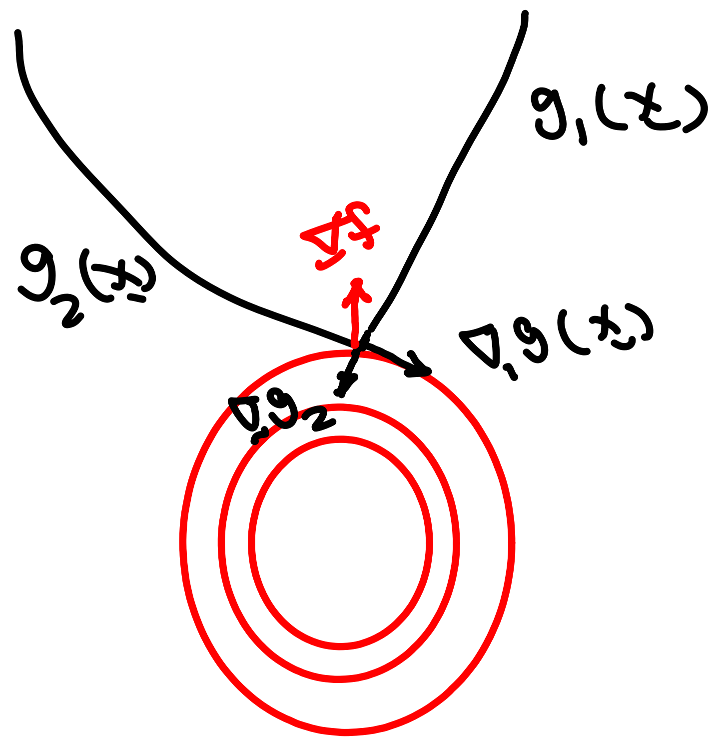
ECE733, Nonlinear Optimization for Electrical Engineers, Dr. Mohamed Bakr

Lecture 3

Inequality constraints, Line search Techniques

Illustration of Equality Constraints



$$\nabla f = \lambda_1 g_1(\underline{x}) + \lambda_2 g_2(\underline{x})$$

λ_1, λ_2 are unconstrained

Interpertation of Lagrange Multipliers

* $g(\underline{x})=0$ is the given constraint
* $g(\underline{x}) = b \Rightarrow b - g(\underline{x}) = 0$ is a relaxed constraint $\Rightarrow db = dg(\underline{x}) = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i} dx_i \leftarrow (1)$
* For the relaxed problem, we have

$$L(\underline{x}, \lambda) = f(\underline{x}) + \lambda (b - g(\underline{x}))$$

$$\frac{\partial L}{\partial \underline{x}} = 0 \Rightarrow \frac{\partial f}{\partial \underline{x}} - \lambda \frac{\partial g}{\partial \underline{x}} = 0$$

$$\Rightarrow \frac{\partial g}{\partial x_i} = \frac{1}{\lambda} \frac{\partial f}{\partial x_i} \longleftarrow (2)$$

Interpertation of Lagrange Multipliers (Cont'd)

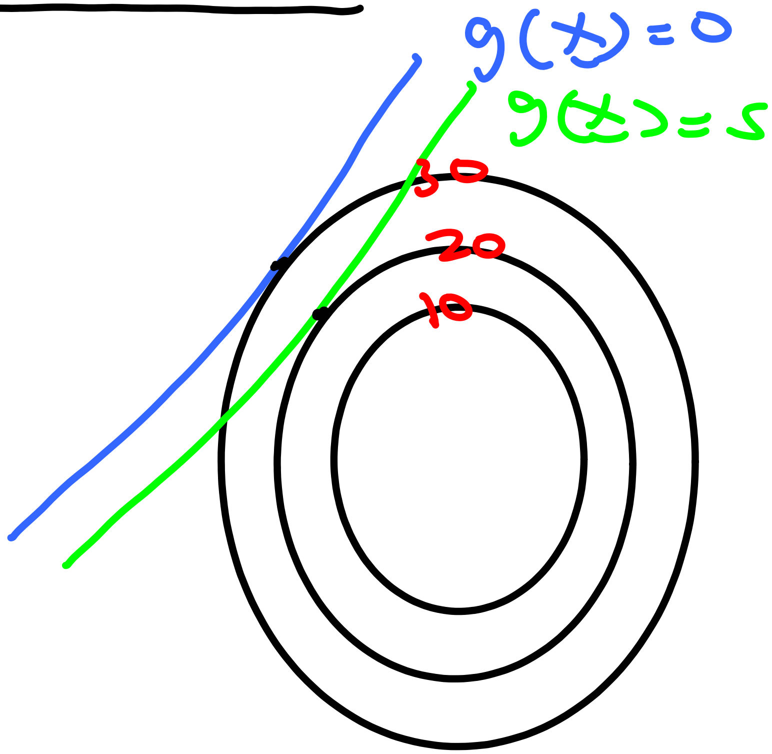
* subsituting from ② into ①

$$db = \frac{1}{\lambda} \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i = \frac{df}{\lambda}$$

$\Rightarrow$ $df = \lambda\, db$ $\Rightarrow$ λ describes the sensitivity of the objective function relative to the relaxation of the constraint!

