^\top$ is a local extremum with $\lambda = -1$. We can parameterise the constraint with the following

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} t \\ 1 - t^2 \end{bmatrix}$$

Note that $t^* = \frac{1}{\sqrt{2}}$ since

$$x(t^*) = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{bmatrix}.$$

Note that $\frac{d}{dt}x(t^*) \neq 0$.

$$\begin{aligned} \left\langle \frac{dx}{dt}(t^*), \nabla f(x^*) \right\rangle &= [2x_1(t^*) \; 2x_2(t^*)] \begin{bmatrix} 1 \\ -2t^* \end{bmatrix} \\ &= \left[\frac{2}{\sqrt{2}} \; 1 \right] \begin{bmatrix} 1 \\ -2/\sqrt{2} \end{bmatrix} = 0 \end{aligned}$$

$$\begin{aligned} \left\langle \frac{dx}{dt}(t^*), \nabla h(x^*) \right\rangle &= [2x_1(t^*) \; 1] \begin{bmatrix} 1 \\ -2t^* \end{bmatrix} \\ &= \left[\frac{2}{\sqrt{2}} \; 1 \right] \begin{bmatrix} 1 \\ -2/\sqrt{2} \end{bmatrix} = 0 \end{aligned}$$

Hence $\nabla h(x^*)$ is parallel to $\nabla f(x^*)$.

Lagrange Multipliers – Interpretation

Suppose we change the rhs of the i^{th} constraint by r

$$\begin{aligned} &\min f(\boldsymbol{x}) \\ \text{s.t } \; &h_1(\boldsymbol{x}) = 0 \\ &\vdots \\ &h_i(\boldsymbol{x}) = r \\ &\vdots \\ &h_m(\boldsymbol{x}) = 0 \end{aligned}$$

Suppose that the optimal solution is $\boldsymbol{x}(r)$ then,

$$\frac{df}{dr}(\boldsymbol{x}(r)) = -\lambda_i$$

where λ_i denotes the Lagrange multiplier of the i^{th} constraint. i.e.
The i^{th} Lagrange multiplier is a measure of the change in the optimal objective function value if the i^{th} constraint is perturbed.

Lagrange Multipliers – Interpretation

Suppose we change the rhs of the i^{th} constraint by r

$$
\begin{aligned}
&\min f(\boldsymbol{x}) \\
\text{s.t } \; & h_1(\boldsymbol{x}) = 0 \\
& \vdots \\
& h_i(\boldsymbol{x}) = r \\
& \vdots \\
& h_m(\boldsymbol{x}) = 0
\end{aligned}
$$

Suppose that the optimal solution is $\boldsymbol{x}(r)$ then,

$$\frac{df}{dr}(\boldsymbol{x}(r)) = -\lambda_i$$

where λ_i denotes the Lagrange multiplier of the i^{th} constraint. i.e. **The i^{th} Lagrange multiplier is a measure of the change in the optimal objective function value if the i^{th} constraint is perturbed.**

Lagrange Multipliers – Interpretation, Proof

Note that at the solution,

$$
\begin{aligned}
h_1(x(r)) &= 0 \\
&\vdots \\
h_i(x(r)) &= r \\
&\vdots \\
h_m(x(r)) &= 0
\end{aligned}
$$

The above implies that,

$$
\frac{dh_j(\boldsymbol{x}(r))}{dr} = \begin{cases} = 1 \text{ if } j = i \\ = 0 \text{ otherwise} \end{cases}
$$

Using the chain rule and the optimality conditions together with the result above we obtain,

$$
\begin{aligned}
\frac{df}{dr}(\boldsymbol{x}(r)) = \boldsymbol{\nabla} f(\boldsymbol{x}(r))^T \frac{d\boldsymbol{x}}{dr} &= \left(-\sum_{j=1}^{m} \lambda_j \boldsymbol{\nabla} \boldsymbol{h}_j(\boldsymbol{x}(r))^T \frac{d\boldsymbol{x}}{dr} \right) \\
&= \left(-\sum_{j=1}^{m} \lambda_j \frac{d\boldsymbol{h}_j(\boldsymbol{x}(r))}{dr} \right) \\
&= -\lambda_i .
\end{aligned}
$$

Example: `runalg.m`, `objfun.m`

$$f(w) = \min_{x_1,x_2} \; -x_1^{2/3} x_2^{1/3};$$
$$\text{s.t } \; p_1 x_1 + p_2 x_2 = w$$

$$f(20) = -0.8076, \;\; \lambda = 0.0404$$

i.e. a change of w by small amount will increase the objective function

$$f(20.1) = -0.8116, \;\; \text{note that } \; \frac{f(20.1) - f(20)}{0.1} \approx -0.0404$$

Example: `runalg.m`, `objfun.m`

$$f(w) = \min_{x_1, x_2} \; -x_1^{2/3} x_2^{1/3};$$
$$\text{s.t } \; p_1 x_1 + p_2 x_2 = w$$

$$f(20) = -0.8076, \; \lambda = 0.0404$$

i.e. a change of w by small amount will increase the objective function

$$f(20.1) = -0.8116, \text{ note that } \frac{f(20.1) - f(20)}{0.1} \approx -0.0404$$

Reminder: Second Order Necessary Conditions

Theorem (Second order necessary conditions)

