Non-Linear Programming

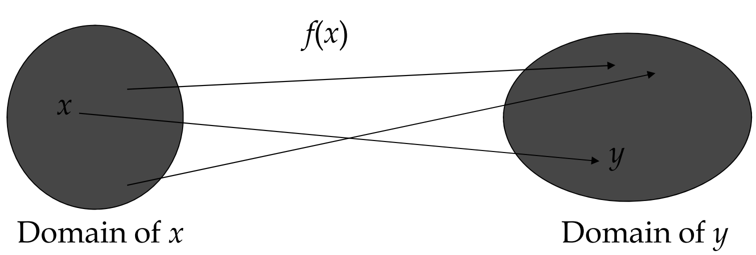
IE 322

Introduction

- A function $f(x)$ is a rule that assigns to every choice of x a unique value $y = f(x)$.

  x is called the independent variable

  y is called the dependent variable



PROPERTIES OF SINGLE-VARIABLE FUNCTIONS

- Let R be the set of the real numbers and S the domain of x.
- When $S = R$, we have an unconstrained function:

  $f(x) = x^3 + 2x^2 - x + 3$, for all $x \in R$

- When $S \subset R$, we have a constrained function:

  $f(x) = x^3 + 2x^2 - x + 3$, for all $x \in S = \{x \mid -5 \le x \le 5\}$

Definition: Continuous Functions

A function $f(x)$ is continuous at $x = c$ if and only if it meets the following conditions:

1. $f(c)$ exists
2. $\lim_{x \to c} f(x)$ exists
3. $\lim_{x \to c} f(x) = f(c)$

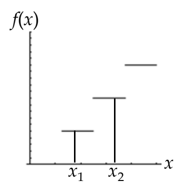
Properties of continuous functions:

1. Sums or products of continuous functions are continuous
2. The ratio of two continuous functions is continuous at all points where the denominator does not vanish

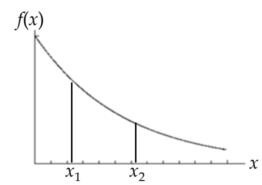
Definition: Monotonic Functions

A function $f(x)$ is monotonic (either increasing or decreasing) if for any two points x_1 and x_2 with $x_1 \leq x_2$, it follows that

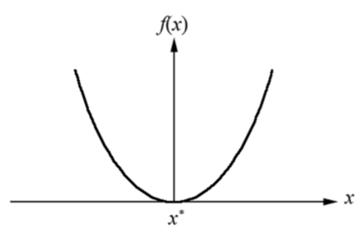
$f(x_1) \leq f(x_2)$	monotonically increasing
$f(x_1) \geq f(x_2)$	monotonically decreasing



Monotonically increasing function



Monotonically decreasing function



Unimodal function

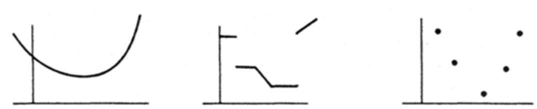
- The function is monotonically decreasing for $x^* \leq 0$ and monotonically increasing for $x^* \geq 0$
- The function attains its minimum at $x = x^*$
- The function is monotonic on either side of the minimum
- The function is said to be a *unimodal function*

Definition

- A function is unimodal on the interval $a \le x \le b$ if and only if it is monotonic on either side of the single optimal point x^*.
- If x^* is the single minimum point of $f(x)$ in the range $a \le x \le b$, then $f(x)$ is unimodal on the interval if and only if for any two points x_1 and x_2:

$$x^* \le x_1 \le x_2 \text{ implies } f(x^*) \le f(x_1) \le f(x_2)$$

$$x^* \ge x_1 \ge x_2 \text{ implies } f(x^*) \le f(x_1) \le f(x_2)$$



(i) continuous (ii) discontinuous (iii) discrete

Unimodal functions.

OPTIMALITY CRITERIA FOR SINGLE-VARIABLE FUNCTIONS

We want to answer this question:

- How can we determine whether a given point x^* is the optimal solution?

- We will develop a set of optimality criteria for determining whether a given solution is optimal.