* λ is +ve $\Rightarrow$ if db is +ve $\Rightarrow$ df is +ve
* λ is −ve $\Rightarrow$ if db is +ve $\Rightarrow$ df is −ve
* λ is 0 $\Rightarrow$ $df = 0$ $\Rightarrow$ unconstrained minimum

Illustration



λ is $-$ve

db is $+$ve

$\Downarrow$

df is $-$ve

Optimization using Inequality Constraint

* $\min\limits_{\underline{x}} f(\underline{x})$ subject to $g_j(\underline{x}) \leq 0$, $j = 1, 2, \ldots, m$
* We can add slack variables to convert inequality constraints into equality constraints
  $g_j(\underline{x}) + y_j^2 = 0$, $j = 1, 2, \ldots, m$
* We can apply the method of Lagrange multipliers for equality constraints on $(n + 2m)$ unknowns

Inequality Constraints (Cont'd)

$$L(\underline{x}, \underline{y}, \underline{\lambda}) = f(\underline{x}) + \sum_{j=1}^{m} \lambda_j \left(g_j(\underline{x}) + y_j^2 \right)$$

$$\frac{\partial L}{\partial \underline{x}} = 0 \Rightarrow \frac{\partial f(\underline{x})}{\partial \underline{x}} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(\underline{x})}{\partial \underline{x}} = 0$$

$\frac{\partial L}{\partial y_j} = 2 \lambda_j y_j = 0 \Longrightarrow$ $y_j = 0$ and λ_j is nonzero $\Rightarrow$ Constraint is active at $\underline{x}^*$

or $\lambda_j = 0$, $y_j \neq 0 \Rightarrow$ Constraint is inactive at $\underline{x}^*$.

Inequality Constraints (Cont'd)

* Constraints are divided into a set of active constraints J_1 and a set of inactive constraints J_2.
* The necessary KKT conditions are:

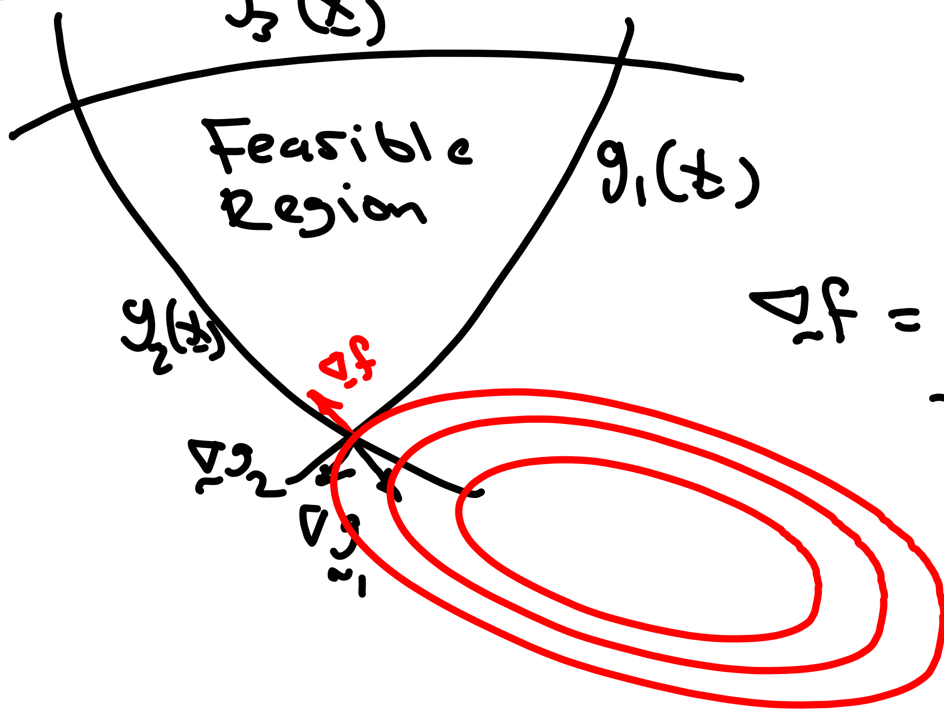
$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \quad i = 1, 2, \dots, n$$

$$\lambda_j g_j = 0, \quad j = 1, 2, \dots, m$$

$$g_j \leq 0, \quad j = 1, 2, \dots, m$$

$$\lambda_j \geq 0, \quad j = 1, 2, \dots, m \qquad (\text{why } \lambda_j \geq 0\text{?})$$

Illustration



$$\nabla f = -\lambda_1 \nabla g_1 - \lambda_2 \nabla g_2$$

$$\lambda_1, \lambda_2 > 0$$

$$\lambda_3 = 0$$

We do not know beforehand the active Constraints!

The General Theory

* For the problem $\min\limits_{\underset{\sim}{x}} f(\underset{\sim}{x})$

subject to $g_j(\underset{\sim}{x}) \leq 0, \; j = 1, 2, \ldots, m$

$h_k(\underset{\sim}{x}) = 0, \; k = 1, 2, \ldots, p$

the necessary conditions are

$$\nabla f + \sum_{j=1}^{m} \lambda_j \nabla g_j - \sum_{k=1}^{p} \beta_k \nabla h_k = \underset{\sim}{0}$$

$$\lambda_j g_j = 0, \; j = 1, 2, \ldots, m, \quad h_k(\underset{\sim}{x}) = 0, \; k = 1, 2, \ldots, p$$

Example

Solve the constrained minimization problem

$$f(\underline{x}) = x_1^2 + x_2^2 - 14x_1 - 6x_2$$

subject to $\quad x_1 + x_2 - 2 \leq 0$

$\quad 2x_2 + x_1 - 3 \leq 0$

Solution: $\nabla f(\underline{x}) = \begin{bmatrix} 2x_1 - 14 \\ 2x_2 - 6 \end{bmatrix}$, $\nabla \underline{g}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and $\nabla \underline{g}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

necessary optimality conditions

$$\nabla f(\underline{x}) + \lambda_1 \nabla \underline{g}_1(\underline{x}) + \lambda_2 \nabla \underline{g}_2(\underline{x}) = 0, \quad g_1(\underline{x}) \leq 0, \quad g_2(\underline{x}) \leq 0$$

4 equations in 4 unknowns.