Let $\Omega = \{\boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = 0\}$*, and* $f \in \mathcal{C}^2$*,* $\boldsymbol{x}^*$ *be a local minimiser of* f *over* Ω *and* $\boldsymbol{d}$ *be a feasible direction at* x^**. If* $\boldsymbol{d}^T \nabla \boldsymbol{f}(x^*) = 0$*, then*

$$\boldsymbol{d}^T \boldsymbol{\nabla}^2 \boldsymbol{f}(x^*) \boldsymbol{d} \geq 0,$$

where $\boldsymbol{\nabla}^2 \boldsymbol{f}$ *is the Hessian matrix of* f*.*

Second Order Necessary Conditions

Theorem (Second order necessary conditions)

Let $\Omega = \{\boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = 0\}$, and $f \in \mathcal{C}^2$, $\boldsymbol{x}^* \in \Omega$ be a local minimiser of $f : \mathbb{R}^n \to \mathbb{R}$ over Ω, and where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ with $m \leq n$ is also in $\mathcal{C}^2$. Suppose that is regular. Then there exists a $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that,

1. $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{\nabla}_x \boldsymbol{f}(\boldsymbol{x}^*) + \nabla_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* = 0$
2. $\boldsymbol{d}^T \boldsymbol{\nabla}^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{d} \geq 0$ *for all* $\boldsymbol{d}$ *such that* $\boldsymbol{\nabla h}(\boldsymbol{x}^*)^T \boldsymbol{d} = 0$

Where $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(x)$ denotes the Lagrangian associated with the problem.

Reminder: Second Order Sufficient Conditions

Theorem (Second Order Sufficient Conditions)

Suppose that $f \in \mathcal{C}^2$ in a region where $\boldsymbol{x}^*$ is an interior point. Suppose that,

1. $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = 0$.
2. $\boldsymbol{\nabla}^2 \boldsymbol{f}(x^*) \succ 0$. *(i.e. the Hessian is positive definite at the point $\boldsymbol{x}^*$)*

Then $\boldsymbol{x}^*$ is a strict local minimiser of f.

Second Order Sufficient Conditions

Theorem (Second Order Sufficient Conditions)

Let $\Omega = \{\boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = 0\}$*, and* $f \in \mathcal{C}^2$*, and where* $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ *with* $m \leq n$ *is also in* $\mathcal{C}^2$*. Suppose that there exists an* $\boldsymbol{x}^* \in \mathbb{R}^n$ *and* $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *such that,*

1. $\boldsymbol{\nabla}_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \nabla_x \boldsymbol{f}(\boldsymbol{x}^*) + \nabla_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* = 0$
2. $\boldsymbol{d}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{d} > 0$ *for all* $\boldsymbol{d}$ *such that* $\boldsymbol{\nabla} \boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{d} = 0$

Then $\boldsymbol{x}^*$ *is a strict local minimiser of* f *over* Ω

Example: Second Order Sufficient Conditions

Consider the problem

$$\min_{x} \; \|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2$$
$$\text{s.t} \;\; \|\boldsymbol{x}\|_2^2 = 9,$$

where $\boldsymbol{x}_0 = [1, \sqrt{3}]^\top$.

(a) Find all the points satisfying the Lagrange condition for the problem.

(b) Using the second order conditions, determine whether or not each of the points in part (a) is a local minimiser.

Example: Second Order Sufficient Conditions

(a) **Find all the points satisfying the Lagrange condition for the problem.**

Example: Second Order Sufficient Conditions

(a) **Find all the points satisfying the Lagrange condition for the problem.**
The Lagrange condition for this problem is:

$$\begin{aligned} 2(x - x_0) + 2\lambda x &= 0 \\ \|x\|^2 &= 9 \end{aligned}$$

Note that the first equation implies that $(1+\lambda)\boldsymbol{x} = \boldsymbol{x}_0$. Therefore $(1+\lambda)^2\|x\|_2^2 = \|x_0\|_2^2$. Therefore $(1+\lambda)^2 = 4/9$. So $\lambda = -1 \pm 2/3$.
If $\lambda = -1 + 2/3 = -1/3$ then $\boldsymbol{x} = \frac{3}{2}[1, \sqrt{3}]^\top$.
If $\lambda = -1 - 2/3 = -5/3$ then $(1+\lambda)\boldsymbol{x} = \boldsymbol{x}_0$ means that $\boldsymbol{x} = -\frac{3}{2}[1, \sqrt{3}]^\top$

Example: Second Order Sufficient Conditions

(b) **Using the second order conditions, determine whether or not each of the points in part (a) is a local minimiser.**

Example: Second Order Sufficient Conditions

(b) **Using the second order conditions, determine whether or not each of the points in part (a) is a local minimiser.**
The Hessian of the Lagrangian is $\nabla^2\mathcal{L}(\boldsymbol{x}, \lambda) = 2(1+\lambda)\boldsymbol{I}$ (where $\boldsymbol{I}$ is the two by two identity matrix). To apply the SONC we first need to check regularity. The gradient of $h(\boldsymbol{x})$ is just $2\boldsymbol{x}$ and it is clearly linearly independent at both points.
If $\lambda = -1/3$ then the SONC does hold. This implies that $\boldsymbol{x} = \frac{3}{2}[1, \sqrt{3}]^\top$ is a strict local minimiser. For $\boldsymbol{x} = -\frac{3}{2}[1, \sqrt{3}]^\top$, $\lambda = -5/3$ and the SONC does not hold.

Example: Mean Variance Optimisation

The mean variance optimisation model attempts to capture the tradeoffs between risk and reward in investments.