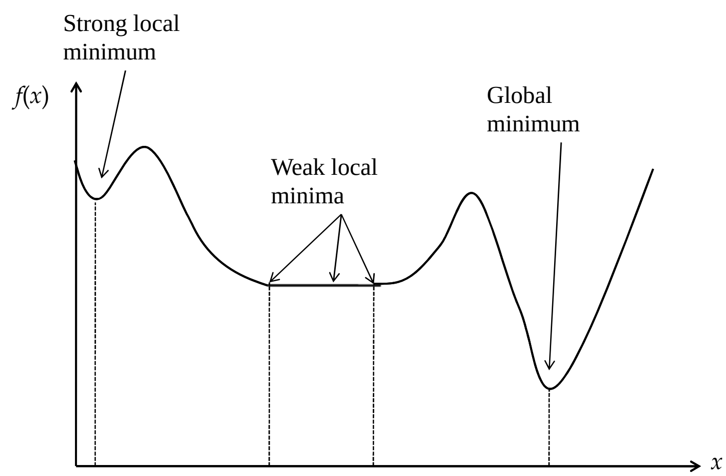
Definition

A function $f(x)$ defined on a set S attains its global minimum at a point $x^{**} \in S$ if and only if

$$f(x^{**}) \leq f(x) \qquad \text{for all } x \in S$$

A function $f(x)$ defined on S has a local minimum at a point $x^{*} \in S$ if and only if

$$f(x^{*}) \leq f(x) \qquad \text{for all } x \text{ satisfying } |x - x^{*}| < \varepsilon$$



Theorem (Necessity Theorem)

The necessary condition for x^* to be a local minimum of f on the interval (a, b), provided that f is differentiable, is

$$\left.\frac{df}{dx}\right|_{x=x^*} = 0$$

Definition

- A stationary point is a point x^* at which $\left.\frac{df}{dx}\right|_{x=x^*} = 0$.
- An inflection point (saddle point) is a stationary point that does not correspond to a local optimum.

Theorem (Sufficiency Theorem)

Suppose at x^* the first derivative is zero and the first non-zero higher order derivative is denoted by n.

1. If n is odd, x^* is an inflection point
2. If n is even, x^* is a local optimal
   - If that derivative is (+), x^* is a local minimum
   - If that derivative is (-), x^* is a local maximum

Example

Consider the function

$$f(x) = 5x^6 - 36x^5 + \tfrac{165}{2}x^4 - 60x^3 + 36$$

Which of the following points are minimizers of $f(x)$?

a) $x_1 = 0$

b) $x_2 = 1$

c) $x_3 = 2$

d) $x_4 = 3$

Solution

We will apply the sufficiency theorem on the points. First, we eliminate any point using the zero-first-derivative test. The first derivative of $f(x)$ is

$$\frac{df}{dx} = 30x^5 - 180x^4 + 330x^3 - 180x^2$$

Clearly all the points make the first derivative equal to zero, hence, all are labeled as critical points which potentially can be minimizers of the function $f(x)$. Now, we test the second derivative and take out any point at which the second derivative is negative:

- $f''(x_1) = 0$
- $f''(x_2) = 60$
- $f''(x_3) = -120$
- $f''(x_4) = 540$

The point x_3 can not be a minimizer, so it should be eliminated.

Since the second derivatives at x_2 and x_4 are positive, they are local minimizers.

The second derivative at x_1 is zero, hence no definitive conclusion can be make and a higher derivative is needed:

$$f'''(x_1) = -360$$

This is an odd-order derivative and is nonzero, hence, x_1 is a saddle point.

Functions of Several Variables

- We will develop conditions for the characterization of optima for functions of several variables.
- The unconstrained multi-variable problem is written as

$$\min f(x) \qquad x \in R^N$$

where x is a vector of the decision variables.

- We assume f and its derivatives exist and are continuous everywhere.
- We assume the minimum does not correspond to a point of discontinuity (either f or the gradient).