Example (Cont'd)

4 Cases to Consider:

a) Both constraints are not active ($\lambda_1 = \lambda_2 = 0$)

$$\nabla f(\underline{x}^*) = 0 \Rightarrow \underline{x}^* = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$ (Both constraints violated)

b) Only First constraint active ($\lambda_2 = 0,\ \lambda_1 \neq 0$)

$$\nabla f(\underline{x}^*) + \lambda \nabla \underline{g}_1 = 0\ , \quad \underline{g}_1(\underline{x}) = 0$$

$$\Rightarrow \begin{bmatrix} 2x_1 - 14 \\ 2x_2 - 6 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad x_1 + x_2 - 2 = 0$$

Example (Cont'd)

substituting in constraint, we get

$$\left(7-\frac{\lambda}{2}\right)+\left(3-\frac{\lambda}{2}\right)=2\ ,\quad \lambda=8 \Rightarrow \underline{x}^{*}=\begin{bmatrix}3\\-1\end{bmatrix}$$

KKT Conditions & second constraint satisfied

c) If only second constraint is active

$$\underline{\nabla} f(\underline{x}^{*})+\lambda_2 \nabla \underline{g}_2(\underline{x}^{*})=\underline{0}\ ,\quad g_2(\underline{x})=0$$

$$\begin{bmatrix}2x_1-14\\2x_2-6\end{bmatrix}+\lambda_2\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix},\quad 2x_2+x_1-3=0$$

Example (Cont'd)

substituting in constraint we get

$$(7 - \lambda_2/2) + 2(3 - \lambda_2) = 3 \Rightarrow \frac{5\lambda_2}{2} = 10 \Rightarrow \lambda_2 = 4$$

$$\Rightarrow \underline{x}^* = \begin{bmatrix} 5 \\ -1 \end{bmatrix} \Rightarrow \text{first constraint violated}$$

d) If both constraints are active, we have

$$x_2^* = 1, \quad x_1^* = 1, \quad \nabla f(\underline{x}^*) + \lambda_1 \nabla \underline{g}_1(\underline{x}^*) + \lambda_2 \nabla \underline{g}_2(\underline{x}^*) = 0$$

$$\begin{bmatrix} -12 \\ -4 \end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \underline{0} \Rightarrow \lambda_2 = -8 \text{ (rejected)}$$

Why Line Search?

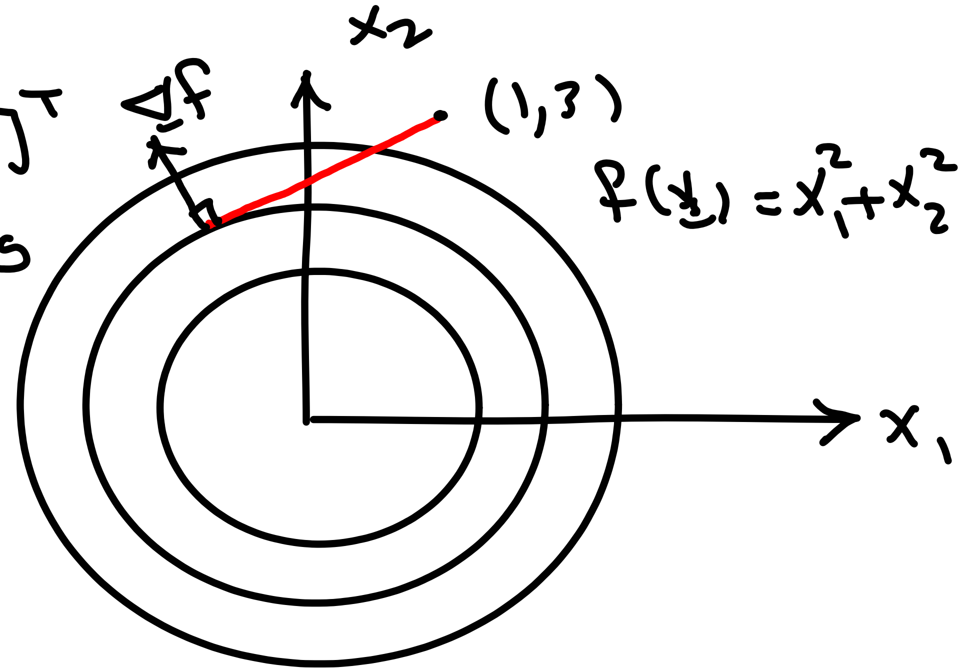
* Many optimization algorithms generate a direction $\underline{s}_k$ at the kth iteration.
* Starting from the current solution $\underline{x}_k$ we search in the direction of $\underline{s}_k$ to find the minimum along that direction

$$\Rightarrow \min_{\lambda} f(\underline{x}_k + \lambda \underline{s}_k)$$

* Get $\underline{x}_{k+1} = \underline{x}_k + \lambda^* \underline{s}_k$

Analytical Example

* starting with $\underline{x}^{(0)} = [1 \quad 3]^T$ find the minimum along the direction $\underline{s} = [-2 \quad -1]^T$.



* $\underline{x} = \underline{x}^{(0)} + \lambda \underline{s} = \begin{bmatrix} 1-2\lambda \\ 3-\lambda \end{bmatrix}$

$f(\lambda) = (1-2\lambda)^2 + (3-\lambda)^2$

$\frac{\partial f}{\partial \lambda} = 0 \Rightarrow -4(1-2\lambda) - 2(3-\lambda) = 0 \Rightarrow \lambda^* = 1$

$\underline{x}^* = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \Rightarrow \nabla f)_{\underline{x}^*} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}_{\underline{x}^*} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$ (why $\nabla f)_{\underline{x}^*} \perp \underline{s}$?)