- Proposed by H. Markowitz in 1952.
- Cornerstone of Modern Portfolio Theory
- Performance measured by expected returns
- Measures risk by variance of portfolio
- Shared with Miller & Sharpe the Nobel Memorial Prize in Economic Sciences (1990)

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1990
Harry M. Markowitz, Merton H. Miller, William F. Sharpe

- The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1990
- Nobel Prize Award Ceremony
- Harry M. Markowitz
- Merton H. Miller
- William F. Sharpe

Autobiography

I was born in Chicago in 1927, the only child of Morris and Mildred Markowitz who owned a small grocery store. We lived in a nice apartment, always had enough to eat, and I had my own room. I never was aware of the Great Depression.

Growing up, I enjoyed baseball and tag football in the nearby empty lot or the park a few blocks away, and playing the violin in the high school orchestra. I also enjoyed reading. At first, my reading material consisted of comic books and adventure magazines, such as *The Shadow*, in addition to school assignments. In late grammar school and throughout high school I enjoyed popular accounts of physics and astronomy. In high school I also began to read original works of serious philosophers. I was particularly struck by David Hume's argument that, though we release a ball a thousand times, and each time, it falls to the floor, we do not have a necessary proof that it will fall the thousand-and-first time. I also read *The Origin of Species* and was moved by Darwin's marshalling of facts and careful consideration of possible objections.

Example: Mean Variance Optimisation

- n assets (stocks) and a single decision period.
- The investment problem consists of deciding in which assets to invest in today, in order to minimize risk for a fixed level of returns at the end of the planning horizon.
- Returns are given as the expected value (mean) of the portfolio.
- Risk is measured by the variance of the portfolio.
- The classical model aims to minimise variance (risk), subject to meeting a performance constraint.

Example: Mean Variance Optimisation

- r_i denotes the (random) return of asset i over the investment period.
- Assume that $\boldsymbol{r} = [r_1, \ldots, r_n]$ is a random variable with mean $\bar{r} \in \mathbb{R}^n$, and a covariance matrix of $\Sigma \in \mathbb{R}^{n \times n}$.
- x_i denotes the amount invested in asset i

Remark: When $x_i > 0$ we are long the asset (i.e. we buy the asset), and $x_i < 0$ means we are short the asset. Short selling means that we can sell the asset today, and we are obliged to buy it in the end of the planing horizon.
Being long means that we want the asset price to increase, being short means we gain if the price falls.

Example: Mean Variance Optimisation

An investor has \$1000 to invest in three stocks, and would like a return on investment of r_m. Short selling is allowed.
The return vector $\bar{r}$ is given by:

$$\bar{r} = [0.14\ 0.11\ 0.10]$$

The covariance matrix of the returns is given by,

$$\Sigma = \begin{bmatrix} 0.2 & 0.05 & 0.02 \\ 0.05 & 0.08 & 0.03 \\ 0.02 & 0.03 & 0.18 \end{bmatrix}.$$

Formulate the Mean Variance Optimisation problem.

Example: Mean Variance Optimisation

Since we have a budget of \$1000, the first constraint is the budget constraint,

$$x_1 + x_2 + x_3 = 1000.$$

The performance constraint is given by:

$$\frac{x_1\bar{r}_1 + x_2\bar{r}_2 + x_3\bar{r}_3}{1000} \geq 0.12$$

Since short selling is allowed there are no further constraints

The objective is to minimise risk. In this framework risk is measured by variance which is given by,

$$\text{var}(x_1r_1 + x_2r_2 + x_3r_3) = [x_1\ x_2\ x_3] \begin{bmatrix} 0.2 & 0.05 & 0.02 \\ 0.05 & 0.08 & 0.03 \\ 0.02 & 0.03 & 0.18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