OPTIMALITY CRITERIA

- The Taylor expansion of a function of several variables is

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T \Delta x + \tfrac{1}{2} \Delta x^T \nabla^2 f(\bar{x}) \Delta x$$

where

$\bar{x}$ = the current point

$\Delta x = x - \bar{x}$ = the change in x

∇f = the gradient (first derivative) of $f(x)$ {vector}

$\nabla^2 f = H_f$ = the Hessian (second derivative) of $f(x)$ {matrix}

- The change in $f(x)$ as a result of change in x is given by:

$$\Delta f(x) = f(x) - f(\bar{x}) = \nabla f(\bar{x})^T \Delta x + \tfrac{1}{2} \Delta x^T \nabla^2 f(\bar{x}) \Delta x$$

- If $\bar{x}$ is a local minimum, then we have to have

$$\Delta f(x) = f(x) - f(\bar{x}) \geq 0$$

- The point $\bar{x}$ is a global minimum if the above relation is true for all $x \in R^N$.

Theorem Necessary Conditions

For x^* to be a local minimum, it is necessary that

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

- When the gradient of $f(x)$ is zero, we have

$$\Delta f(x) = \tfrac{1}{2}\Delta x^T \nabla^2 f(\bar{x})\Delta x$$

- Now, the sign of Δf is dependent on the sign of the quadratic form

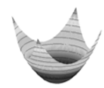
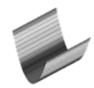
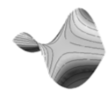
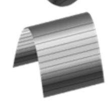
$$Q(x) = \Delta x^T \nabla^2 f(\bar{x})\Delta x$$

- If $Q(x) > 0$, then $\bar{x}$ is a minimum.
- If $Q(x) < 0$, then $\bar{x}$ is a maximum.

- From linear algebra,
  - $Q(x) > 0$ if $\nabla^2 f(\bar{x})$ is positive definite
  - $Q(x) < 0$ if $\nabla^2 f(\bar{x})$ is negative definite

- Based on this, we can say that the stationary point $\bar{x}$ is
  - minimum if $\nabla^2 f(\bar{x})$ is positive definite
  - maximum if $\nabla^2 f(\bar{x})$ is negative definite
  - saddlepoint if $\nabla^2 f(\bar{x})$ is indefinite

Stationary point nature summary

$x^T\mathbf{H}x$	λ_i	**Definiteness H**	**Nature x***	
> 0		Positive d.	Minimum	
≥ 0		Positive semi-d.	Valley	
$\neq 0$		Indefinite	Saddlepoint	
≤ 0		Negative semi-d.	Ridge	
< 0		Negative d.	Maximum	

Theorem Sufficient Conditions

If

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

and

$$\nabla^2 f(x^*) \text{ is positive definite}$$

then x^* is a local minimum.

Example

See the example and the solution in the book.

Constrained Optimality

- In this part, we will develop necessary and sufficient conditions of optimality for constrained problems.
- The conditions we have developed previously no longer hold because of the presence of constraints.

EQUALITY-CONSTRAINED PROBLEMS

- Firstly, we consider the optimization problem involving several equality constraints:

$$\begin{array}{lll} \min & f(x_1, x_2, \ldots, x_N) & \\ \text{subject to} & \mathrm{h}_k(x_1, x_2, \ldots, x_N) = 0 & k = 1, \ldots, K \end{array}$$

LAGRANGE MULTIPLIERS

- We can convert the constrained problem to an equivalent unconstrained problem with the help of certain unspecified parameters known as Lagrange multipliers.
- Consider the problem

$$\begin{array}{ll} \min & f(x_1, x_2, \ldots, x_N) \\ \text{subject to} & \mathrm{h}_1(x_1, x_2, \ldots, x_N) = 0 \end{array}$$

- The method of Lagrange multipliers converts it to the unconstrained problem

$$\min \quad L(x; \lambda) = f(x) - \lambda h_1(x)$$

The Lagrange function

The Lagrange multiplier

- The Lagrange multiplier method can be extended to several equality constraints.

$$\begin{array}{lll} \min & f(x_1, x_2, \ldots, x_N) & \\ \text{subject to} & h_k(x_1, x_2, \ldots, x_N) = 0 & k = 1, \ldots, K \end{array}$$