Numerical Line Search

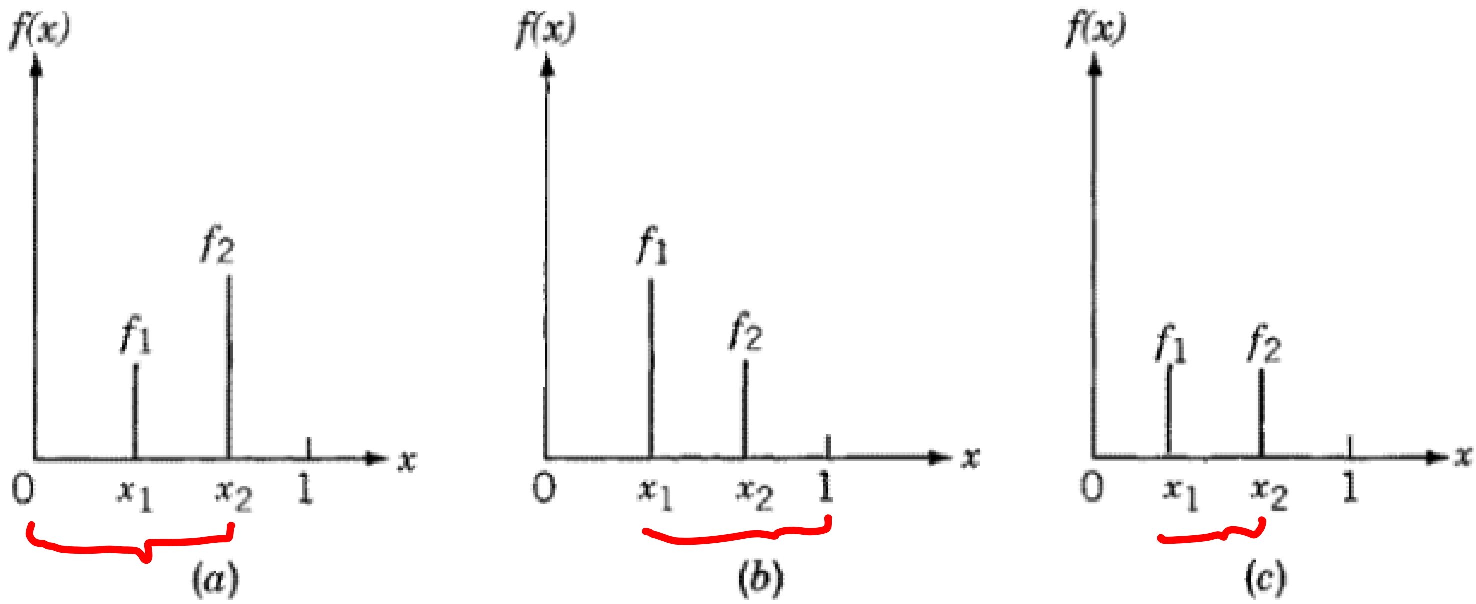
* In this case, the function is not known analytically but rather as values
* Our target is to bracket the solution within a sufficiently small region
* Some of these techniques utilize an interpolation model to predict the minimum location

UNIMODALITY

* A unimodal function has only one minimum within the region of interest $\Rightarrow$ Function values can inform us of how close we are to the solution!

* If $x_1 < x_2 < x^* \Rightarrow f(x_2) < f(x_1)$, and
  if $x_2 > x_1 > x^* \Rightarrow f(x_1) < f(x_2)$.

Unimodality Illustration

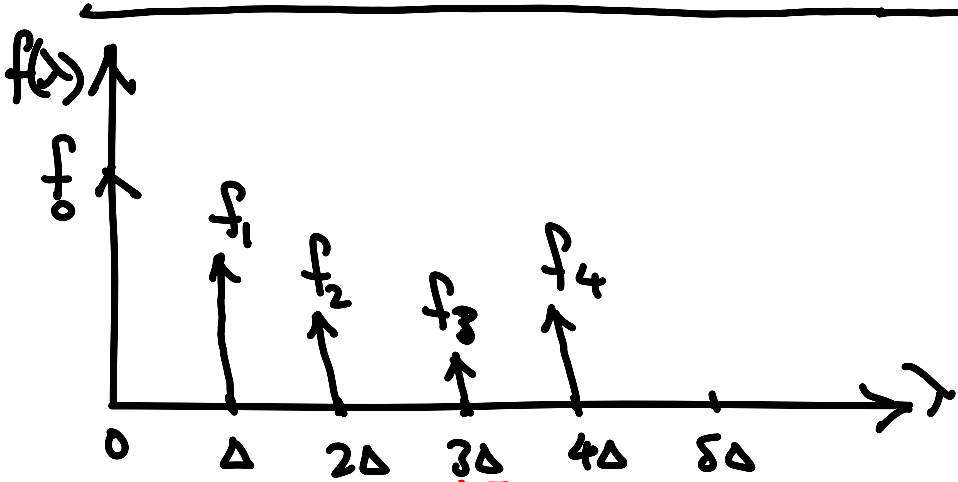


The new solution interval is determined through function values!