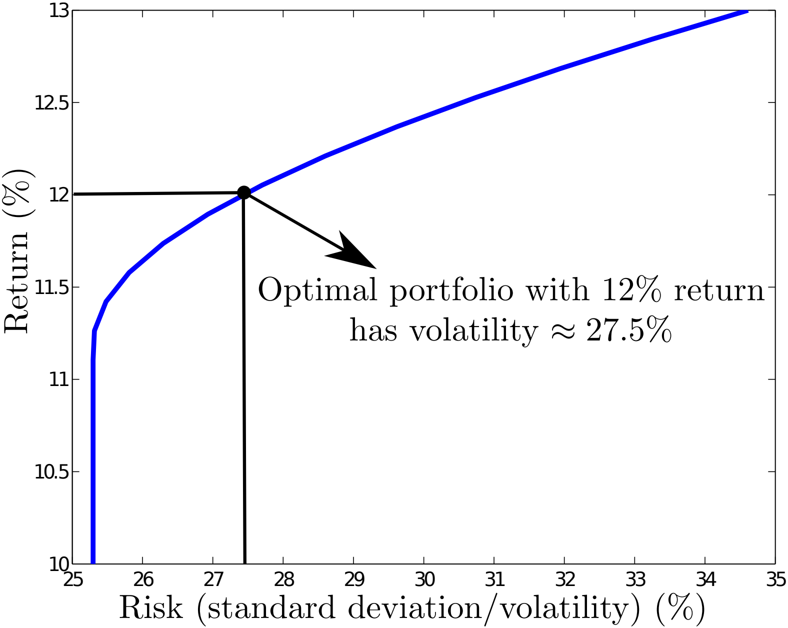
Example: Mean Variance Optimisation

The final model is given by

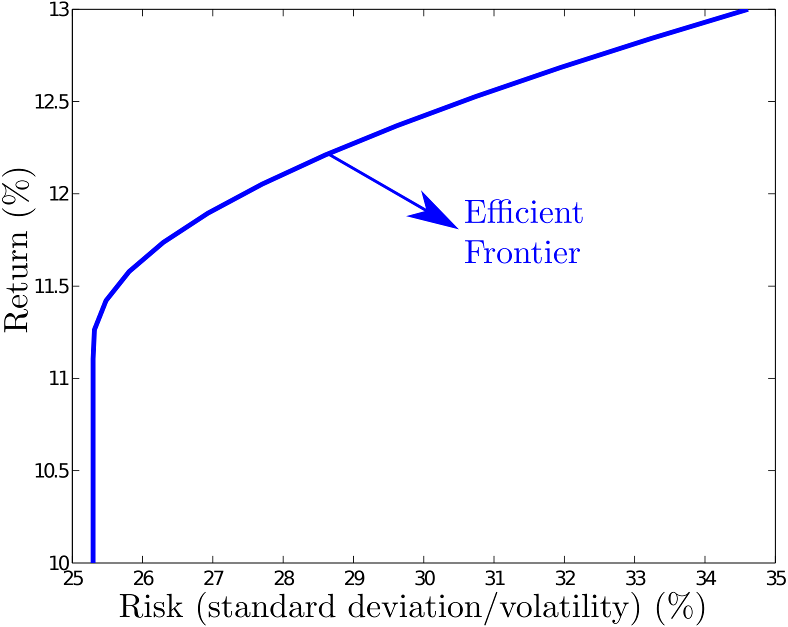
$$\begin{aligned} \min_{x_1,x_2,x_3} \; & \mathsf{var}(x_1 r_1 + x_2 r_2 + x_3 r_3) \\ & x_1 + x_2 + x_3 = 1000 \\ & \frac{x_1 \bar{r}_1 + x_2 \bar{r}_2 + x_3 \bar{r}_3}{1000} \geq 0.12 \end{aligned}$$

By varying the return on investment we obtain the curve in the next slide which is known as the efficient frontier. It is always suboptimal to invest in anything other than the portfolios that make up the efficient frontier (since the efficient portfolio achieves the same return with less risk).

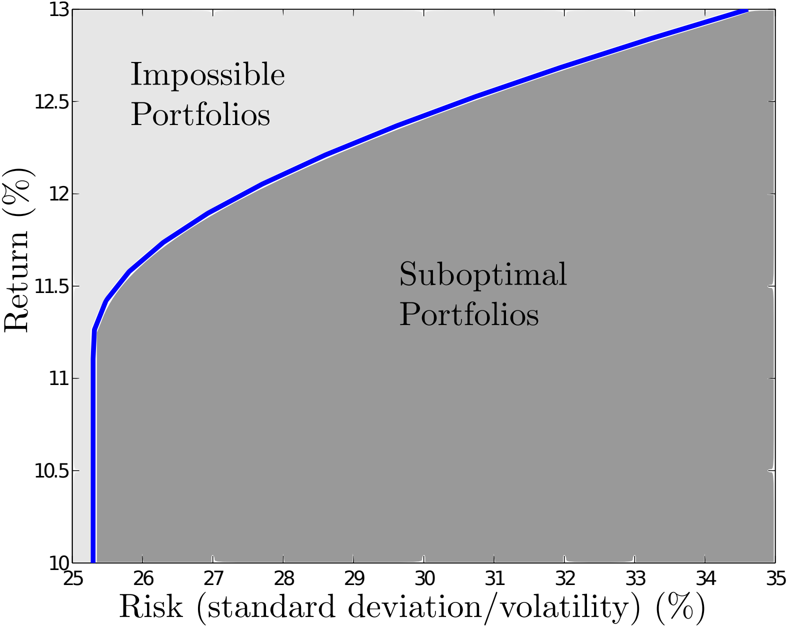
Example: Mean Variance Efficient Frontier



Example: Mean Variance Efficient Frontier



Example: Mean Variance Efficient Frontier

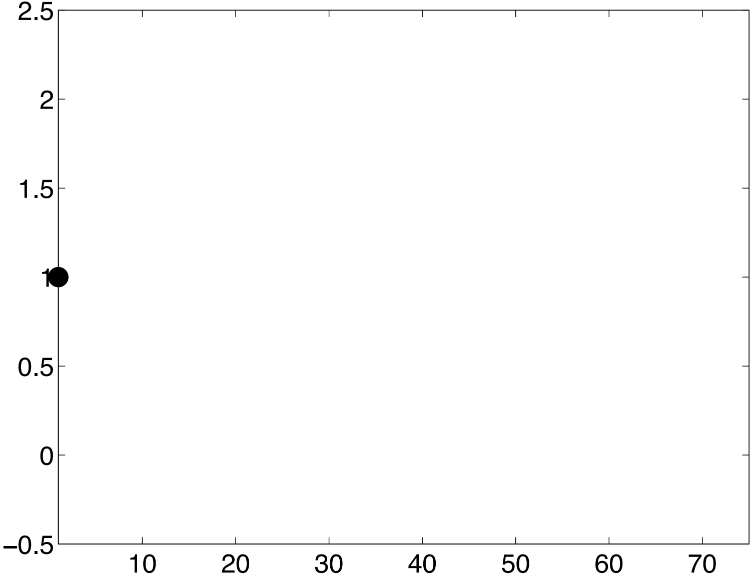


Example: Linear Quadratic Control

Uncontrolled Linear System dynamics

$$x_k = a_{k-1} x_{k-1}, \quad k \geq 1$$

With x_0 given.



2.5
2
1.5
1
0.5
0
−0.5
10
20
30
40
50
60
70

Example: Linear Quadratic Control

Uncontrolled Linear System dynamics

$$x_k = a_{k-1}x_{k-1}, \quad k = 1, \ldots, N$$

With x_0 given.