- The Lagrange function becomes

$$\min \qquad L(x;\lambda) = f(x) - \sum_{k=1}^{K} \lambda_k h_k$$

- Here, λ_1, λ_2, …, λ_k are the Lagrange multipliers

- To solve the Lagrange problem, we set the partial derivatives of L with respect to x equal to zero and add to them the constrains:

$$\begin{aligned} &\frac{\partial L(x;\lambda)}{\partial x_1} = 0 \\ &\frac{\partial L(x;\lambda)}{\partial x_2} = 0 \\ &\vdots \\ &\frac{\partial L(x;\lambda)}{\partial x_N} = 0 \\ &h_1(x) = 0 \\ &h_2(x) = 0 \\ &\vdots \\ &h_K(x) = 0 \end{aligned}$$

- These equations form a system of nonlinear equations of $N+K$ equations in $N+K$ variables.
- The critical points of L are tested by computing the Hessian matrix of L with respect to x.

KUHN-TUCKER CONDITIONS

- Lagrange multipliers can be used to develop optimality criteria for the equality constrained optimization problems.
- Kuhn and Tucker can develop optimality criteria and include equality and inequality constrained problems:

$$
\begin{aligned}
\text{min} \quad & f(x) \\
\text{subject to} \quad & g_j(x) \le 0 \quad \text{for } j = 1,2,\ldots,J \\
& h_k(x) = 0 \quad \text{for } \; k = 1,2,\ldots,K
\end{aligned}
$$

- The general NLP problem is given by

$$\begin{aligned} &\min \quad f(x) \\ &\text{subject to} \quad g_j(x) \le 0 \quad \text{for } j = 1,2,\ldots,J \\ &\qquad\qquad h_k(x) = 0 \quad \text{for } k = 1,2,\ldots,K \end{aligned}$$

- To solve this problem, we write the Lagrange function:

$$\min \; L(x,\mu,\lambda) = f(x) - \sum_{j=1}^{J} \mu_j g_j(x) - \sum_{k=1}^{K} \lambda_k h_k(x)$$

- To find the critical points of L,

$$\frac{\partial L}{\partial x_i} = 0 \quad \text{for } i = 1,\ldots,N$$

- We should adjust the values of u_j and υ_k until

min L = min f

Kuhn-Tucker Conditions or Kuhn-Tucker Problem

- The KT conditions are stated as:

Find vectors $x_{(N\times1)}, u_{(1\times J)}$, and $\upsilon_{(1\times K)}$ that satisfy

$$\nabla f(x) - \sum_{j=1}^{J} \mu_j \nabla g_j(x) - \sum_{k=1}^{K} \lambda_k \nabla h_k(x) = 0$$

$$\begin{aligned} &g_j(x) \le 0 && \text{for } j = 1,2,\ldots,J \\ &h_k(x) = 0 && \text{for } k = 1,2,\ldots,K \\ &\mu_j g_j(x) = 0 && \text{for } j = 1,2,\ldots,J \\ &\mu_j \le 0 && \text{for } j = 1,2,\ldots,J \end{aligned}$$

Example

See the example and the solution in the book

KUHN-TUCKER THEOREMS

Theorem 5.1 Kuhn-Tucker Necessity Theorem

Consider the NLP problem

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0 \quad \text{for } j = 1,2,\ldots,J \\ & h_k(x) = 0 \quad \text{for } k = 1,2,\ldots,K \end{aligned}$$

Let f, g, and h be differentiable functions and x^* be a feasible solution to the NLP problem. Let $I = \{j \mid g_j(x^*) = 0\}$. Further, $\nabla g_j(x^*)$ for $j \in I$ and $\nabla h_k(x^*)$ for $k = 1,\ldots, K$ are linearly independent. If x^* is an optimal solution to the NLP, then there exists a (u^*, υ^*) such that (x^*, u^*, υ^*) solves the KT problem.

- The conditions $\nabla g_j(x^*)$ for $j \in I$ and $\nabla h_k(x^*)$ for $k = 1, \ldots, K$ are linearly independent at the optimum is known as the constraint qualification.
- The constraint qualification holds in the following cases:
  1. When all the constraints are linear.
  2. When all the inequality constraints are concave functions and the equality constraints are linear and there exists at least one feasible x that is strictly inside the feasible region of the inequality constraints.
     In other words, there exists an $\bar{x}$ such that $g_j(\bar{x}) > 0$ for $j = 1,\ldots, J$ and $h_k(\bar{x}) = 0$ for $k = 1,\ldots, K$.