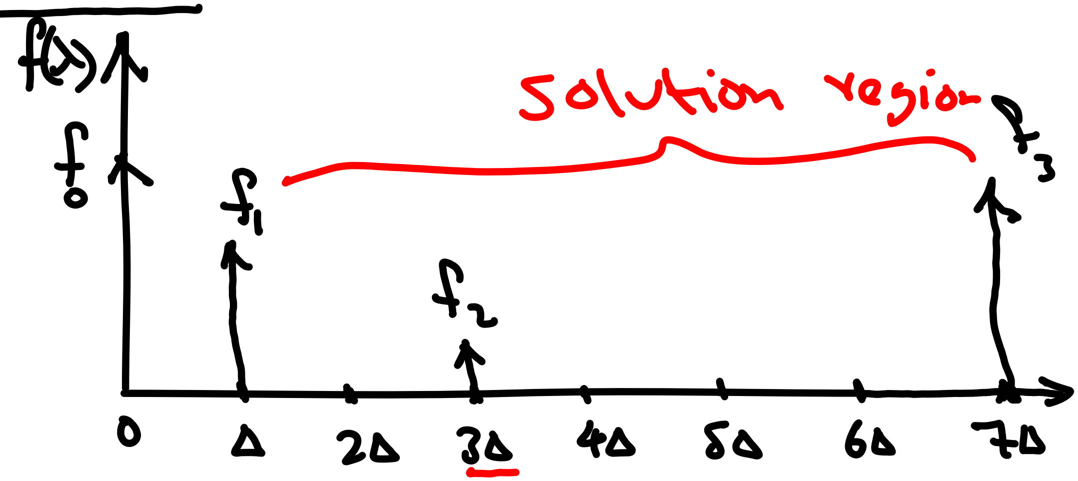
Unrestricted Search

* In this case, we do not know beforhand an interval for the solution (e.g. $0 \leq \lambda \leq 1$).
* In most practical cases, we do know a useful interval for the solution
* The main difficulty with this type of technique is the selection of the step size Δ.

Unrestricted Search (Cont'd)



Fixed step size

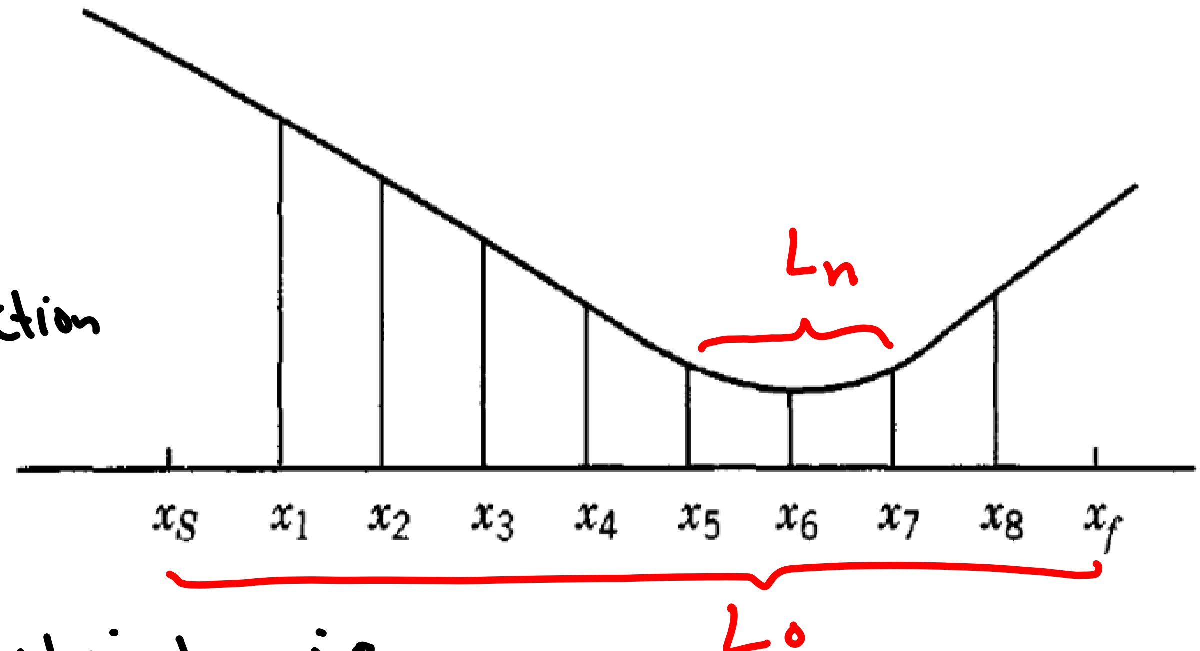


Accelerated step size

* Fixed step size requires a guess for a suitable Δ.
* Accelerated step size requires a further bracketing technique.

Exhaustive Search

* Start with an initial Interval L_0
* Evaluate the function at n equally spaced points