Aim: Keep x as close to 0 as possible, so we introduce a control u

Controlled Dynamics

$$x_k = a_{k-1}x_{k-1} + b_k u_k, \quad k = 1, \ldots, N$$

Example: Linear Quadratic Control

Uncontrolled Linear System dynamics

$$x_k = a_{k-1} x_{k-1}, \quad k = 1, \ldots, N$$

With x_0 given.

Aim: Keep x as close to 0 as possible, so we introduce a control u

Controlled Dynamics

$$x_k = a_{k-1} x_{k-1} + b_k u_k, \quad k = 1, \ldots, N$$

Example: Linear Quadratic Control

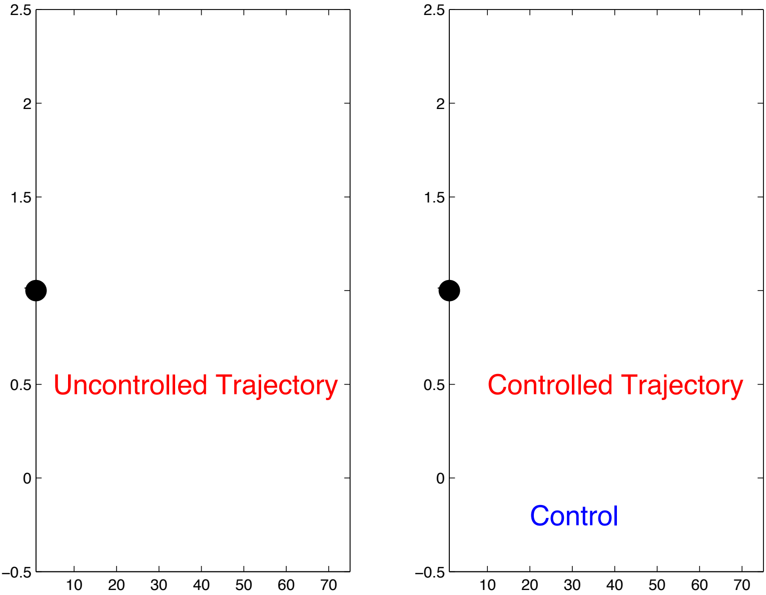
- Keep x as close to 0 as possible, but control is not free!
- We need to balance the objective to keep x close to zero with objective of keeping control costs down.

Linear Quadratic Control

$$\min_u \sum_{k=1}^{N} \frac{1}{2}(qx_k^2 + ru_k^2)$$

$$x_k = a_{k-1}x_{k-1} + b_k u_k, \quad k = 1, \ldots, N$$

- r and q are positive scalars.
- Many variations of this basic model exist.
- Linear Quadratic Control is well known because it has a closed form solution & useful in practice



2.5
2
1.5
0.5
0
−0.5
10
20
30
40
50
60
70
Uncontrolled Trajectory
Controlled Trajectory
Control

Example: Linear Quadratic Control

Write down the Linear Quadratic Control problem in the following form

$$\min_{z} \frac{1}{2} \boldsymbol{z}^\top \boldsymbol{Q} \boldsymbol{z}$$
$$\text{s.t. } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}$$

Find the closed form solution of the Linear Quadratic Control problem
You may assume that $\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top$ is invertible

Example: Linear Quadratic Control

$$Q = \begin{bmatrix} qI_N & 0_N \\ 0_N & rI_N \end{bmatrix} \quad z = [x_1, \ldots, x_N, u_1, \ldots, u_N]^\top$$

Where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix & 0_N is an $N \times N$ matrix of zeros.

$$A = \begin{bmatrix} 1 & & \cdots & 0 & -b_1 & & \cdots & 0 \\ -a_1 & 1 & & \vdots & & -b_2 & & \vdots \\ & \ddots & \ddots & & \vdots & & \ddots & \\ 0 & & -a_{N-1} & 1 & 0 & \cdots & & -b_N \end{bmatrix} \in \mathbb{R}^{N \times 2N}$$

$$b = [a_0 x_0, 0, \ldots, 0]^\top$$

Linear Quadratic Control in vector form

$$\min_z \frac{1}{2} z^\top Q z$$

$$\text{s.t. } Az = b$$

Remark: The $1/2$ in the objective function is used for convenience and does not change the solution.

Linear Quadratic Control in vector form

Note that Q is symmetric and positive definite. (why?).
We write down the Lagrangian as follows,

$$\mathcal{L}(z, \lambda) = \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{Q}\boldsymbol{z} + \boldsymbol{\lambda}^\top(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{z})$$

The Lagrange condition gives,

$$\nabla_z \mathcal{L}(\boldsymbol{z}, \boldsymbol{\lambda}) = \boldsymbol{Q}\boldsymbol{z} - \boldsymbol{A}^\top \boldsymbol{\lambda} = 0$$

Therefore,

$$\boldsymbol{z} = \boldsymbol{Q}^{-1}\boldsymbol{A}^\top \lambda$$

Multiplying both sides by A and using the condition $Az = b$ we get

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top \boldsymbol{\lambda}.$$

Therefore, $\boldsymbol{\lambda} = (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top)^{-1}\boldsymbol{b}$ (using the given assumption). So the optimal $\boldsymbol{z}$ is $\boldsymbol{z}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top)^{-1}b$. This is the only minimum point since the Hessian of the Lagrangian is positive definite and the problem is convex.