- When the constraint qualification is not met at the optimum, there may not exist a solution to the Kuhn-Tucker problem.
- Therefore, do not apply the Kuhn-Tucker optimality conditions when the constraint qualification is not met.

Example

Consider the NLP problem

$$\min f(x) = 1 - x^2$$
$$\text{s.t.} \quad -1 \le x \le 3$$

Check the optimality of $x = 2$.

Solution

- First we have to write the problem in our standard form:

$$\begin{aligned} &\min f(x) = 1 - x^2 \\ &\text{s.t.} \quad g_1(x) = -x - 1 \le 0 \\ &\qquad\;\; g_2(x) = -3 + x \le 0 \end{aligned}$$

- Before we can apply the necessity theorem, we have to check that the constraint qualification holds at $x = 2$.
- Since g_1 and g_2 are linear function, the constraint qualification holds. Therefore, we can apply the theorem.
- We need to check that there is u such that x and u are a solution to the KT problem.
- The KT conditions are

$$\begin{array}{ll} (1) & -2x - u_1 + u_2 = 0 \\ (2) & -x - 1 \le 0 \\ & -3 + x \le 0 \\ (3) & \text{NA} \\ (4) & u_1(x+1) = 0 \\ & u_2(3-x) = 0 \\ (5) & u_1, u_2 \le 0 \end{array}$$

- For $\xi = 2$, Condition (4) requires $\upsilon_1 = \upsilon_2 = 0$.
- Condition (1) is violated ($-4 \neq 0$).
- There is no solution to the KT problem for $\xi = 2$.
- Hence, $\xi = 2$ cannot be optimal.

Example

Consider the optimization problem

$$\begin{array}{ll} \min & f(x) = (x_1 - 1)^2 + x_2^2 \\ \text{s.t.} & g_1(x) = x_1 - x_2^2 \leq 0 \end{array}$$

Apply the KKT necessity theorem to check the optimality of $x = [0, 0]^T$.

Solution

We have one constraint, the constraint qualification holds.

For $x = [0, 0]^T$ and $u_1 = 2$, the KKT conditions are satisfied.

Therefore, we cannot make a definite decision on the optimality of x.

Kuhn-Tucker Sufficiency Theorem

Consider the NLP problem

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & g_j(x) \le 0 \quad \text{for } j = 1,2,\ldots,J \\ & h_k(x) = 0 \quad \text{for } \quad k = 1,2,\ldots,K \end{array}$$

Let the objective function $f(x)$ be convex, the inequality constraints $g_j(x)$ be all concave function for $j = 1, \ldots, J$, and the equality constraints $h_k(x)$ for $k = 1, \ldots, K$ be linear. If there exists a solution (x^*, u^*, υ^*) that satisfies the Karush-Kuhn-Tucker conditions, then x^* is an optimal solution to the NLP problem.

Example

Consider the optimization problem

$$\begin{array}{ll} \min & f(x) = x_1^2 - x_2 \\ \text{s.t.} & g_1(x) = -x_1 + 1 \le 0 \\ & g_2(x) = -26 + x_1^2 + x_2^2 \le 0 \\ & h_1(x) = x_1 + x_2 - 6 = 0 \end{array}$$

We want to prove the point $x^* = [1, 5]^T$ is optimal.

$f(x)$ is convex because $\boldsymbol{H}_f$ is positive semidefinite.

$g_1(x)$ is linear, hence it is both convex and concave.

$g_2(x)$ is concave because $\boldsymbol{H}_{g2}$ is negative definite.

h_1 is linear.

The conditions of the theorem are satisfied.

The KT conditions are

$$
\begin{aligned}
2x_1 - u_1 + 2x_1u_2 - \upsilon_1 &= 0 \\
-1 \qquad + 2x_2u_2 - \upsilon_1 &= 0 \\
-x_1 + 1 &\le 0 \\
-26 + x_1^2 + x_2^2 &\le 0 \\
x_1 + x_2 - 6 &= 0 \\
u_1(x_1 - 1) &= 0 \\
u_2(26 - x_1^2 - x_2^2) &= 0 \\
u_1, u_2 &\le 0
\end{aligned}
$$

Substitute in the value of x^*.