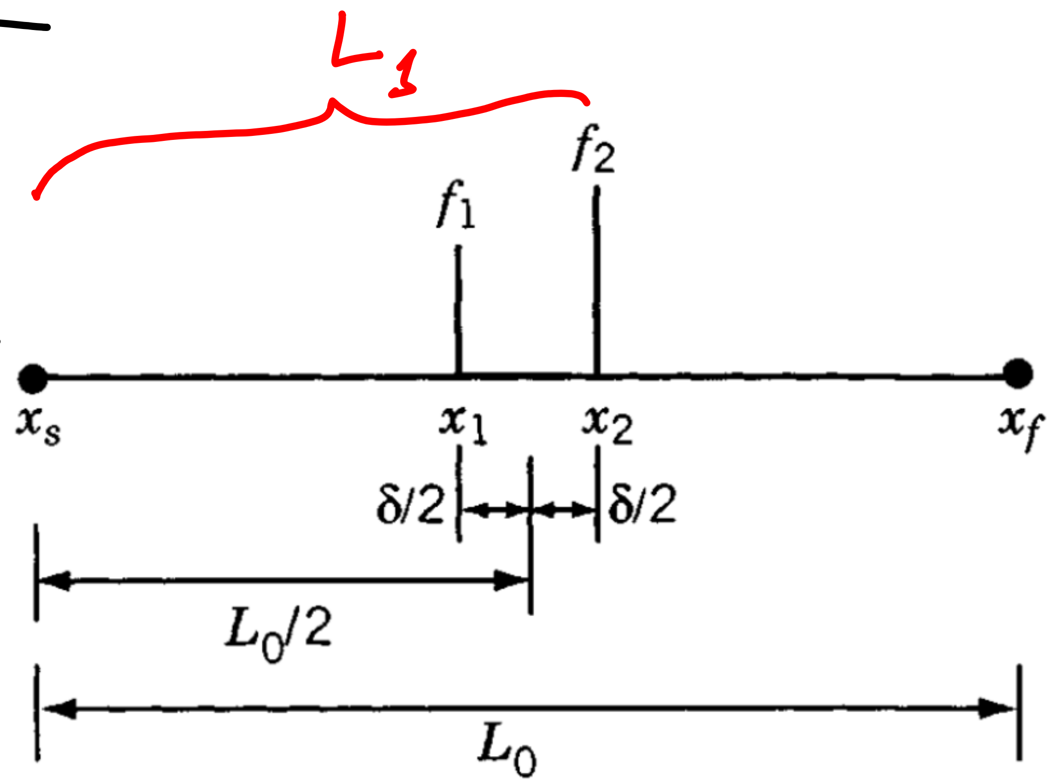


* Interval of uncertainty is

$$L_n = \frac{2}{(n+1)} L_0$$

Dichotomous Search

* At every step, two function values are evaluated around the centre of the region

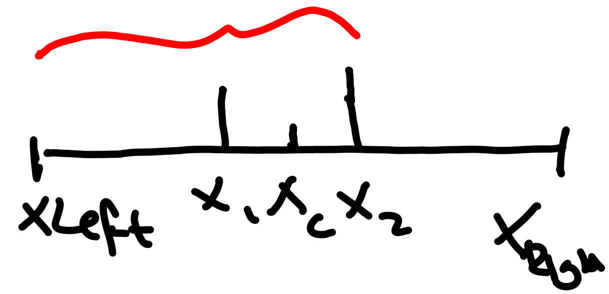


* Depending on these two values, a new interval is determined
* One function value is wasted per iteration.

Matlab Code

```matlab
%an implementation of the
Dichotomous search
xLeft=0;
xRight=1.0
delta=1.0e-3;
xCenter=0.5*(xLeft+xRight)
x1=xCenter-0.5*delta;
x2=xCenter+0.5*delta;
fLeft=xLeft*(xLeft-1.5);
fRight=xRight*(xRight-1.5);
f1=x1*(x1-1.5);
f2=x2*(x2-1.5);
L=xRight-xLeft;
Epsilon=1.0e-2;
while(L>Epsilon)
   if(f2>=f1)
       xRight=x2;
       fRight=f2;
       xCenter=0.5*(xLeft+xRight);
       x1=xCenter-0.5*delta;
       x2=xCenter+0.5*delta;
       f1=x1*(x1-1.5);
       f2=x2*(x2-1.5);
       L=xRight-xLeft;
else

       xLeft=x1;
       fLeft=f1;
       xCenter=0.5*(xLeft+xRight);
       x1=xCenter-0.5*delta;
       x2=xCenter+0.5*delta;
       f1=x1*(x1-1.5);
       f2=x2*(x2-1.5);
       L=xRight-xLeft;
   end
end
xLeft
xRight
```



final result

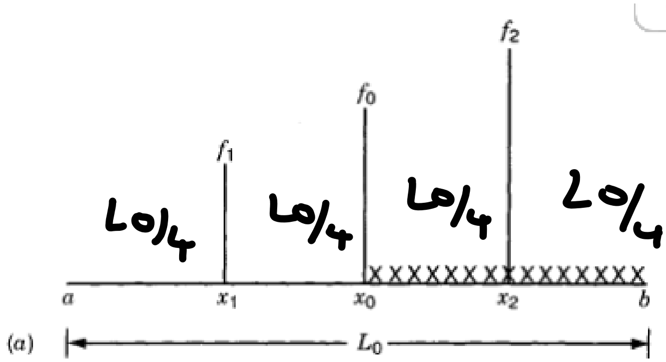
xLeft =

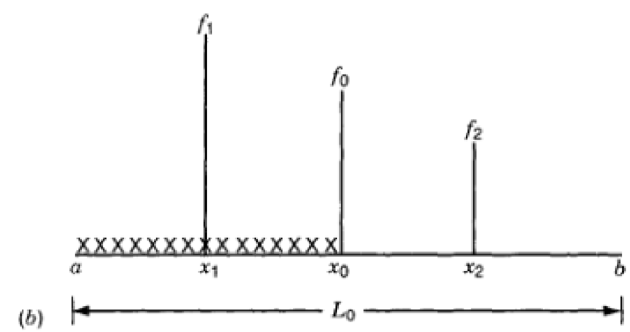
0.7493

xRight =

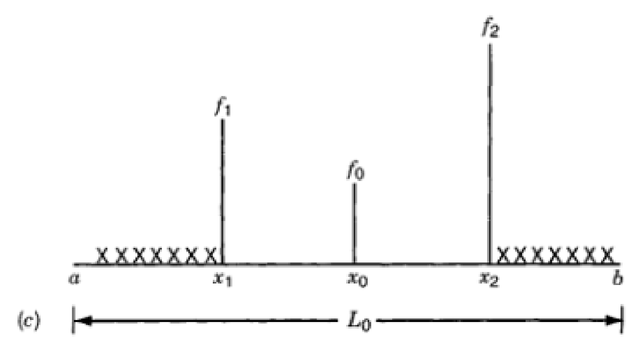
0.7581

Interval Halving Method





* Using equally spaced values, interval is halved
* $L_n = \left(\frac{1}{2}\right)^n L_0$
* Function values are reused