Outline

1. **Equality Constrained Optimisation**
    - Regular Points
    - First Order Necessary Conditions (Lagrange's Theorem)
    - Simple Proof of Lagrange's Theorem
    - Lagrange Multipliers – Interpretation
    - Second Order Necessary Conditions
    - Second Order Sufficient Conditions
    - Examples: Portfolio Optimisation, Optimal Control

2. **Problems with Inequality Constraints**
    - Definitions
    - Karush-Kuhn-Tucker Conditions
    - Second Order Conditions

Problem Formulation

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

- Where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$, $m \leq n$ and $h : \mathbb{R}^n \to \mathbb{R}^m$
- As before $h_i(\boldsymbol{x}) = 0$, $i = 1, \ldots, m$ are equality constraints
- $g_i(\boldsymbol{x}) \leq 0$, $i = 1, \ldots, p$ are inequality constraints
- The feasible region is $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n \mid h(\boldsymbol{x}) = 0,\ g(\boldsymbol{x}) \leq 0\}$.

Some definitions

Definition (Active constraint)

An inequality constraint $g_j(\boldsymbol{x}) \leq 0$ is said to be active at the point $\boldsymbol{x}^*$ if $g_j(\boldsymbol{x}^*) = 0$. It is inactive if $g_j(\boldsymbol{x}) < 0$.

Definition (Regular Points)

Let $\boldsymbol{x}^*$ satisfy $h(\boldsymbol{x}^*) = 0$ and $g(\boldsymbol{x}^*) \leq 0$, and let $J(\boldsymbol{x}^*)$ be the index set of active inequality constraints:

$$J(\boldsymbol{x}^*) \triangleq \{j \mid g_j(\boldsymbol{x}^*) = 0\}.$$

Then we say that $\boldsymbol{x}^*$ is a regular point if the vectors,

$$\nabla h_i(\boldsymbol{x}^*), \ \ \nabla g_j(\boldsymbol{x}^*), \ \ 1 \leq i \leq m, \ j \in J(\boldsymbol{x}^*).$$

are linearly independent.

Karush-Kuhn-Tucker Conditions

- 2^{nd} Berkeley Symposium on Mathematical Statistics and Probability, 1950.
  - A. W. Tucker mathematician (Princeton), Kuhn (PhD student)
  - Became known as the Kuhn-Tucker theorem
- Kuhn-Tucker not the first ones to study optimality conditions:
  - Karush (U. Chicago) proved the theorem in 1939 (in his MSc thesis)
  - Fritz John (1948), tried to publish it earlier, paper rejected

NONLINEAR PROGRAMMING

H. W. KUHN AND A. W. TUCKER

PRINCETON UNIVERSITY AND STANFORD UNIVERSITY

1. Introduction

Linear programming deals with problems such as (see [4], [5]): to maximize a linear function $g(x) \equiv \sum c_i x_i$ of n real variables $x_1, \ldots, x_n$ (forming a vector x) constrained by $m + n$ linear *inequalities*,

$$f_h(x) \equiv b_h - \sum a_{hi} x_i \geqq 0, \; x_i \geqq 0, \qquad h = 1, \ldots, m; \; i = 1, \ldots, n.$$

This problem can be transformed as follows into an equivalent saddle value (minimax) problem by an adaptation of the calculus method customarily applied to constraining *equations* [3, pp. 199–201]. Form the Lagrangian function

$$\phi(x, u) \equiv g(x) + \sum u_h f_h(x).$$

Then, a particular vector x^0 maximizes $g(x)$ subject to the $m + n$ constraints if, and only if, there is some vector u^0 with nonnegative components such that

$$\phi(x, u^0) \leqq \phi(x^0, u^0) \leqq \phi(x^0, u) \quad \text{for all nonnegative } x, u.$$

Such a saddle point (x^0, u^0) provides a solution for a related zero sum two person game [8], [9], [12]. The bilinear symmetry of $\phi(x, u)$ in x and u yields the characteristic duality of linear programming (see section 5, below).

Karush-Kuhn-Tucker (KKT) Theorem

Theorem (First Order Necessary Conditions (KKT Theorem))

Suppose that f, h, and g are $\mathcal{C}^1$. Let $\boldsymbol{x}^*$ be a regular and a local minimiser of,

$$\begin{aligned} \min\ & f(\boldsymbol{x}) \\ \textit{s.t. } & h(\boldsymbol{x}) = 0 \\ & g(\boldsymbol{x}) \leq 0 \end{aligned}$$

Then there exists a $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and a $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that:

a. $\boldsymbol{\mu}^* \geq 0$

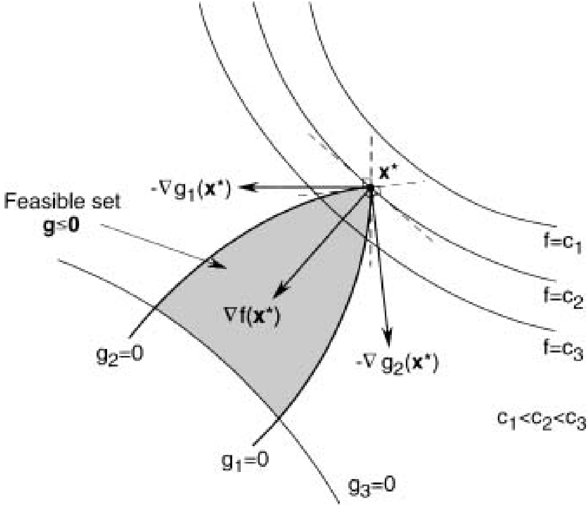
b. $\nabla_x f(\boldsymbol{x}^*) + \boldsymbol{\nabla}_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* + \boldsymbol{\nabla}_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = 0$

c. $\mu_i^* g_i(\boldsymbol{x}^*) = 0,\ i = 1, \ldots, p$

d. $h(\boldsymbol{x}^*) = 0$, $g(\boldsymbol{x}^*) \leq 0$

Remark: Note that because $\boldsymbol{\mu}^* \geq 0$ and $g(\boldsymbol{x}^*) \leq 0$ the last condition implies that if $g_j(\boldsymbol{x}^*) < 0$ then $\boldsymbol{\mu}_j^* = 0$.