The KT conditions become

$$
\begin{aligned}
2 - u_1 + 2u_2 - \upsilon_1 &= 0 \\
-1 \quad + 10u_2 - \upsilon_1 &= 0 \\
1 - 1 &\le 0 \\
-26 + 1 + 25 &\ge 0 \\
1 + 5 - 6 &= 0 \\
u_1(1-1) = 0 &\rightarrow u_1 \le 0 \\
u_2(-26 + 1 + 25) = 0 &\rightarrow u_2 \le 0 \\
u_1, u_2 &\le 0
\end{aligned}
$$

For $\upsilon_1 = 1$, u = [1.4, 0.2]. Hence, x^* solves the KT problem.

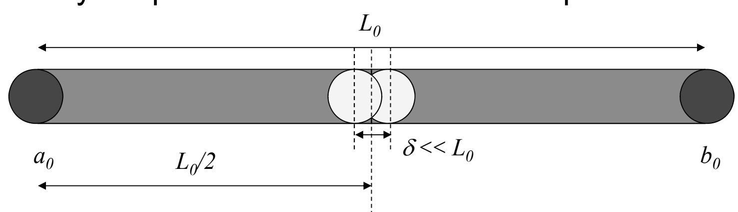
Therefore, x^* is an optimal solution.

Dichotomous search

- Conceptually simple idea:

> Main Entry: **di·chot·o·mous**
> Pronunciation: `dI-'kät-&-m&s` *`also`* `d&-`
> Function: *`adjective`*
> `: dividing into two parts`

  - Try to split interval in half in each step

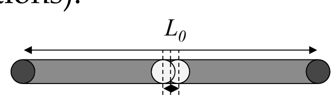


$f(\leftarrow) > f(\rightarrow)$:

Dichotomous search (2)

- Interval size after 1 step (2 evaluations):

$$L_1 = \frac{1}{2}(L_0 + \Delta)$$



- Interval size after m steps ($2m$ evaluations):

$$L_m = \frac{L_0}{2^m} + \Delta\left(1 - \frac{1}{2^m}\right)$$

- Proper choice for Δ:

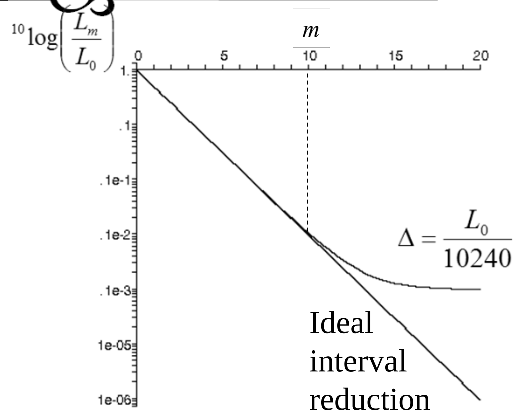
$$L_m^{\text{ideal}} = \frac{L_0}{2^m} \qquad \Rightarrow \qquad \Delta < \frac{L_m}{10} \le \frac{L_m^{\text{ideal}}}{10} = \frac{L_0}{10 \cdot 2^m}$$

Dichotomous search (3)

Example: m = 10

$$L_{10}^{\text{ideal}} = \frac{L_0}{2^{10}} = \frac{L_0}{1024}$$

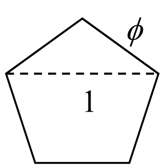
$$\Rightarrow \Delta = \frac{L_{10}^{\text{ideal}}}{10} = \frac{L_0}{10240}$$