Example

Find the minimum of the function $f(\lambda) = \lambda^5 - 5\lambda^3 - 20\lambda + 5$ using interval halving in the interval $(0, 5)$

Analytical solution:

$$\frac{\partial f}{\partial \lambda} = 0 \Rightarrow 5\lambda^4 - 15\lambda^2 - 20 = 0 \Rightarrow \lambda^4 - 3\lambda^2 - 4 = 0$$

$$(\lambda^2 - 4)(\lambda^2 + 1) = 0 \Rightarrow \lambda^* = +2 \text{ in interval}$$

$$\left.\frac{\partial^2 f}{\partial \lambda^2}\right)_{\lambda^*} = 20\lambda^3 - 30\lambda\Big)_{\lambda^*} = 100 > 0 \quad \text{local minimum}$$

Example (cont'd)

Results from a Matlab Code

xLeft = 0   x1=1.2500   xCenter=2.5000   x2=3.7500 xRight=5

xLeft = 0   x1 = 0.6250 xCenter =1.2500 x2 = 1.8750 xRight =2.5000

xLeft =1.2500 x1 =1.5625 xCenter=1.8750 x2 =2.1875 xRight =2.5000

xLeft =1.5625 x1 =1.7969   xCenter =1.8750 x2 =1.9531 xRight =2.1875

xLeft =1.9922 x1 =1.9971   xCenter =2.0068 x2 =2.0166 xRight =2.0313

Fibonacci Numbers

* Fibonacci sequence is a sequence of numbers satisfying the recursion relationship

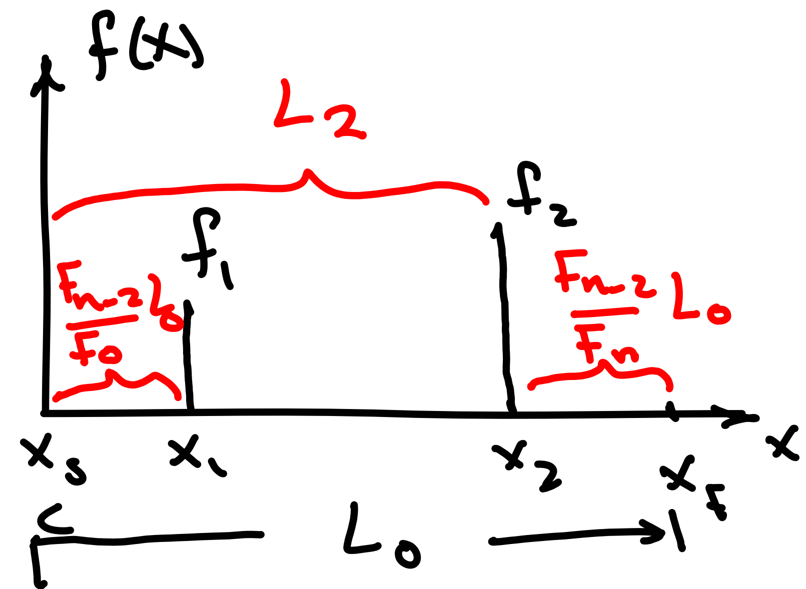
$$F_n = F_{n-1} + F_{n-2}, \text{ with } F_0 = 1,\ F_1 = 1$$

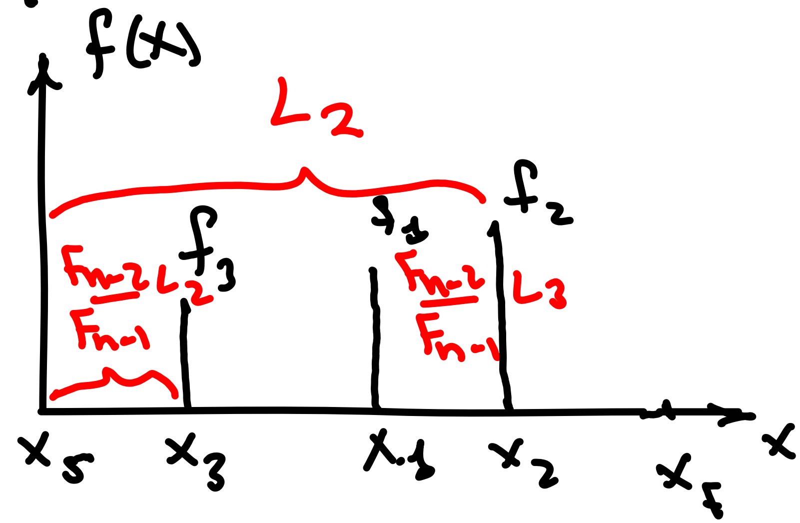
1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, ·

* This sequence is used to divide each interval to two sub-intervals.

Fibonacci Method

* Start with an interval L_0 and choose n
* Get $L_2 = \frac{F_{n-1}}{F_n} L_0$
* Use modality to remove part of the interval
* These steps are repeated until all n steps are done.