KKT Theorem Graphical Illustration



Example 2: KKT Theorem

Use the KKT conditions to solve the following problem

$$\min \ -\frac{400x}{(10+x)^2}$$

$$\text{s.t. } x \geq 0$$

Example 2: KKT Theorem

The KKT conditions for this problem are:

$$
\begin{aligned}
&\frac{-400(10+x)^2 + 800x(10+x)}{(10+x)^4} - \mu = 0 \\
&\mu \geq 0 \\
&\mu x = 0 \\
&-x \leq 0
\end{aligned}
$$

If $\mu > 0$, then $x = 0$ and the first condition implies that $\mu < 0$. Therefore we cannot have $\mu > 0$ and $x = 0$. The other case is $x > 0$ and $\mu = 0$, in this case the first condition implies that,

$$-4(10+x)^2 + 8x(10+x) = 0$$

The only solution to the above is $x = 10$. Therefore the only KKT point is $\mu = 0$ and $x = 10$.

Maximisation problems

$$\begin{aligned} \max\ & f(\boldsymbol{x}) \\ \text{s.t. } & \boldsymbol{h}(\boldsymbol{x}) = 0 \\ & g(\boldsymbol{x}) \leq 0 \end{aligned}$$

In this case the KKT optimality conditions are,

a. $\boldsymbol{\mu}^* \geq 0$

b. $-\boldsymbol{\nabla}_x \boldsymbol{f}(\boldsymbol{x}^*) + \nabla_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* + \boldsymbol{\nabla}_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = 0$

c. $\mu_i^* g_i(\boldsymbol{x}^*) = 0,\ i = 1, \ldots, p$

d. $\boldsymbol{h}(\boldsymbol{x}^*) = 0$

e. $\boldsymbol{g}(\boldsymbol{x}^*) \leq 0$

This can be seen by the maximisation problem to a minimisation problem by multiplying the objective function with -1.

Maximisation problems

$$\begin{aligned} \max\ & f(\boldsymbol{x}) \\ \text{s.t. } \boldsymbol{h}(\boldsymbol{x}) &= 0 \\ \boldsymbol{g}(\boldsymbol{x}) &\leq 0 \end{aligned}$$

An alternative set of conditions are,

a. $\boldsymbol{\mu}^* \leq 0$

b. $\boldsymbol{\nabla}_x \boldsymbol{f}(x^*) + \boldsymbol{\nabla}_x \boldsymbol{h}(x^*)\lambda^* + \boldsymbol{\nabla}_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = 0$

c. $\mu_i^* g_i(\boldsymbol{x}^*) = 0$

d. $\boldsymbol{h}(\boldsymbol{x}^*) = 0$

e. $\boldsymbol{g}(\boldsymbol{x}^*) \leq 0$

Problems with $\geq$ constraints

$$\begin{aligned} \min\ f(\boldsymbol{x}) \\ \text{s.t. } \boldsymbol{h}(\boldsymbol{x}) = 0 \\ \boldsymbol{g}(\boldsymbol{x}) \geq 0 \end{aligned}$$

For this case we can multiply the constraints with -1 to obtain

a. $\boldsymbol{\mu}^* \geq 0$

b. $\boldsymbol{\nabla}_x f(\boldsymbol{x}^*) + \boldsymbol{\nabla}_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda} - \nabla_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu} = 0$

With conditions c., d. and e. the same as before.
Or we can change the sign of the multiplies and get the following conditions

a. $\boldsymbol{\mu}^* \leq 0$

b. $\boldsymbol{\nabla}_x \boldsymbol{f}(\boldsymbol{x}^*) + \nabla_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* + \nabla_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = 0$

With conditions c., d. and e. the same as before.

Example

Use the KKT conditions to solve the following problem

$$\min f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2 - 3x_1$$
$$\text{s.t. } x_1 \geq 0,\ x_2 \geq 0$$

Example

The KKT conditions for this problem are

a. $\boldsymbol{\mu} \leq 0$

b. $[2x_1 + x_2 - 3,\ x_1 + 2x_2]^\top + \mu = 0$

c. $\boldsymbol{\mu}^\top \boldsymbol{x} = 0$

d. $\boldsymbol{x} \geq 0$

We have four equations with four unknowns. To satisfy condition c. we first try $\mu_1 = 0$ and $x_2 = 0$ and solve to find

$$x_1 = \frac{3}{2} \quad \mu_2 = -\frac{3}{2}$$

Therefore $x_1 = \frac{3}{2}$, $x_2 = 0$, $\mu_1 = 0$ and $\mu_2 = -\frac{3}{2}$ do satisfy the KKT conditions.

On the other hand choosing $x_1 = 0$ and $\mu_2 = 0$ gives $x_2 = 0$ and $\mu_1 = 3$. Therefore the KKT conditions are not satisfied at this point.