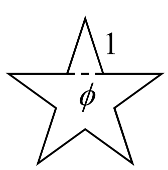


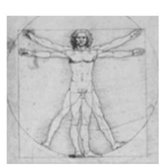
Sectioning – Golden Section

Golden section ratio ϕ =0.618034…

- Golden section method uses this *constant* interval reduction ratio ϕ

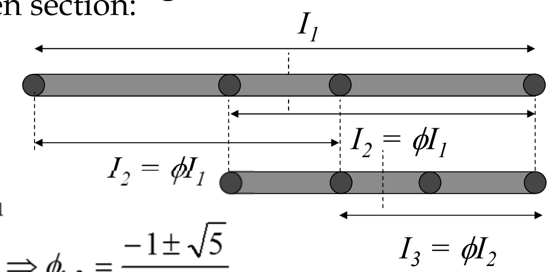






Sectioning - Golden Section

- Origin of golden section:



$$I_1 = I_2 + I_3$$

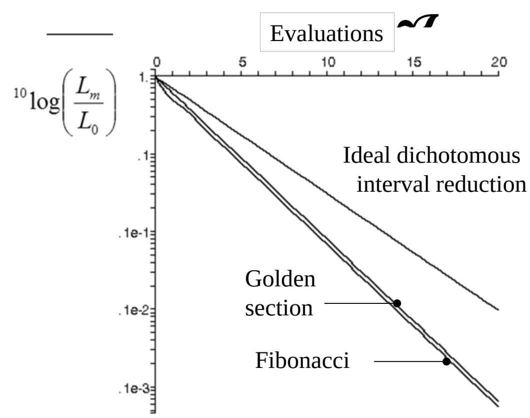
$$\Rightarrow I_1 = \phi I_1 + \phi^2 I_1$$

$$\phi^2 + \phi - 1 = 0 \Rightarrow \phi_{1,2} = \frac{-1 \pm \sqrt{5}}{2}$$

$$\phi = \frac{\sqrt{5}-1}{2} \approx 0.618034$$

- Final interval: $I_N = \phi^N I_1$

Comparison sectioning methods



Example: reduction to 2% of original interval:

	N
Dichotomous	12
Golden section	9

- Conclusion: Golden section simple and near-optimal

Example 19.1-1

$$\text{Maximize } f(x) = \begin{cases} 3x, & 0 \le x \le 2 \\ \frac{1}{3}(-x + 20), & 2 \le x \le 3 \end{cases}$$

The maximum value of *f(x)* occurs at *x* = 2. The following table demonstrates the calculations for iterations 1 and 2 using the dichotomous and the golden section methods. We will assume Δ = 0.1

Example 19.1-1 (cont'd)

Dichotomous method	Golden section method
Iteration 1	*Iteration 1*
$I_0 = (0, 3) \equiv (x_L, x_R)$	$I_0 = (0, 3) \equiv (x_L, x_R)$
$x_1 = 0 + .5(3 - 0 - .1) = 1.45,\ f(x_1) = 4.35$	$x_1 = 3 - .618(3 - 0) = 1.146,\ f(x_1) = 3.438$
$x_2 = 0 + .5(3 - 0 + .1) = 1.55,\ f(x_2) = 4.65$	$x_2 = 0 + .618(3 - 0) = 1.854,\ f(x_2) = 5.562$
$f(x_2) > f(x_1) \Rightarrow x_L = 1.45,\ I_1 = (1.45, 3)$	$f(x_2) > f(x_1) \Rightarrow x_L = 1.146,\ I_1 = (1.146, 3)$
Iteration 2	*Iteration 2*
$I_1 = (1.45, 3) \equiv (x_L, x_R)$	$I_1 = (1.146, 3) \equiv (x_L, x_R)$
$x_1 = 1.45 + .5(3 - 1.45 - .1) = 2.175,\ f(x_1) = 5.942$	$x_1 = x_2$ in iteration 0 $= 1.854,\ f(x_1) = 5.562$
$x_2 = \frac{3 + 1.45 + .1}{2} = 2.275,\ f(x_2) = 5.908$	$x_2 = 1.146 + .618(3 - 1.146) = 2.292,\ f(x_2) = 5.903$
$f(x_1) > f(x_2) \Rightarrow x_R = 2.275,\ I_2 = (1.45, 2.275)$	$f(x_2) > f(x_1) \Rightarrow x_L = 1.854,\ I_1 = (1.854, 3)$

Continuing in the same manner, the interval of uncertainty will eventually narrow down to the desired Δ-tolerance.