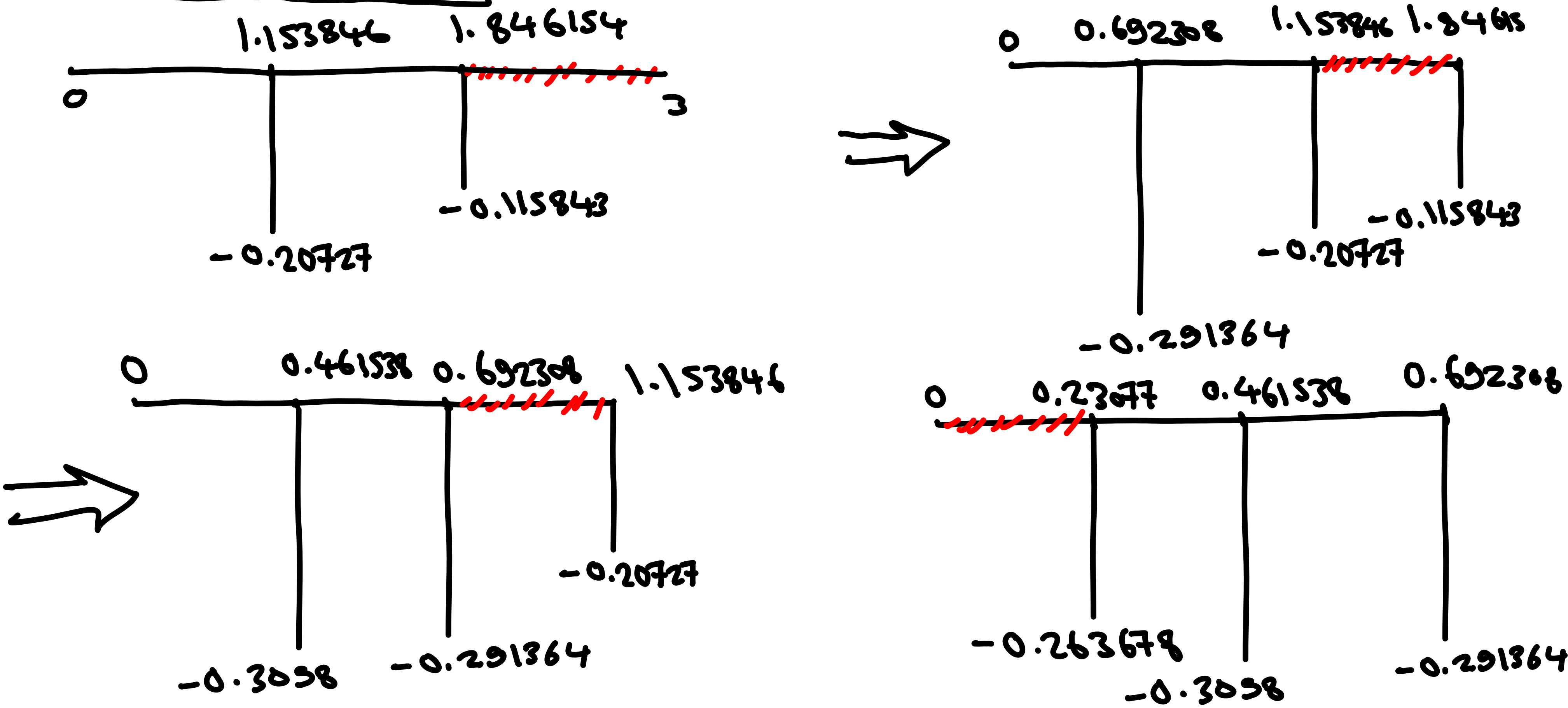
Example :

minimize $f(x) = 0.65 - \frac{0.75}{1+x^2} - 0.65x\tan^{-1}\left(\frac{1}{x}\right)$ in the interval $[0, 3]$ using the Fibonacci method with $N=6$.

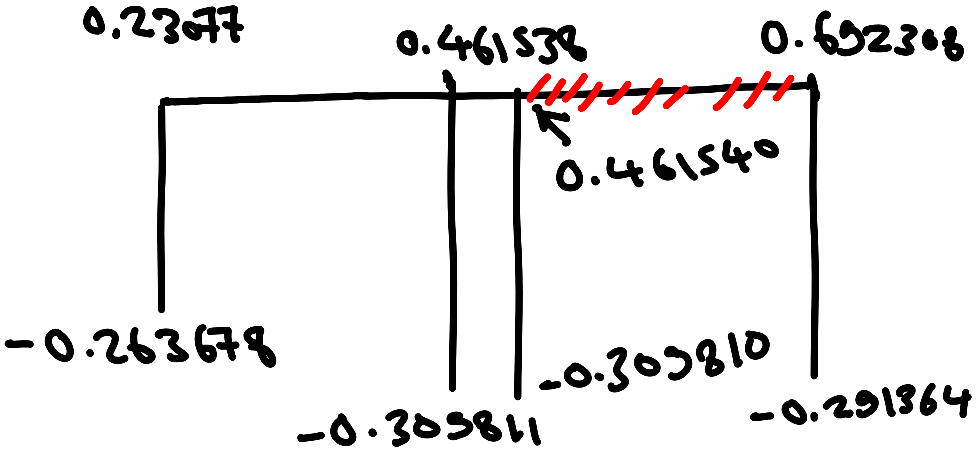
Solution: First two experiments are placed at a distance $d = \left(\frac{F_4}{F_6}\right) * 3 = \frac{5}{13} * 3 = 1.153846$ from the interval edges.

$\Rightarrow x_1 = 1.153846, \quad x_2 = 3 - 1.153846 = 1.846154$

Example (Cont'd)



Example (Cont'd)



$$L_6 = [0.23077, 0.46540]$$

$$\frac{L_6}{L_1} = \frac{0.461540 - 0.23077}{3.0} = 0.076923$$

The Golden Section Method

* The Fibonacci sequence is given by
  1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, ...

$$F_8 / F_9 = \frac{34}{55} = 0.618$$

$$F_9 / F_{10} = \frac{55}{89} = 0.6179$$

actually $\lim_{N \to \infty} \frac{F_{N-1}}{F_N} \approx 0.618$

* The ratio 0.618 is then used at all steps!

Example: Use the Golden Section search to find the minimum of the function

$$f(x) = x^4 - 14x^3 + 60x^2 - 70x$$

in the rang $[0,2]$. Locate the value x^* to within a range of 0.3

Solution: $0.3 = 2(0.61803)^N \Rightarrow N = 4$

four iterations are reqvired

Example (Cont'd)