Second Order Necessary Conditions

Theorem (Second Order Necessary Conditions)

Suppose that f, $\boldsymbol{h}$, and $\boldsymbol{g}$ are $\mathcal{C}^2$. Let $\boldsymbol{x}^*$ be a regular and a local minimiser of $f(\boldsymbol{x})$ over $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = 0,\ \boldsymbol{g}(\boldsymbol{x}) \leq 0\}$. Then there exists a $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and a $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that:

a. $\boldsymbol{\mu}^* \geq 0$, $\boldsymbol{\nabla}_x \boldsymbol{f}(\boldsymbol{x}^*) + \boldsymbol{\nabla}_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda} + \boldsymbol{\nabla}_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = 0$, $\mu_i^* g_i(\boldsymbol{x}^*) = 0$, $\boldsymbol{h}(\boldsymbol{x}^*) = 0$, $\boldsymbol{g}(\boldsymbol{x}^*) \leq 0$

b. $\boldsymbol{d}^\top \boldsymbol{H}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{d} \geq 0$ *for all* $\boldsymbol{d}$ *such that*

$$\nabla_x \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{d} = 0, \nabla_x g_j(x^*)^\top \boldsymbol{d} = 0,\ j \in J(\boldsymbol{x}^*)$$

Where $\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is the Hessian of the Lagrangian defined as,

$$\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{\nabla}^2_{xx} \boldsymbol{f}(x) + \sum_{i=1}^{m} \lambda_i \boldsymbol{\nabla}^2_{xx} \boldsymbol{h}_i(x) + \sum_{i=1}^{p} \mu_i \nabla^2_{xx} \boldsymbol{g}_i(x)$$

Second Order Sufficient Conditions

Theorem (Second Order Sufficient Conditions)

Suppose that f, $\boldsymbol{h}$, and $\boldsymbol{g}$ are $\mathcal{C}^2$. Let $\boldsymbol{x}^*$ be a regular and a local minimiser of $f(\boldsymbol{x})$ over $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = 0,\ \boldsymbol{g}(\boldsymbol{x}) \leq 0\}$. Then there exists a $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and a $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that:

a. $\boldsymbol{\mu}^* \geq 0$, $\boldsymbol{\nabla}_x \boldsymbol{f}(\boldsymbol{x}^*) + \boldsymbol{\nabla}_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda} + \boldsymbol{\nabla}_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = 0$, $\mu_i^* g_i(\boldsymbol{x}^*) = 0$, $\boldsymbol{h}(\boldsymbol{x}^*) = 0$, $\boldsymbol{g}(\boldsymbol{x}^*) \leq 0$

b. $\boldsymbol{d}^\top \boldsymbol{H}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{d} > 0$ *for all* $\boldsymbol{d} \neq 0$ *such that*

$$\nabla_x \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{d} = 0, \nabla_x g_j(x^*)^\top \boldsymbol{d} = 0,\ j \in J(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$$

where $J(x^*, \mu^*) = \{i \mid g_i(x^*) = 0, \mu_i^* > 0\}$. (note that $J(x^*, \mu^*)$ is a subset of $J(x^*)$).

Then x^* is a strict local minimiser of f over $\{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \leq 0\}$.

Example 1

Consider the following problem:

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

a. Write down the KKT conditions for this problem.
b. Find all the points (and KKT multipliers) satisfying the KKT condition. In each case, determine if each point is regular.
c. Find all the points in part (b) that also satisfy the SONC.
d. Find all the points in part (c) that also satisfy the SOSC.

Example 1a: Solution

a.

Example 1a: Solution

a. The KKT conditions are,

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

Example 1b,c: Solution

b.

c.

Example 1b,c: Solution

b. Note that $\mu_1 > 0$ and $\mu_2 > 0$ is not possible. Also $\mu_1 = \mu_2 = 0$ is not possible. And $\mu_1 = 0$ $\mu_2 > 0$ also leads to a contradiction. So we are left with $\mu_1 > 0$ and $\mu_2 = 0$. In this case we find the unique solution to be $x_1^* = x_2^* = 1$ and $\mu_1^* = 1$, $\mu_2^* = 0$. For this point we have the gradient of the first constraint is $[-1,\ -1]^\top$ and the gradient of the second constraint is $[1,\ -1]^\top$. Hence the point x^* is regular.

c. Both constraints are active. Note that at the point x^* we need to ensure that the Hessian of the Lagrangian is positive definite for all vectors d such that,

$$[-1,\ -1]\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 0$$

$$[1,\ -1]\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 0$$

This is only possible when $d_1 = d_2 = 0$. From which we conclude that the second order necessary conditions are satisfied.

Example 1d: Solution

d.

Example 1d: Solution

d. The Hessian of the Lagrangian is,

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

And it needs to be positive definite on all vectors $d \neq 0$ that satisfy,

$$-d_1 - d_2 = 0.$$

If we pick $d = [1,\ -1]^\top$ then,

$$[1,\ -1]^\top \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 < 0.$$

Therefore the SOSC is not satisfied.

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

Theorem (First Order Necessary & Sufficient Conditions (KKT Conditions the convex case))

Suppose that the point $(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ *satisfies the following:*

a. $\boldsymbol{\mu}^* \geq 0$
b. $\boldsymbol{\nabla}_x f(\boldsymbol{x}^*) + \boldsymbol{\nabla}_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* + \boldsymbol{\nabla}_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = 0$
c. $\mu_i^* g_i(\boldsymbol{x}^*) = 0,\ i = 1, \ldots, p$
d. $h(\boldsymbol{x}^*) = 0$, $g(\boldsymbol{x}^*) \leq 0$

In addition assume that f *is convex and that the feasible region is convex, and that there exists a point* $\bar{\boldsymbol{x}}$ *such that* $\boldsymbol{g}(\bar{\boldsymbol{x}}) < 0$ *(Slater's constraint qualification) then* x^* *is a global minimiser.*